UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Scuola di Dottorato di Scienze

Corso di Dottorato di Ricerca in Fisica ed Astronomia

"Search for new particles decaying into vector boson pairs at LHC with the CMS detector and prospects for the vector boson scattering at the future High Luminosity LHC"

Coordinatore Scuola di Dottorato : **Prof. Giberto Chirico**

Supervisor : **Prof. Alessio Ghezzi**

Assistant Supervisor: **Dott. Pietro Govoni**

Tesi di Dottorato di:

Raffaele Gerosa

Matricola : 709921

Ciclo XXVIII

Anno Accademico 2015-2016

Contents

Executive Summary

This thesis presents a three year work inside the Compact Muon Solenoid (CMS) collaboration, mainly focused on data analysis from proton-proton collisions collected by the Large Hadron Collider (LHC) at a centre of mass energy of 8 TeV (LHC Run-I).
My efforts were dedicated to searches for heavy particles decaying into di-boson pairs exploring boosted topologies, where the fragmentation products of hadronically decaying vector bosons ($V \to qq'$) overlap in the detector, preventing their identification as resolved jets. Several benchmark scenarios for new physics have been considered, from heavy Standard Model (SM) like Higgs boson to beyond the SM (BSM) extensions of the Higgs sector and massive gravitons predicted by Warped Extra Dimension (WED) models.
In order to explore these boosted signatures, I developed specific algorithms to identify boosted vector bosons decaying hadronically (V-jets), which are also applied in other searches for new physics characterized high p_T jets originating from W/Z-boson, Higgs boson or top quark decays. These algorithms have been improved and new methods have been tested on detailed simulation, reproducing the conditions of the present data taking (LHC Run-II), and finally commissioned with first 13 TeV collisions.
Eventually, I performed feasibility studies for the Vector Boson Scattering (VBS) in view of the CMS detector upgrade, foreseen for the high luminosity LHC (HL-LHC) data taking. If no evidence for new physics will be observed with Run-II collisions, this rare process would become crucial to search for BSM physics via precision electroweak measurements.

The CMS detector is one of the two general purpose experiments placed along the LHC ring at Cern, which achieved the Higgs boson discovery. Although no clear hints of BSM physics have been observed during Run-I, a number of puzzles, such as the origin of dark matter, the matter-antimatter asymmetry, the hierarchy problem and the electroweak vacuum stability, suggest that the SM cannot be considered the ultimate theory of nature. The LHC Run-II represents a unique opportunity for particle physics, where the machine will give access to an energy range never explored before, since the beam collisions will happen at a centre of mass energy of 13 TeV, allowing searches for new physics over a broad phenomenology, as well as to continue the precise measurements of SM processes.

The work presented here is organised in six chapters, where the first two are dedicated to a

brief introduction to the SM theory and BSM extensions, as well as to a description of the CMS experiment, while the analyses I performed are detailed in the following ones.
In particular, Chapter 1 presents the fundamental principles of the SM of elementary particles, focusing also on the experimental measurements inside the Higgs sector. A description of possible BSM models is reported to provide a concise theoretical background for the searches presented in the thesis. Finally, the VBS process is described highlighting its role in testing the electroweak symmetry breaking (EWSB) mechanism.
Chapter 2 is dedicated to the description of the LHC machine and to the key aspects of particle phenomenology at hadron colliders. The CMS experiment and each sub-detector it is composed of are also illustrated. Finally, identification algorithms and performances of the different physics objects (muons, electrons, jets and missing energy) are summarized.
In Chapter 3, I present my studies on identification techniques and properties of hight p_T jets produced by boosted hadronically decaying vector bosons, based on the 8 TeV data collected during Run-I, which still represent the reference benchmark inside the CMS community. Then, my work on vector boson jet tagging in view of the LHC Run-II is reported, with a particular emphasis on jet substructure correlations and performances in a high pileup environment. In fact, with the increase of the beam intensity foreseen during Run-II, a larger number of additional strong interactions (pileup), happening together with a hard scattering event, will degradate the performances and the resolution of jets and missing transverse energy. For these reasons, I worked also on the commissioning of a powerful pileup rejection method, called PileUp Per Particle Identification (PUPPI), for high p_T jets and missing energy, looking at the first Run-II data.
Chapter 4 is dedicated to a search for heavy resonances decaying into vector boson pairs exploring semi-leptonic signatures with 8 TeV data. I performed the WW $\to \ell\nu q\overline{q}$ analysis, as well as the combination with other exclusive searches in di-boson final states (ZZ $\to \ell\ell q\overline{q}$ and VV $\to q\overline{q}q'\overline{q}'$). With no significant excess observed above the predicted background, upper limits are produced for massive gravitons predicted by Bulk and Randall-Sundrum WED models. Eventually, a model independent re-interpretation of the analysis is also performed.
In Chapter 5, a search for a SM-like Higgs boson at high mass, $600 < m_{\mathrm{H}} < 1000$ GeV, in the boosted semi-leptonic WW final state is presented. The analysis is categorised in exclusive jet bins to disentangle the gluon fusion production from the vector boson fusion (VBF) mode. The result is also interpreted in the context of a BSM heavy Higgs scenario, called electroweak singlet model, which predicts the existence of two Higgs scalars with the mass of the lighter one being around 125 GeV. Upper limits on these models are set through an unbinned fit to the reconstructed di-boson invariant mass spectrum (m_{WW}), where a deviation with a local significance of about 2.5 σ has been observed in the VBF category for $700 < m_{\mathrm{WW}} < 800$ GeV.
Finally, Chapter 6 presents a preliminary performance evaluation for a Vector Boson Scattering (VBS) analysis in view of the high luminosity LHC data taking, considering two different topologies represented by the scattering of same sign W-boson pairs and the fully leptonic WZ boson one. The ultimate goal is to assess the future CMS sensitivity to the longitudinal electroweak scattering, to partial unitarized scenarios and to new physics in the EWSB sector, which is parametrized through anomalies in the vector boson self couplings. In addition, the difference in the performance between the proposed upgraded CMS detector and the present one, which will undergo a strong degradation due to radiation damage, is investigated showing how the upgraded technology allows better performances in all the considered benchmarks.

Chapter 1

Theoretical introduction: from the Higgs boson to BSM physics

Elementary particles and their interactions are described by a fundamental theory called Standard Model (SM). It is a renormalizable gauge theory developed in the framework of quantum field physics, which has been confirmed experimentally, up to the TeV scale, by several experiments: from the precision measurements performed by the Large Electron Positron machine (LEP) [1] to the recent Large Hadron Collider (LHC) era [2, 3, 4, 5]. In this Chapter, the electroweak sector of the SM is presented in Section 1.1, together with the Brout-Englert-Higgs (BEH) mechanism for the electroweak symmetry breaking (EWSB) in Section 1.2. In Section 1.3, a summary of the CMS experimental results on the SM Higgs boson are exposed.
In Section 1.4, a brief description of beyond the SM Higgs sector extensions is presented, while, in Section 1.5, an introduction to extra-dimensions models, considered in searches for heavy gravitons, are described. Eventually, in Section 1.6, the Vector Boson Scattering (VBS) process is presented, describing its importance in probing the unitarization of the EWSB mechanism.

1.1 The Gauge theory of weak interactions

Weak interactions at low energy are described by the Fermi effective Lagrangian, which consists of the product of a vector current and an axial vector one. For example, the transfer operator responsible for a nucleon β decay is given by [6]:

$$\mathcal{L} = -\frac{G_F^{(\beta)}}{\sqrt{2}} \cdot \left[\bar{\psi}_p \gamma^\mu (1 - \alpha \cdot \gamma^5)\psi_n\right]\left[\bar{\psi}_e \gamma_\mu (1 - \gamma^5)\psi_{\nu_e}\right] \tag{1.1}$$

where the parameters $G_F^{(\beta)}$ and α are estimated studying hyperon decays. This theory cannot be renormalized just from a simple dimensional point of view. In fact, since the action must to be dimensionless, the Fermi operator contains dimension-six mass terms that don't guarantee the unitary of the S-Matrix expansion:

$$[\mathcal{S}] = \left[\int dx^4 L(\psi, \partial_\mu \psi)\right] = [\text{mass}]^{-4} \times [L] \quad [\psi] = [\text{mass}]^{\frac{3}{2}} \rightarrow G_F = [\text{mass}]^{-2} \Leftrightarrow \hat{O}_F = [\text{mass}]^6$$

In order to build a correct theory describing weak couplings, it should be developed in the framework of local invariant gauge theories, in analogy with Quantum Electro Dynamics (QED).

This gauge theory has to reproduce, in the low energy limit, the local four-fermion interaction described via the Fermi Lagrangian.
This happens when the interaction, described in a gauge theory by the exchange of a massive vector boson with mass m_{V}, is characterized by a transferred momentum $q^2 \ll m_{\mathrm{V}}^2$. Since gauge theories are known to be renormalizable, the mass of this intermediate vector boson acts as a cut-off to the growth of cross section with energy [7]. The theory of electroweak interactions is defined to be local invariant under the gauge group $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$. The matter fields, which enter in this theory, are 1/2 spin particles (fermions), whose free Lagrangian is given by:

$$\mathcal{L}(\psi, \partial_\mu \psi) = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi = \bar{\psi}(\not{p} - m)\psi \tag{1.2}$$

where an operator ψ in the spinor representation of the Lorentz group[1] is assigned to each fermion. To build a $\mathrm{SU(2)_L}$ symmetry, the following chiral projectors are introduced:

$$\mathrm{P_{R/L}} = \frac{1}{2}(1 \pm \gamma^5) \rightarrow \begin{cases} \mathrm{P_R} + \mathrm{P_L} = 1 \\ \mathrm{P_R} \cdot \mathrm{P_L} = \mathrm{P_L} \cdot \mathrm{P_R} = 0 \\ \mathrm{P_R^2} = \mathrm{P_R} \quad \mathrm{P_L^2} = \mathrm{P_L} \end{cases} \Rightarrow \psi = \frac{1}{2}(1-\gamma^5)\psi + \frac{1}{2}(1+\gamma^5)\psi = \psi_{\mathrm{L}} + \psi_{\mathrm{R}}$$

where ψ_{L} and ψ_{R} are the left and the right chiral component of ψ. For sake of simplicity, one starts considering a system composed by a charged lepton and its neutrino (ψ^ℓ, ψ^{ν_ℓ}):

$$\begin{aligned} \mathcal{L} = \bar{\psi}^\ell(\not{p} - m_\ell)\psi^\ell + \bar{\psi}^{\nu_\ell}(\not{p} - m_{\nu_\ell})\psi^{\nu_\ell} = \bar{\psi}^\ell_{\mathrm{L}} \not{p}\, \psi^\ell_{\mathrm{L}} + \bar{\psi}^\ell_{\mathrm{R}} \not{p}\, \psi^\ell_{\mathrm{R}} + \bar{\psi}^{\nu_\ell}_{\mathrm{L}} \not{p}\, \psi^{\nu_\ell}_{\mathrm{L}} + \bar{\psi}^{\nu_\ell}_{\mathrm{R}} \not{p}\, \psi^{\nu_\ell}_{\mathrm{R}} \\ - m_\ell(\bar{\psi}^\ell_{\mathrm{L}}\psi^\ell_{\mathrm{R}} + \bar{\psi}^\ell_{\mathrm{R}}\psi^\ell_{\mathrm{L}}) - m_{\nu_\ell}(\bar{\psi}^{\nu_\ell}_{\mathrm{L}}\psi^{\nu_\ell}_{\mathrm{R}} + \bar{\psi}^{\nu_\ell}_{\mathrm{R}}\psi^{\nu_\ell}_{\mathrm{L}}) \end{aligned}$$

Neglecting the mass terms to reproduce the left handed Fermi current, one can assign the left handed fields to an SU(2) doublet Ψ_{L}, while the right handed ones remain scalars with respect to the action of the $\mathrm{SU(2)_L}$ group:

$$\mathcal{L} = \bar{\Psi}_{\mathrm{L}} \not{p}\, \Psi_{\mathrm{L}} + \bar{\psi}^\ell_{\mathrm{R}} \not{p}\, \psi^\ell_{\mathrm{R}} + \bar{\psi}^{\nu_\ell}_{\mathrm{R}} \not{p}\, \psi^{\nu_\ell}_{\mathrm{R}} \quad \begin{cases} \Psi_{\mathrm{L}} = \begin{bmatrix} \psi^{\nu_\ell}_{\mathrm{L}} \\ \psi^\ell_{\mathrm{L}} \end{bmatrix} \rightarrow \mathrm{SU(2)_L} \rightarrow \Psi'_{\mathrm{L}} = \Psi_{\mathrm{L}} \cdot e^{-(ig\vartheta_i\tau_i)/2} \\ \psi^\ell_{\mathrm{R}}, \psi^{\nu_\ell}_{\mathrm{R}} \rightarrow \mathrm{SU(2)_L} \rightarrow \psi^\ell_{\mathrm{R}}, \psi^{\nu_\ell}_{\mathrm{R}} \end{cases} \tag{1.3}$$

where g is a general coupling constant to the SU(2) group, while τ_i are the SU(2) generators in its fundamental representation which satisfy the commutation rule: $[\tau_i, \tau_j] = i\epsilon_{ijk}\tau_k$.

1.1.1 Local non-abelian Gauge symmetries

Considering N fields ϕ_i (scalars or spinors) and a SU(N) local transformation, a generic gauge theory has to be based on a Lagrangian invariant under such group:

$$\phi_i(x) \rightarrow \mathrm{SU(N)} \rightarrow \phi'_i(x) = U_{ij}(x)\phi_j(x) \tag{1.4}$$

In `QED`, where the invariance gauge group is abelian $[\mathrm{U(1)}_{em} \rightarrow \mathrm{U(x)} = e^{iq\vartheta(x)}]$, the kinetic term of the Dirac Lagrangian is not invariant under local phase transformations. The symmetry is then restored through the substitution of the ordinary derivate with the covariant one [8]:

$$D_\mu = \partial_\mu - igA_\mu \quad ; \quad D'_\mu = U(x)D_\mu U(x)^{-1} = D_\mu \quad ; \quad A'_\mu = A_\mu - \frac{i}{g}U(x)^{-1}\partial_\mu U(x)$$

[1] γ^μ are defined as a base of the Clifford algebra. In this chapter, the Chiral or Weyl representation is considered: $\gamma^0 = I_{2x2} \otimes \sigma_1$, $\gamma^i = \sigma_i \otimes i\sigma_2$ and $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$.

where A_μ is a massless vector field that interacts with fermion through the covariant derivative. To obtain an invariant Lagrangian, it is necessary to introduce a free kinetic term for the gauge field A_μ:

$$[D'_\mu, D'_\nu] = U[D_\mu, D_\nu]U^{-1} = -igU(\partial_\mu A_\nu - \partial_\nu A_\mu)U^{-1} = [D_\mu, D_\nu] \quad [D_\mu, D_\nu] = -igF_{\mu\nu}$$

Thus, in an Abelian gauge theory, the commutator of the covariant derivate is locally invariant and proportional to the strength $F_{\mu\nu}$ of the field A_μ.
In the non-Abelian case, like for the SU(N) group, the gauge field A_μ is an N×N matrix that undergoes the following transformation rule [8]:

$$D'_\mu = U(x)D_\mu U(x)^{-1} \rightarrow A'_\mu = U(x)A_\mu U^{-1}(x) - ig^{-1}U(x)\partial_\mu U(x)^{-1} \tag{1.5}$$

In analogy with the Abelian case, the gauge field strength is defined to be:

$$F_{\mu\nu} = -ig^{-1}[D_\mu, D_\nu] = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu] \tag{1.6}$$

that is no longer gauge invariant, but its trace meets the invariance condition:

$$F'_{\mu\nu} = UF_{\mu\nu}U^\dagger \neq F_{\mu\nu} \quad \mathrm{Tr}(F'_{\mu\nu}F'^{\mu\nu}) = \mathrm{Tr}(UF_{\mu\nu}U^\dagger UF^{\mu\nu}U^\dagger) = \mathrm{Tr}(U^\dagger UF_{\mu\nu}F^{\mu\nu}) = \mathrm{Tr}(F_{\mu\nu}F^{\mu\nu})$$

The field A_μ is then expanded over a base of generators t^a for the gauge group, obtaining:

$$A_\mu = A^a_\mu t^a \quad F_{\mu\nu} = F^a_{\mu\nu}t^a \quad ; \quad \begin{cases} 1)\ (t^a)^\dagger = t^a \\ 2)\ \mathrm{Tr}(t^a) = 0 \\ 3)\ \mathrm{Tr}(t^a t^b) = t^a_{ij}t^b_{ji} = \frac{\delta_{ab}}{2} \rightarrow \\ 4)\ [t^a, t^b] = if^{abc}t^c \end{cases} \quad \begin{cases} A^a_\mu = 2\cdot\mathrm{Tr}(A_\mu t^a) \\ F^a_{\mu\nu} = 2\cdot\mathrm{Tr}(F_{\mu\nu}t^a) \end{cases}$$

Writing $F_{\mu\nu}$ as a function of A_μ, as in Equation 1.6, and using the Lie algebra definition, one finds:

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu$$

where a set of gauge fields A^a_μ, introduced in the theory, are massless. Mass terms, which have the form $\frac{1}{2}m_A A^a_\mu A^\mu_a$, cannot be added without spoiling the local symmetry, as indicated by the transformation rule reported in Equation 1.5.

1.1.2 Local invariance in the electroweak sector

Starting from the Lagrangian defined in Equation 1.3 and requiring local invariance under the $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ group, a covariant derivate dependent on four additional gauge fields must be introduced [9, 10]:

$$D_\mu = \partial_\mu - \frac{ig}{2}\tau^i W^i_\mu - \frac{ig'Y}{2}B_\mu \rightarrow L = L_{\mathrm{Kin}} + L_{\mathrm{CC}} + L_{\mathrm{NC}} \tag{1.7}$$

$$\mathcal{L}_{\mathrm{CC}} = \frac{g}{2}\bar{\Psi}_L\gamma^\mu\left[\tau^1 W^1_\mu + \tau^2 W^2_\mu\right]\Psi_L \quad \mathcal{L}_{\mathrm{NC}} = \frac{g}{2}\bar{\Psi}_L\gamma^\mu\tau^3 W^3_\mu\Psi_L + \frac{g'Y}{2}B_\mu\left[\bar{\psi}^\ell\gamma^\mu\psi^\ell + \bar{\psi}^{\nu_\ell}\gamma^\mu\psi^{\nu_\ell}\right]$$

where the triplet W_μ^i is related to the SU(2) generators, while the singlet B_μ is connected with the U(1) symmetry. The charged current Lagrangian $\mathcal{L}_{\rm CC}$ is then written in terms of charged vector fields $W_\mu^\pm$:

$$\begin{cases} \tau^\pm = \frac{1}{2}[\tau^1 \pm i\tau^2] \\ W_\mu^\pm = \frac{1}{\sqrt{2}}[W_\mu^1 \pm iW_\mu^2] \end{cases} \qquad \mathcal{L}_{\rm CC} = \frac{g}{\sqrt{2}}[\bar{\Psi}_L \gamma^\mu \tau^\pm \Psi_L W_\mu^\mp] = \frac{g}{\sqrt{2}}[J_{\rm CC}^{\mu+} W_\mu^- + J_{\rm CC}^{\mu-} W_\mu^+] \tag{1.8}$$

where $J_{\rm CC}^{\mu\pm}$ represents the weak charged current through which left handed fermions are coupled with charged vector bosons $\rm W^+$ and $\rm W^-$:

$$J_{\rm CC}^{\mu+} = \bar{\psi}_{\rm L}^{\nu_\ell} \gamma^\mu \psi_{\rm L}^\ell \qquad\qquad J_{\rm CC}^{\mu-} = \bar{\psi}_{\rm L}^{\ell} \gamma^\mu \psi_{\rm L}^{\nu_\ell} \tag{1.9}$$

The local gauge invariance under the chosen group originates also a neutral current $\mathcal{L}_{\rm NC}$, which is not predicted inside the Fermi model. This contribution cannot be directly identified with the electromagnetic current, since it contains neutrino terms and it does not involve left-handed and right-handed fermions with the same weight. For this reason, $\mathcal{L}_{\rm NC}$ is translated according to the following notation:

$$\Psi = \begin{bmatrix} \psi_{\rm L}^{\nu_\ell} \\ \psi_{\rm L}^{\ell} \\ \psi_{\rm R}^{\nu_\ell} \\ \psi_{\rm R}^{\ell} \end{bmatrix} \qquad T_3 = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad Y = \begin{bmatrix} Y(L) & 0 & 0 & 0 \\ 0 & Y(L) & 0 & 0 \\ 0 & 0 & Y(\nu_{\rm R}) & 0 \\ 0 & 0 & 0 & Y(\ell_{\rm R}) \end{bmatrix}$$

$$\mathcal{L}_{\rm NC} = g\bar{\Psi}\gamma^\mu T_3 \Psi W_\mu^3 + g'\bar{\Psi}\gamma^\mu \frac{Y}{2} \Psi B_\mu \tag{1.10}$$

In order to recover the electromagnetic current term, an orthogonal transformation is performed in the two-dimensional space of the neutral gauge fields W_μ^3 and B_μ:

$$\begin{bmatrix} B_\mu \\ W_\mu^3 \end{bmatrix} = \begin{bmatrix} \cos\vartheta_{\rm W} & \sin\vartheta_{\rm W} \\ -\sin\vartheta_{\rm W} & \cos\vartheta_{\rm W} \end{bmatrix} \begin{bmatrix} A_\mu \\ Z_\mu \end{bmatrix} \tag{1.11}$$

This depends only on one mixing angle, called Weinberg angle, measured to be $\sin^2\vartheta_{\rm W} = 0.2315 \pm 0.0002$. In terms of Z_μ and A_μ, Equation 1.10 takes the following form:

$$\mathcal{L}_{\rm NC} = \bar{\Psi}\gamma^\mu \left[g\sin\vartheta_{\rm W} T_3 + g'\cos\vartheta_{\rm W} \frac{Y}{2} \right] \Psi A_\mu + \bar{\Psi}\gamma^\mu \left[g\cos\vartheta_{\rm W} T_3 - g'\sin\vartheta_{\rm W} \frac{Y}{2} \right] \Psi Z_\mu \tag{1.12}$$

where one can identify A_μ as the photon field providing:

$$eQ = g\sin\vartheta_{\rm W} T_3 + g'\cos\vartheta_{\rm W} \frac{Y}{2} \tag{1.13}$$

The conventional adopted choice is to set the hypercharge of the left chiral fields as Y(L) =-1. In this way, from the left-handed terms in Equation 1.13, one obtains the Gell-Mann-Nishijima formula:

$$\begin{cases} \nu_l \ : \ \frac{g}{2}\sin\vartheta_{\rm W} - \frac{g'}{2}\cos\vartheta_{\rm W} = 0 \\ l \ : \ -\frac{g}{2}\sin\vartheta_{\rm W} - \frac{g'}{2}\cos\vartheta_{\rm W} = -e \end{cases} \rightarrow g\sin\vartheta_{\rm W} = g'\cos\vartheta_{\rm W} = e \quad \rightarrow \quad Q = T_3 + \frac{Y}{2} \tag{1.14}$$

Table 1.1 reports all the different matter fields (leptons and quarks), that enter in the SM picture of elementary particle interactions.

Leptons			Q	T_3	Y	Interaction	Quarks				Q	T_3	Y	Interaction
ν_{L}^{e}	ν_{L}^{μ}	ν_{L}^{τ}	0	$\frac{1}{2}$	-1	Weak	$u_{\mathrm{L}}^{i} =$	u_{L}	c_{L}	t_{L}	$\frac{2}{3}$	$+\frac{1}{2}$	$\frac{1}{3}$	Weak, EM, Strong
e_{L}	μ_{L}	τ_{L}	-1	$-\frac{1}{2}$	-1	Weak, EM	$d_{\mathrm{L}}^{i} =$	d_{L}	s_{L}	b_{L}	$-\frac{1}{3}$	$-\frac{1}{2}$	$\frac{1}{3}$	Weak, EM, Strong
ν_{R}^{e}	ν_{R}^{μ}	ν_{R}^{τ}	0	0	0	No Interaction	$u_{\mathrm{R}}^{i} =$	u_{R}	c_{R}	t_{R}	$\frac{2}{3}$	0	$\frac{4}{3}$	Weak, EM, Strong
e_{R}	μ_{R}	τ_{R}	-1	0	-2	Weak, EM	$d_{\mathrm{R}}^{i} =$	d_{R}	s_{R}	b_{R}	$-\frac{1}{3}$	0	$-\frac{2}{3}$	Weak, EM, Strong

Table 1.1: The building blocks of the SM: leptons and quarks.

1.1.3 Vector boson sector

The electroweak Lagrangian must include also the kinetic terms for the gauge fields:

$$\mathcal{L}_B = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^a_{\mu\nu}W_a^{\mu\nu} \qquad \begin{cases} B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \\ W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g f^{abc} W^b_\mu W^c_\nu \quad f^{abc} = \epsilon^{abc} \end{cases}$$

The Lagrangian $\mathcal{L}_B$ is usually expressed, through Equation 1.8 and Equation 1.11, in terms of the physical fields A_μ, Z_μ and $W^\pm_\mu$, showing trilinear and quadrilinear self interactions among vector bosons, which is a consequence of the chosen non-Abelian gauge group ($\mathrm{SU(2)_L}$).
Even if first studies of these interactions have been already performed also at the LHC [11, 12], precision measurements of trilinear (WWZ, WWγ) and quadrilinear (WW$\gamma\gamma$, WWZZ, WWγZ, WWWW) gauge couplings represent an important milestones for the future physics at hadron colliders. So far, fermions and vector bosons are forced to be massless inside the theory, since mass terms will spoil the local invariance. A possible solution of this conflict between massless particles, as required by the theory, and massive fermions and vector bosons, as observed experimentally, is provided by the EWSB mechanism.

1.2 The Higgs model for the electroweak symmetry breaking

One necessary condition for the electroweak theory, to reproduce in the low energy limit the Fermi model, consists of a non zero-mass prediction for the W-bosons. In fact, one can set a lower bound to m_{W} comparing the μ decay amplitude, expressed in the Fermi theory, with the SM one expressed in the low energy limit [$q^2 \ll m^2_{\mathrm{W}} \Rightarrow q^2 \to 0$] [13]:

$$\frac{G_F}{\sqrt{2}} = \left(\frac{g}{2\sqrt{2}}\right)^2 \frac{1}{m^2_{\mathrm{W}}} \to m_{\mathrm{W}} \geq 37.3 \text{ GeV}$$

which is much larger than the actual photon mass upper bound $m_\gamma \leq 1.1 \cdot 10^{-18}$ eV [14].
To introduce gauge boson masses in the Lagrangian, the symmetry can be explicitly broken adding a mass term by hand, which would lead to a non-renormalizable theory.
In a quantum field theory, to any continuous global symmetry of the Lagrangian, a conserved current and a consequent conservation law is associated. This principle is known as Nother's theorem [15]:

$$\phi \to \phi + i\delta\vartheta^a t^a \phi \quad \to \quad \mathcal{L}'(\phi', \partial_\mu\phi') = \mathcal{L}(\phi, \partial_\mu\phi) \quad \Rightarrow \quad J^\mu = \frac{\partial L}{\partial(\partial_\mu\phi)}\delta\phi \ : \ \partial_\mu J^\mu = 0$$

where the Lorentz scalar charge $Q = \int d^3x \ J^0(\vec{x}, t)$ is conserved ($\dot{Q} = 0$). Taking the Hamiltonian $\mathcal{H}$, a global symmetry of the system and its related charge operator $\hat{Q}$, then:

- **Wigner-Weyl symmetry:** if $[\hat{Q}, \mathcal{H}] = 0$ and $\hat{Q}|0\rangle = 0$, the spectrum falls in multiplets of the symmetry group, therefore particles belonging to the same multiplet have the same mass.
- **Nambu-Goldstone symmetry:** if $[\hat{Q}, \mathcal{H}] = 0$ and $\hat{Q}|0\rangle \neq 0$, the symmetry is not manifested in the spectrum and is said to be spontaneously broken.

The most common way through which a spontaneous symmetry breaking is realized consists of introducing a set of field operators ϕ_k, which satisfy a non-trivial transformation under the gauge group and a non vanishing vacuum expectation value (`VEV`) $\langle 0|\phi_k|0\rangle = v_k \neq 0$.
Since the space is isotropic, the fields ϕ_k must be scalars otherwise the `VEV` would be frame dependent. Furthermore, we must be careful not to break the U(1) invariance related to `QED`, to keep the photon massless achieving a unified theory.
Definitely, the easiest way is represented by the minimal Higgs mechanism where, introducing a scalar field (ϕ) in the doublet representation of SU(2), a new Lagrangian term $\mathcal{L}_H$ invariant under the electroweak gauge symmetry is added as [16, 17, 18, 19, 20]:

$$\phi = \begin{bmatrix}\phi_1 \\ \phi_2\end{bmatrix} \quad ; \quad D_\mu = \partial_\mu - \frac{ig}{2}\tau^i W^i_\mu - \frac{ig'Y}{2}B_\mu \quad ; \quad V(\phi, \phi^\dagger) = \mu^2\phi\phi^\dagger + \lambda(\phi\phi^\dagger)^2 \tag{1.15}$$

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi, \phi^\dagger) \tag{1.16}$$

The chosen Higgs potential $V(\phi, \phi^\dagger)$[2] depends on μ^2: for $\mu^2 > 0$ the scalar potential has a global minimum at $|\phi_0|^2 = \langle 0|\phi|0\rangle = 0$ while, for $\mu^2 < 0$, it has an infinite number of global minima characterized by:

$$|\phi_0|^2 = \langle 0|\phi_0|0\rangle = -\frac{\mu^2}{2\lambda} = \frac{1}{2}v^2 \quad ; \quad \phi^0 = \frac{1}{\sqrt{2}}\begin{bmatrix}v_1 \\ v_2\end{bmatrix} \quad ; \quad |v_1|^2 + |v_2|^2 = v^2$$

where v is the vacuum expectation value for the Higgs field ϕ. The expected form of the potential is sketched in Figure 1.1 where, for $\mu^2 > 0$, all the minimum configurations are connected by a gauge transformation, changing the complex phase of ϕ without affecting its modulus. When a particular configuration is selected, the Higgs field is expanded around the minimum obtaining a renormalizable perturbation theory. As a consequence, the symmetry is said to be spontaneously broken. To keep the vacuum unchanged under $\mathrm{U}(1)_{em}$ transformations, the Higgs field hypercharge should be fixed as follows:

$$e^{ieQ\vartheta(x)}\phi_0 \approx (I_{2\times2} + ieI_{2\times2}\cdot\vec{Q})\phi_0 \Rightarrow \begin{bmatrix}Q_1 & 0 \\ 0 & Q_2\end{bmatrix}\begin{bmatrix}v_1 \\ v_2\end{bmatrix} = \begin{bmatrix}\frac{1}{2} + \frac{Y(\phi)}{2} & 0 \\ 0 & -\frac{1}{2} + \frac{Y(\phi)}{2}\end{bmatrix}\begin{bmatrix}v_1 \\ v_2\end{bmatrix} = \begin{bmatrix}0 \\ 0\end{bmatrix}$$

where there are two possible solutions to preserve $\mathrm{U}(1)_{em}$ symmetry:

$$1)\ v_1 = 0 \quad ; \quad |v_2| = v \quad ; \quad Y(\phi) = 1 \qquad 2)\ v_2 = 0 \quad ; \quad |v_1| = v \quad ; \quad Y(\phi) = -1 \tag{1.17}$$

Adopting the first choice $\mathrm{Y}(\phi) = 1$, $\mathrm{Q}_1 = 1$ and $\mathrm{Q}_2 = 0$, one obtains $v_2 = v$ real and positive. Expanding ϕ around this minimum, one finds:

$$\phi = \begin{bmatrix}\phi^+ \\ \phi^0\end{bmatrix} = \begin{bmatrix}\vartheta_1(x) + i\vartheta_2(x) \\ \frac{1}{\sqrt{2}}(v + h(x) + i\vartheta_3(x))\end{bmatrix} = \frac{1}{\sqrt{2}}e^{-\frac{i}{v}\tau^i\vartheta^i(x)}\begin{bmatrix}0 \\ h(x) + v\end{bmatrix} \tag{1.18}$$

[2] $V(\phi, \phi^\dagger) = \mu^2\phi\phi^\dagger + \lambda(\phi\phi^\dagger)^2$ is the most general assumption for the Higgs potential which is consistent with gauge invariance and renormalizability.

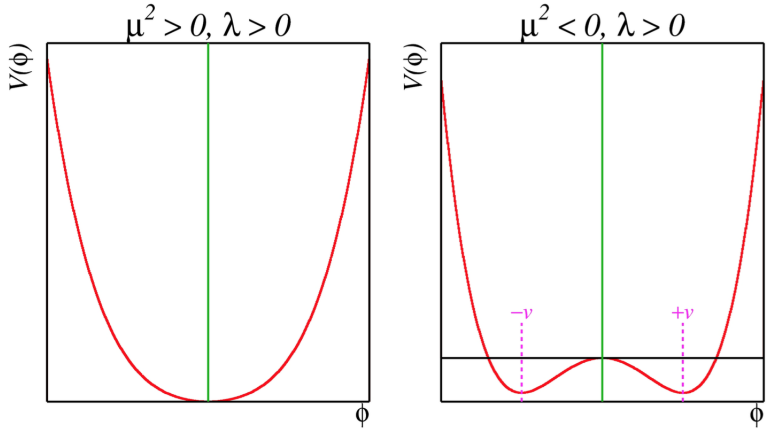

Figure 1.1: Form of the Higgs potential $V(\phi)$ as a function of on the μ^2 sign, positive on the left and negative on the right.

where the Goldstone modes $\vartheta^i(x)$ are reabsorbed through an $\mathrm{SU(2)_L}$ transformation, resulting in the so called unitary gauge. Inside the unitary gauge framework, the Higgs potential takes the following form:

$$V(\phi,\phi^\dagger) = \mu^2\phi\phi^\dagger + \lambda(\phi\phi^\dagger)^2 = \frac{1}{2}(2\lambda v^2)\cdot h(x)^2 + \lambda v\cdot h(x)^3 + \frac{\lambda}{4}\cdot h(x)^4 - \frac{\lambda}{4}v^4 \tag{1.19}$$

where the scalar real Higgs field $h(x)$ acquires a rest mass given by $m_h^2 = 2\lambda v^2$, which depends on the Higgs self coupling constant λ and the VEV v. In addition, there are also cubic ($\propto h^3(x)$) and quartic ($\propto h^4(x)$) Higgs self couplings and a constant term, called cosmological constant ($\rho_h = \lambda v^4/4$) that is not relevant for the SM. Substituting this gauge choice inside the covariant derivative in Equation 1.15, one finds:

$$D^\mu\phi = \frac{1}{\sqrt{2}}\Big(\begin{bmatrix}0\\ \partial_\mu h(x)\end{bmatrix} - \frac{i}{2}\Big(g\begin{bmatrix}W_3^\mu & W_1^\mu - iW_2^\mu\\ W_1^\mu + iW_2^\mu & -W_3^\mu\end{bmatrix} + g'B^\mu\Big)\begin{bmatrix}0\\ h(x)+v\end{bmatrix}\Big) =$$

$$\frac{1}{\sqrt{2}}\begin{bmatrix}0\\ \partial_\mu h(x)\end{bmatrix} - \frac{i}{2\sqrt{2}}\begin{bmatrix}v+h(x)\end{bmatrix}\begin{bmatrix}g(W_1^\mu - iW_2^\mu)\\ -gW_3^\mu + g'B^\mu\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}0\\ \partial_\mu h(x)\end{bmatrix} - \frac{i}{\sqrt{2}}\Big(1+\frac{h(x)}{v}\Big)\begin{bmatrix}gvW^{\mu+}\\ -v\sqrt{\frac{g^2+(g')^2}{2}}Z^\mu\end{bmatrix}$$

therefore:

$$(D_\mu\phi)^\dagger(D^\mu\phi) = \frac{1}{2}\partial_\mu h(x)\partial^\mu h(x) + \Big[\frac{g^2v^2}{4}W_\mu^+W^{\mu-} + \frac{1}{2}\frac{(g^2+g'^2)v^2}{4}Z^\mu Z_\mu\Big]\Big[1+\frac{h(x)}{v}\Big]^2 \tag{1.20}$$

Thanks to the spontaneous symmetry breaking and the unitary gauge choice, the three degrees of freedom related to the Goldstone modes $\vartheta^i(x)$ are transferred to the vector bosons ($W_\mu^\pm$ and Z_μ) through their mass terms, as reported in Equation 1.20. Thus, each vector boson becomes massive acquiring one additional polarization mode, keeping unchanged the total number of degrees of freedom. In particular:

$$m_\mathrm{W}^2 = \frac{g^2v^2}{4} \quad ; \quad m_\mathrm{Z}^2 = \frac{v^2(g^2+g'^2)}{4} \quad ; \quad m_\gamma = 0 \quad ; \quad v = \sqrt{\frac{1}{\sqrt{2}G_F}} \approx 246.22\,\mathrm{GeV} \tag{1.21}$$

where W and Z-boson masses are quadratically dependent on both the VEV and the weak coupling constants (g and g'), such that $m_\mathrm{Z}^2 = m_\mathrm{W}^2 + v^2g'^2/4 = m_\mathrm{W}^2/\cos^2\vartheta_\mathrm{W}$. Always from

Equation 1.20, cubic and quartic couplings of the Higgs field to pairs of massive vector bosons (W^+W^- or ZZ) are extracted proportional to $2h(x)/v$ and $h(x)^2/v^2$, respectively. The cubic coupling strength is proportional to the vector boson rest mass and requires a non null `VEV`, in fact:

$$g_{h\to WW} = g \cdot m_W \quad ; \quad g_{h\to ZZ} = \frac{1}{2}\frac{g \cdot m_Z}{\cos\vartheta_W} \quad \to \quad \frac{\Gamma(h\to WW)}{\Gamma(h\to ZZ)} = 4\cos^2\vartheta_W \frac{m_W^2}{m_Z^2} \approx 2.7 \quad (1.22)$$

Finally, since the Higgs field is coupled to W/Z bosons, the di-boson mass spectrum becomes sensitive to the `SM` Higgs presence and properties.

1.2.1 Fermion mass generation

As already described in Section 1.1.2, a Dirac mass term is not locally invariant under the electroweak gauge group, but it is possible to accommodate an invariant Dirac mass, for both leptons and quarks, trough the Higgs field. This is performed introducing a generic Yukawa term, which respects both Lorentz and $SU(2)_L \otimes U(1)_Y$ invariance:

$$\mathcal{L}_Y = +\bar\Psi_L^i \Gamma_\ell^{ij} \phi(\psi_R^\ell)^j + \bar\Psi_L^i \Gamma_\nu^{ij} \tilde\phi(\psi_R^\nu)^j + \bar Q_L^i \Gamma_d^{ij} \phi(\psi_R^d)^j + \bar Q_L^i \Gamma_u^{ij} \tilde\phi(\psi_R^u)^j + h.c. \quad (1.23)$$

where $\mathcal{L}_Y$ is reported as function of the interaction eigenstates and (Γ_l, Γ_ν, Γ_u, Γ_d) are generic 3×3 complex matrices in the three-dimensional generation space of quarks and leptons. Since neutrino mass is $m_\nu \approx 0$, the correspondent term is neglected:

$$\bar Q_L^i \Gamma_d^{ij} \phi(\psi_R^d)^j = \begin{bmatrix} (\bar\psi_L^u)^i & (\bar\psi_L^d)^i \end{bmatrix} \Gamma_d^{ij} \begin{bmatrix} 0 \\ \frac{v+h(x)}{\sqrt{2}} \end{bmatrix} (\psi_R^d)^j = \frac{v+h(x)}{\sqrt{2}} \cdot \left[(\bar\psi_L^d)^i \Gamma_d^{ij} (\psi_R^d)^j \right]$$

$$\bar Q_L^i \Gamma_u^{ij} \tilde\phi(\psi_R^u)^j = \begin{bmatrix} (\bar\psi_L^u)^i & (\bar\psi_L^d)^i \end{bmatrix} \Gamma_u^{ij} \begin{bmatrix} \frac{v+h(x)}{\sqrt{2}} \\ 0 \end{bmatrix} (\psi_R^u)^j = \frac{v+h(x)}{\sqrt{2}} \cdot \left[(\bar\psi_L^u)^i \Gamma_u^{ij} (\psi_R^u)^j \right]$$

Thanks to the previous equations, one finds:

$$\mathcal{L}_Y = \left[1 + \frac{h(x)}{v}\right] \left[(\bar\psi_L^d)^i M_d^{ij} (\psi_R^d)^j + (\bar\psi_L^u)^i M_u^{ij} (\psi_R^u)^j + (\bar\psi_L^\ell)^i M_\ell^{ij} (\psi_R^\ell)^j \right] + h.c.$$

Each complex squared matrix M can be diagonalized, obtaining real positive entries, through a bi-unitary transformation. This implies that it is always possible to find a pair of unitary transformations (V,U) which map M_ℓ, M_u or M_d into a diagonal form:

$$\mathcal{L}_Y = \left[1 + \frac{h(x)}{v}\right] \left[(\bar\psi_L^d)^k V_{d,L}^{ki} (V_{d,L}^{ih})^\dagger M_d^{hm} U_{d,R}^{mj} (U_{d,R}^{jn})^\dagger (\psi_R^d)^n .. \right] + h.c. \begin{cases} (\bar\psi_L^{d'})^i = (\bar\psi_L^d)^k (V_{d,L}^{ki}) \\ M_{d'}^{ij} = (V_{d,L}^{ih})^\dagger M_d^{hm} U_{d,R}^{mj} \\ (\psi_R^{d'})^j = (U_{d,R}^{jn})^\dagger (\psi_R^d)^n \end{cases}$$

$$\mathcal{L}_Y = \left[1 + \frac{h(x)}{v}\right] \left\{ \sum_i \left[m_{d'}^i (\bar\psi_L^{d'})^i (\psi_R^{d'})^i + m_{u'}^i (\bar\psi_L^{u'})^i (\psi_R^{u'})^i + m_\ell^i (\bar\psi_L^\ell)^i (\psi_R^\ell)^i + h.c. \right] \right\} \quad (1.24)$$

Eventually, the Yukawa term inside Equation 1.23 has been re-written in terms of mass eigenstates instead of weak interaction ones. In this way, both fermion masses and Yukawa couplings

between fermions and the Higgs field have been produced, which are proportional to the ratio between the fermion mass and the VEV.
The kinetic part of the electroweak Lagrangian, as well as the neutral current, remain unchanged moving from interaction to mass eigenstates, thanks to the unitary of $V_{\mathrm{R,L}}^{u,d}$ and $U_{\mathrm{R,L}}^{u,d}$ transformations. This guarantees, at least at three level, the suppression of flavor changing neutral currents, which are possible only at higher orders in the perturbation expansion. The only piece affected by this transformation is the weak charged current interaction, because up and down components, belonging to a SU(2) doublet, transform in different ways:

$$J_{\mathrm{CC}}^{\mu} = \bar{Q}_{\mathrm{L}}^{i}\gamma^{\mu}\tau^{+}Q_{\mathrm{L}}^{i} = (\bar{\psi}_{\mathrm{L}}^{u})^{i}\gamma^{\mu}(\psi_{\mathrm{L}}^{d})^{i} = (\bar{\psi}_{\mathrm{L}}^{u'})^{i}(V_{u,\mathrm{L}}^{ik})^{\dagger}\gamma^{\mu}U_{d,\mathrm{L}}^{kj}(\psi_{\mathrm{L}}^{d'})^{j} = (\bar{\psi}_{\mathrm{L}}^{u'})^{i}\gamma^{\mu}V_{\mathrm{CKM}}^{ij}(\psi_{\mathrm{L}}^{d'})^{j} \quad (1.25)$$

where the unitary matrix $V_{\mathrm{CKM}} = V_{u,\mathrm{L}}^{\dagger}U_{d,\mathrm{L}}$ is called Cabibbo-Kobayashi-Maskawa matrix. It is responsible for flavor changing in weak charge currents at three level experimentally discovered, for the first time, in weak decays without conservation of strangeness. The matrix elements have been measured by different experiments [21].

1.2.2 Theoretical constraints on the Higgs boson mass

Through the BEH mechanism, a new real massive scalar field $h(x)$ is introduced in the electroweak theory. Its mass value is not predicted by the SM, since the Higgs coupling constant λ is a free parameter of the model. However, there are different theoretical arguments that give an indirect constraint on the Higgs boson mass, in particular:

- **Perturbative unitary bound:** the unitary of the longitudinally polarized VBS amplitude, $\mathrm{V_L V_L} \to \mathrm{V_L V_L}$, gives a direct theoretical constraint on m_h [22]:

  $$m_h \lesssim 800 \text{ GeV}$$

  The relationship between the Higgs boson and the unitary bound of the VBS will be better described in Section 1.6.

- **Running of λ:** one loop renormalization calculations state that the Higgs self coupling constant λ must be a function of the renormalization scale μ of the analyzed process. This dependence is extracted by solving the evolution equation (RGE) [23] as follows:

  $$\frac{d\lambda(\mu)}{dlog(\mu^2)} = \frac{1}{16\pi^2}\left[12\lambda^2 + \frac{3}{8}g^4 + \frac{3}{16}(g^2+g'^2)^2 - 3h_t^2 - 3\lambda g^2 - \frac{3}{2}\lambda(g^2+g'^2) + 6\lambda h_t^2\right] \quad h_t = \frac{\sqrt{2}m_t}{v}$$

  This solution, together with the other ones for the fundamental SM parameters ($g(\mu)$, $g'(\mu)$, $g_s(\mu)$, $h_t(\mu)$)[3], brings to the results reported in Figure 1.2 (left), where $\lambda(\mu)$ is shown for different initial condition. If the initial condition $\lambda(\mu_0)$ at $\mu_0 = v$ or $\mu_0 = m_{\mathrm{Z}}$ is too small, $\lambda(\mu)$ becomes negative for large renormalization scale values. Since $\lambda(\mu)$ is a coupling constant, it has to be positive along the whole energy range in which the SM validity aims to be extended. This is equivalent to the requirement that the Higgs potential remains bounded from below, which allows to set a lower limit to m_h.
  Another important feature is the increase of the coupling constant $\lambda(\mu)$ with the scale μ as a function of the Higgs mass hypothesis. When λ becomes grater than one, the theory

[3] These are usually obtained with the $\overline{\mathrm{MS}}$ scheme

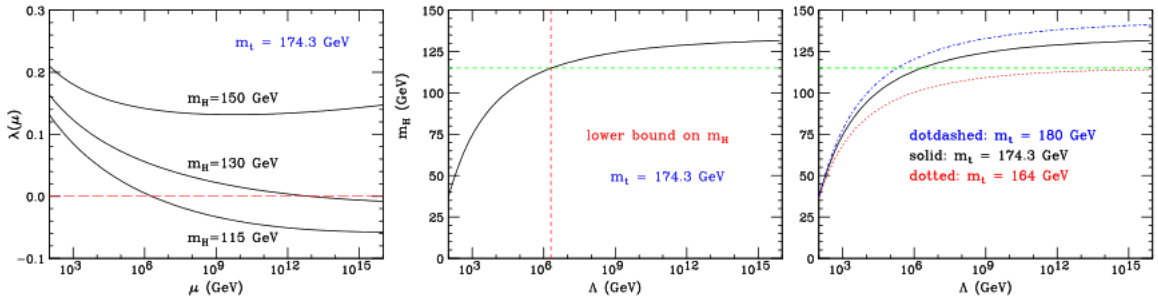

Figure 1.2: Left: the running of coupling constant $\lambda(\mu)$ for different initial values of $\lambda(m_Z)$. Middle: lower limit on m_h as function of the energy scale Λ. Right: lower limit on m_h as a function of m_t.

leaves the perturbative domain giving a Landau singularity. Considering a simplified model, without gauge and Yukawa couplings, the evolution of λ is given by:

$$\lambda(\mu) = \frac{\lambda(\mu_0)}{1 - \frac{3}{4\pi^2}\lambda(\mu_0)\log\frac{\mu^2}{\mu_0^2}} \quad ; \quad \lambda \to \infty \Leftrightarrow \mu^2 \approx \Lambda_L^2 = \mu_0^2 \exp\frac{4\pi^2}{3\lambda(\mu_0)} \tag{1.26}$$

where, for every initial condition $\lambda(\mu_0)$, the theory has an upper Λ_H and a lower scale Λ_L of validity, which correspond to a lower and a upper limit on m_h, respectively. Taking m_h to lie between [125,150] GeV, the theory is valid approximately up to the Planck scale $\approx 10^{19}$ GeV as shown in Figure 1.2 (middle and right). Therefore, the SM should be considered as an effective low energy theory, where the difference between the electroweak scale, $\approx 10^2$ GeV, and the Plank one, $\approx 10^{19}$ GeV, is usually called hierarchy problem.

1.3 Higgs boson measurements at the LHC

The SM Higgs boson production cross section at a proton-proton (pp) colliders is shown in Figure 1.3 as a function of the Higgs mass hypothesis (m_h). On the left, the total cross section is reported for different $\sqrt{s}$ values (7, 8 and 14 TeV), while, on the right, the contribution of the different leading production mechanisms is displayed [24]. In addition, in Figure 1.4, the corresponding Leading Order (LO) Feynman diagrams are drawn.

- **Gluon fusion** ($gg \to h$) is the dominating Higgs production mechanism over the entire mass range accessible at the LHC. It proceeds mainly with a heavy quark loop (top quark), as shown in Figure 1.4, since the Higgs coupling to fermions is proportional to m_f/v, as stated in Section 1.2.1. The inclusive cross section is known with NNLO+NNLL QCD and NLO EWK accuracy [24].

- **Vector Boson Fusion** ($qq' \to qq'h$) is about one order of magnitude weaker than the gluon fusion, where the Higgs boson in produced through a direct coupling with vector bosons (W or Z), which are irradiated by a pair of incoming quarks from the proton beams. The cross section is known at NNLO QCD + NLO EWK reducing the scale dependence at the level of 1-2% [24]. In addition, the fragmentation of the outgoing quarks produces two forward hard jets with high invariant mass, which are commonly used to tag the event.

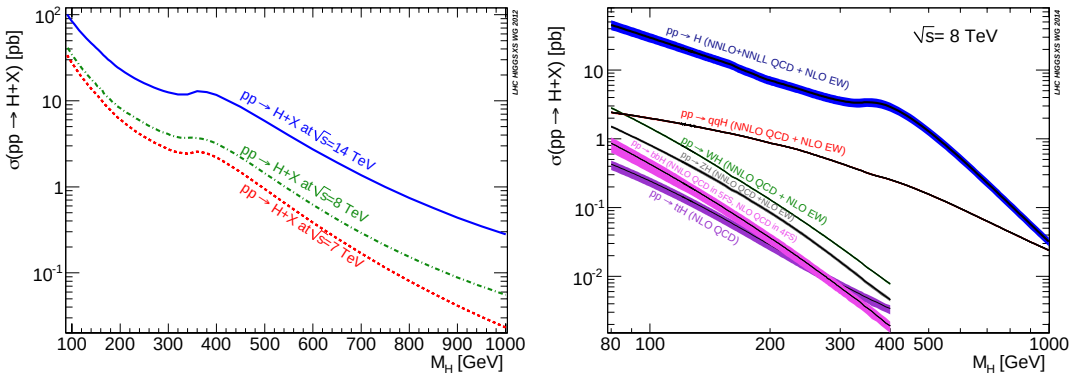

Figure 1.3: (Left) The SM Higgs total production cross section, as a function of its mass (m_h), for three different center of mass energy: $\sqrt{s} = 7$, 8 and 14 TeV. (Right) The SM Higgs production cross-sections at $\sqrt{s} = 8$ TeV for the different production mechanisms: gluon fusion (blue), vector boson fusion (red), associate production with a W or a Z-boson (green, grey), $t\bar{t}$ and $b\bar{b}$ associated production (violet and magenta respectively). When available, NNLO QCD and NLO EKW corrections are taken into account.

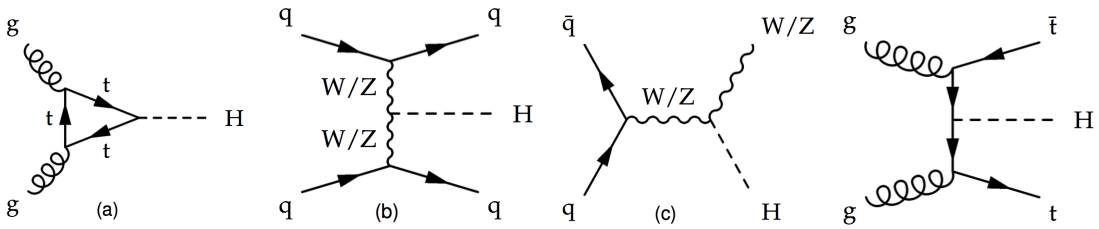

Figure 1.4: Leading order Feynman diagrams for the most important production processes of the SM Higgs boson: (a) gluon fusion, (b) vector boson fusion, (c) Higgs-strahlung and (d) $t\bar{t}$ associated production.

- **Associated production** with **W** or **Z-boson** ($q\bar{q}' \to \mathrm{Wh}$, $q\bar{q} \to \mathrm{WZ}$) have a smaller cross section than the previous mechanisms but, the presence of additional vector bosons helps in tagging the events reducing the contamination from other SM processes. This production mode is sensitive, as well as VBF, only to the Higgs coupling to vector bosons and the cross section is predicted with NNLO QCD + NLO EWK accuracy [24].

- **Associated production** with **$\mathbf{t\bar{t}}$ pairs** ($qq, gg \to t\bar{t}h$) has the smallest cross section, only known at NLO QCD [24]. This represents an important process since it is directly sensitive to the Higgs coupling properties to the top quark.

Depending on the Higgs boson mass hypothesis, different decay channels can be exploited to detect it. This is reported in Figure 1.5 (left), where the decay branching fractions, including NLO QCD and EWK corrections, are shown. New channels becomes available once the Higgs mass hypothesis exceeds a di-lepton/quark or di-boson threshold. Light-fermion decay modes contribute only in the low mass region, up to m_h around 150 GeV, while once the decay into pairs of vector bosons is accessible, this quickly dominates.
The Higgs boson does not couple to photons and gluons at Leading Order (LO), but such couplings can arise via fermion or vector boson loops, giving a sizable contribution in the low mass region. At the same time, the Higgs total width is shown in Figure 1.5 (right). It quickly increases with the mass due to the opening of new channels, becoming almost as large

as the Higgs mass itself around 1 TeV and making the definition of the Higgs boson itself as a particle matter to debate.

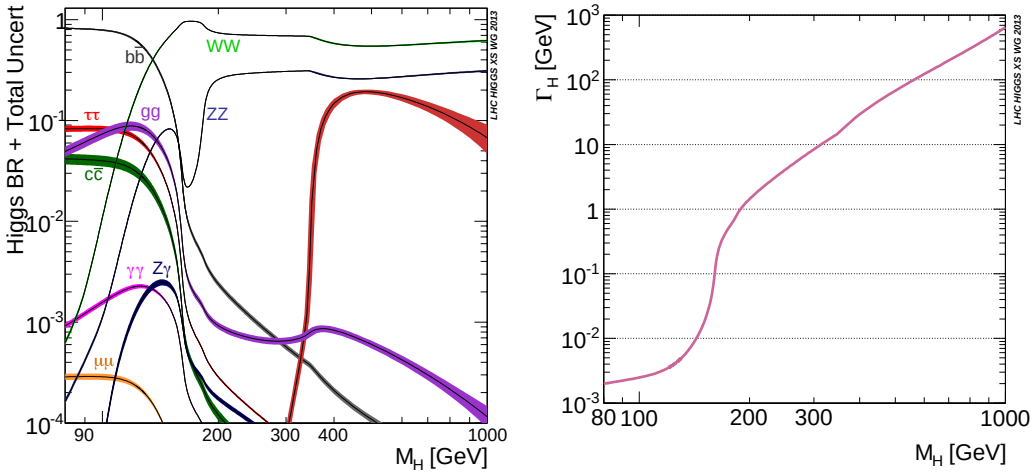

Figure 1.5: (Left) Decay branching ratios of the SM Higgs boson in the different channels as a function of the mass hypothesis. (Right) Total decay width of the SM Higgs boson as a function of its mass.

In the year 2012, the ATLAS and the CMS Collaborations announced the observation of a new boson with mass close to 125 GeV [25, 26]. Subsequent studies on production and decay rates [27, 28, 29, 30, 31] and spin-parity [32, 33] of the new boson showed that its properties are compatible with those expected for the SM Higgs boson. The CMS measurements targeting its decay into fermions ($b\bar{b}$ and $\tau\tau$) or vector bosons (ZZ^*, WW^*, $\gamma\gamma$) and different production modes ($gg \to h$, VBF, associated production with a vector boson and $qq, gg \to t\bar{t}h$) have been combined to extract the mass of the resonance and its couplings to SM particles [34] [4]. In Figure 1.6 (left), the 68% CL confidence regions, obtained from a two-dimensional likelihood scan, for the Higgs signal strength relative to the SM expectation ($\mu = \sigma/\sigma_{\rm SM}$) and its mass ($m_h$) are shown for $h \to ZZ \to 4\ell$ and $h \to \gamma\gamma$ searches, as well as for their combination. Only these two channels are considered for the mass measurement thanks to their optimal resolution, coming from the precise measurements of lepton momentum and photon energy. In addition, again in Figure 1.6 (right), a likelihood scan as a function of m_h is obtained introducing independent signal strengths for each specific production or decay mode, profiling them in the statistical interpretation. The Higgs boson mass is measured to be $m_h = 125.02^{+0.29}_{-0.31}$ GeV.
Fixing the mass value to the best-fit one, the signal strengths have been evaluated grouping different analysis categories and trying to separate each production or decay mode minimizing mutual contaminations. Results are shown in Figure 1.7 for both production (left) and decay modes (middle). In addition, since each Higgs boson production mechanism can be associated to either coupling to fermions (ggh, tth) or vector bosons (VBF, VH), a proper decorrelation of channels allows to extract relative signal strengths of the Higgs coupling to fermions and vector bosons. Figure 1.7 (right) shows the likelihood scan of the data for $\mu_{\rm VBF,Vh}/\mu_{\rm ggh,tth}$, where the best-fit corresponds to $1.25^{+0.62}_{-0.44}$ which is compatible with the SM expectation.

[4] Many studies have been developed to give a complete characterization of the Higgs boson properties but, in this Section, only few of them, which are connected with the topics later explored in this thesis, will be described.

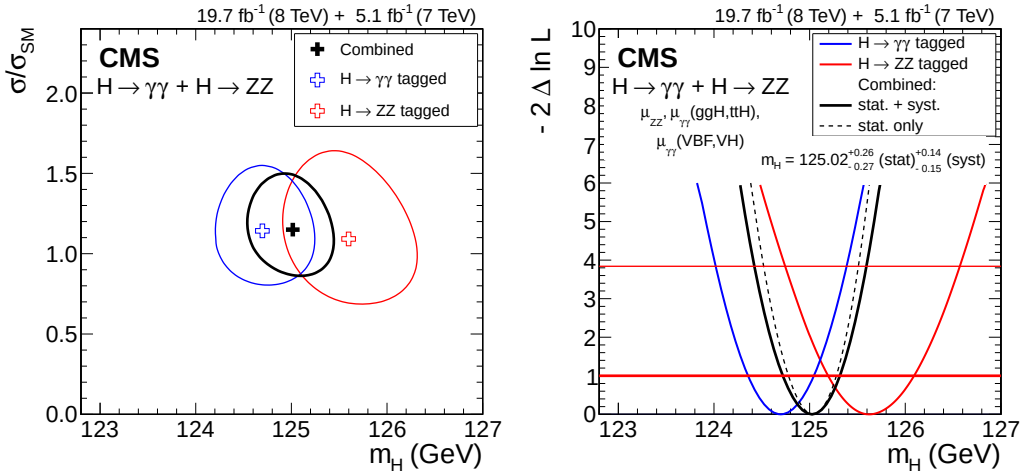

Figure 1.6: (Left) The 68% CL confidence regions for the signal strength $\sigma/\sigma_{\mathrm{SM}}$ versus the mass of the boson m_h for the $h \to \gamma\gamma$ and $h \to \mathrm{ZZ} \to 4\ell$ final states and their combination. (Right) Scan of the test statistics as a function of the Higgs mass m_h for to the same channels.

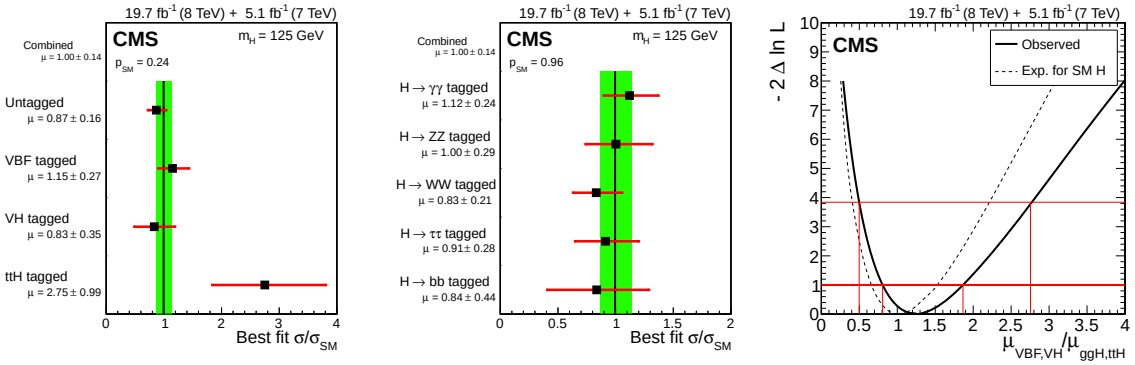

Figure 1.7: Values of the best-fit $\sigma/\sigma_{\mathrm{SM}}$ for the combined analysis (solid vertical line) and for separate combinations grouped by production mode (left) and predominant decay state (middle). The green band shows the total uncertainty on the fitted strength. The horizontal red bars indicate the ± 1 standard deviation in the best-fit values for the individual combinations. (Right) Likelihood scan as a function of the ratio $\mu_{\mathrm{VBF,Vh}}/\mu_{\mathrm{ggh,tth}}$ combined for all channels. The solid curve represents the observed result in data, while the dashed one indicates the expected median result in the presence of the SM Higgs boson.

Finally, the Higgs coupling to SM particles is investigated simultaneously in different production and decay processes, including the possibility of the SM Higgs to be coupled to BSM particles. To test possible deviations from the SM predictions, production rate scale factors $\kappa_i^2 = \sigma_i/\sigma_i^{\mathrm{SM}}$ and decay ones $\kappa_i^2 = \Gamma_{ii}/\Gamma_{ii}^{\mathrm{SM}}$ are fitted to data, where reference values are provided in [35]. The Higgs coupling to massive vector bosons is protected from large radiative corrections by the electroweak custodial symmetry [36]. Thus, testing the compatibility of these couplings with the SM expectation is an indirect check of the presence of new physics in the EWSB sector. In Figure 1.8 (left), a two-dimensional scan in the $(\kappa_{\mathrm{V}}, \kappa_f)$ parameter space is shown, where data are compatible within 68% CL with the SM prediction $(\kappa_{\mathrm{V}}, \kappa_f) = (1,1)$. Moreover, Figure 1.8 (right) reports the coupling scale factors to single particles, where the small uncertainty on κ_t is achieved since the Higgs boson is mainly coupled to top quarks in the ggh production.

Overall, couplings to vector bosons, fermions, gluons and photons[5] are compatible with the SM prediction within 95% CL. The current precision in the Higgs couplings to SM particles is around 15% for vector bosons (W,Z,γ), 20% for gluons and 30% for fermions.

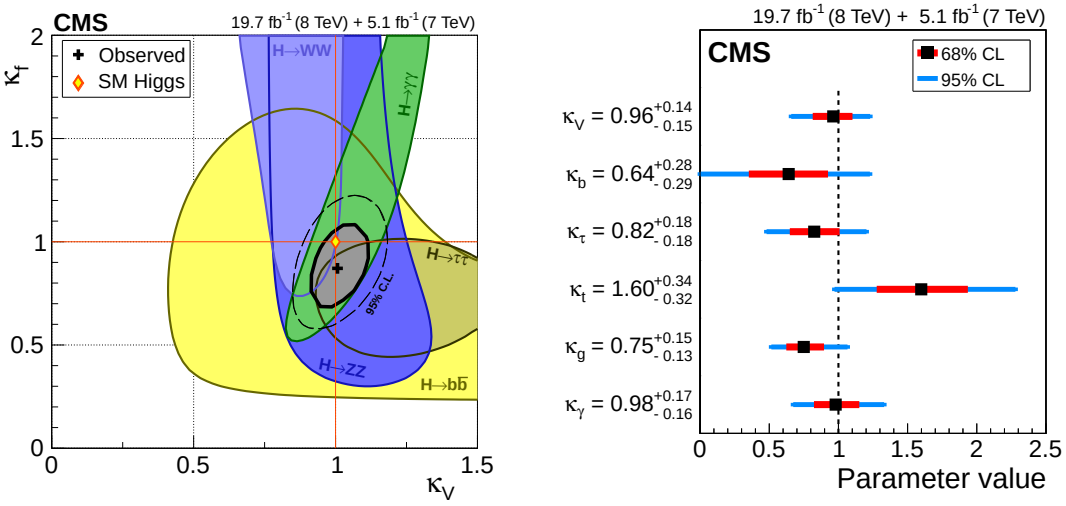

Figure 1.8: (Left) The 68% CL confidence regions for individual channels and for their combination on κ_V and κ_f parameters. The cross indicates the global best-fit value, while the dashed contour bounds the 95% CL confidence region for the combination. The diamond represents the SM expectation, $(\kappa_V,\kappa_f) = (1,1)$. (Right) Likelihood scans for the coupling scaling factors, one coupling at a time while profiling the remaining ones: κ_V (W and Z boson), κ_b (bottom quark), κ_τ (τ-lepton), κ_t (top quark), κ_g (gluon, "effective coupling") and κ_γ (photon, "effective coupling").

1.4 BSM extensions of the Higgs sector

The properties of the observed Higgs boson, briefly summarized in Section 1.3, are compatible with the SM expectation, within the present experimental and theoretical uncertainties, and consistent with the unitarity constraint of the vector boson scattering amplitude. Because of the uncertainties, it is still possible that the discovered resonance represents only one piece of a more extended Higgs sector, only partially responsible for the EWSB. BSM extensions can be realized in several scenarios, such as the electroweak singlet [37, 38, 39, 40, 41] or the two Higgs doublet model (2HDM) [42, 43].
In this Section, the simplest case in which an additional Higgs scalar field ϕ_H is added to the SM Lagrangian, transforming under a hidden $U(1)_{hid}$ gauge symmetry broken by a non null VEV, is investigated. This implies that ϕ_H must be neutral to all the quantum numbers related to the SM gauge groups. Only renormalizable couplings with original Higgs field are allowed through a term $\phi_H^2\phi_h^2$, where h stands for the interaction eigenstate of the observed boson with mass of about 125 GeV. Under these constraints, the Higgs Lagrangian becomes:

$$\mathcal{L}_{\text{Higgs}} = |D_\mu\phi_h|^2 + |D_\mu\phi_H|^2 + \mu_{\phi_h}^2|\phi_h|^2 + \mu_{\phi_H}^2|\phi_H|^2 - \lambda|\phi_h|^4 - \rho|\phi_H|^4 - \eta|\phi_h|^2|\phi_H|^2 \quad (1.27)$$

[5] Gluons and photons don't couple directly with the Higgs boson. An effective coupling is possible only with heavy quark or vector boson loops, as in shown for the gluon fusion production in Figure 1.4

where the Higgs doublet ϕ_h and the singlet ϕ_{H} are written as:

$$\phi_h = \begin{bmatrix} \vartheta_1(x) + i\vartheta_2(x) \\ \frac{1}{\sqrt{2}}(v + h(x) + i\vartheta_3(x)) \end{bmatrix} \qquad \phi_{\mathrm{H}} = \frac{1}{\sqrt{2}}[H(x) + \xi + i\vartheta_4(x)] \tag{1.28}$$

As already described in Section 1.2, Goldstone degrees of freedom can be transferred to vector boson masses with the unitary gauge choice, while v and ξ are the respective VEV used for breaking the related symmetries:

$$v^2 = \frac{1}{\lambda}(-\mu^2_{\phi_h} - \frac{1}{2}\eta^2\xi^2) \qquad \xi = \frac{1}{\rho}(-\mu^2_{\phi_{\mathrm{H}}} - \frac{1}{2}\eta^2 v^2) \tag{1.29}$$

Moreover, combining Equation 1.27 and Equation 1.28, the following mass matrix is obtained:

$$\mathrm{M} = \begin{bmatrix} 2\lambda v^2 & \eta v\xi \\ \eta v\xi & 2\rho\xi^2 \end{bmatrix} \tag{1.30}$$

where by means of a diagonalization, one gets the following mass eigenstates:

$$\begin{bmatrix} h'(x) \\ H'(x) \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \cdot \begin{bmatrix} h(x) \\ H(x) \end{bmatrix} \tag{1.31}$$

while the mass eigenvalues are given by:

$$\mathrm{M}^2_{h',H'} = (\lambda v^2 + \rho\xi^2) \pm \sqrt{(\lambda v^2 - \rho\xi^2)^2 + \eta^2 v^2 \xi} \tag{1.32}$$

In the most generic case, h' and H' couple to both the SM Higgs field $h(x)$ and the hidden sector $H(x)$. We can reasonably assume from direct Higgs boson measurements that the two potential parameters (λ,ρ) and the VEVs (v,ξ) are of similar size, while the off-diagonal parameter η is moderate. In this way, the properties of h' remain dominated by the SM component, while H' is characterized primarily by the hidden Higgs field. The phenomenology of this model depends on whether the SM Higgs particle is lighter or heavier than the new companion. The most interesting scenario is obtained placing the new scalar H' in the high mass region, $m_{\mathrm{H}} \in [130, 1000]$ GeV, assuming that h' matches the novel Higgs boson observed at 125 GeV, decaying only into SM particles at a rate reduced by the mixing parameter. Possible decays of the heavier boson H' into the hidden sector are suppressed.
In this framework, h' couplings to SM particles are universally suppressed by the mixing parameter $\cos\alpha$, which reduces the production cross section σ and total width Γ^{tot} with respect to the SM predictions (σ^{SM}, Γ^{SM}), according to:

$$\sigma = \cos^2\alpha \cdot \sigma^{\mathrm{SM}} \quad \Gamma^{\mathrm{SM}} = \cos^2\alpha \cdot (\Gamma^{\mathrm{SM}}_{\mathrm{vis}} + \Gamma^{\mathrm{SM}}_{\mathrm{inv}}) \quad \Gamma^{\mathrm{tot}} = \Gamma^{\mathrm{SM}} + \Gamma^{\mathrm{new}} \tag{1.33}$$

where Γ^{new} and $\cos^2\alpha$ are already constrained by SM Higgs measurements, as described in Section 1.3, while Γ^{inv} refers to the Higgs invisible width due to $h \to \mathrm{ZZ} \to \nu\nu$ decays. It is common to express Equation 1.33 in terms of the coupling scale factor (κ) and the branching fraction of new physics ($\mathrm{BR}_{\mathrm{new}}$):

$$\kappa = \frac{\Gamma_i\Gamma_j}{\Gamma^{tot}} \cdot \frac{\Gamma^{\mathrm{SM}}}{\Gamma^{\mathrm{SM}}_i\Gamma^{\mathrm{SM}}_j} = \frac{\cos^2\alpha}{1 + \Gamma^{\mathrm{new}}/(\Gamma^{\mathrm{SM}}\cos^2\alpha)} = \frac{\cos^2\alpha}{1 + B_{\mathrm{new}}} \quad B_{\mathrm{new}} = \frac{\mathrm{BR}_{\mathrm{new}}}{\mathrm{BR}_{\mathrm{SM}}} = \frac{\mathrm{BR}_{\mathrm{new}}}{1 - \mathrm{BR}_{\mathrm{new}}}$$

where i and j indicate the initial and final state of the considered process. Since κ is measured by SM Higgs direct searches, upper bounds on the mixing angle α and the hidden branching fraction are set:

$$\sin^2\alpha \leq 1-\kappa^2 \qquad \mathrm{BR}_{\mathrm{new}} \leq 1-\kappa \quad \rightarrow \quad \frac{\Gamma^{\mathrm{new}}}{\Gamma^{\mathrm{SM}}} = \cos^2\alpha\left(\frac{\cos^2\alpha}{\kappa}-1\right) \tag{1.34}$$

which leads to:

$$\cos^2\alpha = \frac{\kappa}{1-\mathrm{BR}_{\mathrm{new}}} \qquad \frac{\Gamma^{\mathrm{hid}}}{\Gamma^{\mathrm{SM}}} = \frac{\kappa\cdot\mathrm{BR}_{\mathrm{new}}}{(1-\mathrm{BR}_{\mathrm{new}})^2} \tag{1.35}$$

Looking at the mixing parameter, there are two interesting cases:

- $|\cos\alpha| = 1$ and $\mathrm{BR}_{\mathrm{new}} = 0$, where the heavier Higgs is completely decoupled from the SM one, i.e. the SM couplings and width are not modified by the high mass sector. This scenario is equivalent to search for a SM-like Higgs boson at high mass.

- $|\cos\alpha| \in (0,1)$, where the low and the high mass bosons are coupled, thus the SM branching fractions and couplings are modified according to the properties of the new state. The value of $\mathrm{BR}_{\mathrm{new}}$ depends only on the amplitude of the $H \rightarrow hh$ decay since no decays into hidden particles are considered.

From direct measurements, an indirect upper limit at 95% CL is measured to be $\cos^2\alpha \leq 0.28$, which is obtained using the signal strength fits to the $h(125)$ boson, as reported in [34]. On the other hand, experimental results from direct searches of the extended Higgs sector will be widely described in Chapter 5.

1.5 Gravitons in warped extra dimensions models

Additional heavy particles, with respect to those expected within the SM, are also predicted by unified models describing the gravitational interaction inside the framework of quantum field theory. The SM of elementary particles describes precisely four fundamental interactions (electromagnetic, weak, strong and Higgs field), while the gravitational one remains out of this picture. The mass scale that weakens the gravitational force is known as Planck scale, $\overline{m}_{pl} = m_{pl}/\sqrt{8\pi} = 2.4\times10^{18}$ GeV, while the one for the weak interactions is around 100 GeV. In the Warped Extra Dimension (WED) theory [44, 45], the effective difference between the Plank and the weak scale is explained by the existence of an additional finite spatial dimension in our world that, in this picture, would have a non plain five-dimensional metric. This five-dimensional space is referred to as the Bulk. Quantum fluctuations of the metric are interpreted as particles, where fluctuations around its infinite four-dimensional part correspond to the Graviton field (spin-2), while the ones around the finite fifth dimension give rise to a spin-0 field called Radion. Therefore, the finitness of the extra dimension introduces in the four-dimensional effective theory excitation modes, which appear as heavy resonances, usually called Kaluza-Klein (KK) modes. The mass of the lightest KK-mode of the gravitational field is expected to be in the TeV range, depending on the size of the ratio between the extra dimension curvature and the Plank scale $(k/\overline{m}_{pl})$.
The metric commonly chosen in WED models brings to a Universe scenario in which, once the compactification scheme is applied, the extra dimension is described as a finite line connecting

two four-dimensional infinite branes. Starting from the most generic Lorentz invariant solution of the Einstein equation in the four-dimensional space:

$$ds^2 = e^{-2\sigma(\phi)}\eta_{\mu\nu}dx^\mu dx^\nu + r_c^2 d\phi^2 = g_{\mathrm{MN}}dx^{\mathrm{M}}dx^{\mathrm{N}} \tag{1.36}$$

the following action can be written:

$$\mathcal{S} = \mathcal{S}_{\mathrm{Gravity}} + \mathcal{S}_{\mathrm{TeV}} + \mathcal{S}_{\mathrm{Plank}} + \mathcal{S}_{\mathrm{Matter}} \tag{1.37}$$

where $\mathcal{S}_{\mathrm{Gravity}}$ is the gravitational action, $\mathcal{S}_{\mathrm{TeV}}$ and $\mathcal{S}_{\mathrm{Plank}}$ are the gravitational actions confined in the two four-dimensional branes, ϕ is the fifth dimension and g_{MN} is the five-dimensional tensor metric. The gravitational actions are written as:

$$\mathcal{S}_{\mathrm{Gravity}} = \int d^4x \int_{-\pi}^{+\pi} d\phi \sqrt{g}(\Lambda_{\mathrm{Bulk}} + 2\mathrm{M}_5^3 R) \tag{1.38a}$$

$$\mathcal{S}_{i=\mathrm{TeV/Plank}} = -\int d^4x \sqrt{g(\phi = 0, \pi)}\Lambda_i \tag{1.38b}$$

From the Einstein equation, taking $\Lambda_{\mathrm{bulk}} = \Lambda_{\mathrm{Plank}} = -\Lambda_{\mathrm{TeV}} = \Lambda$, the exponential warped factor $\sigma(\phi)$ is expressed as:

$$\sigma(\phi) = r_c|\phi|\sqrt{\frac{-\Lambda}{24\mathrm{M}_5^2}} = r_c|\phi|k \tag{1.39}$$

Finally, integrating out the extra dimension, the Plank mass is defined as follows:

$$\overline{m}_{pl}^2 = \frac{\mathrm{M}_5^3}{k}(1 - e^{-2\pi k r_c}) \tag{1.40}$$

where k and r_c, respectively known as curvature parameter and compactification radius, are free parameters in the theory ruling the hierarchy between the weak and the gravitational scales. When the Higgs sector is added to the `WED` context, the Higgs doublet is confined to live in the TeV brane, which corresponds to the `SM` world, and the four dimensional `VEV` of the Higgs field becomes:

$$v = e^{-\pi k r_c} v_0 \tag{1.41}$$

where v_0 is the five-dimensional `VEV`, usually assumed to be of the order of the Plank scale. Therefore, a reasonable separation between weak and gravitational scales is foreseen when $kr_c \sim 11$. Moreover, gravity particles are generated by quantum fluctuation around the classical solution for the metric, usually decomposed in the following way:

$$\delta g_{\mathrm{MN}}(x, \phi) = \begin{bmatrix} h_{\mu\nu|}(x, \phi) & h_{\mu,5}(x, \phi) \\ h_{5\nu|}(x, \phi) & h_{5,5}(x, \phi) \end{bmatrix} \tag{1.42}$$

where selecting $h_{\mu,5} = h_{5,\nu} = 0$, also known as axial gauge choice, tensor and scalar perturbations are decoupled. Tensor fluctuations obtained from the Fourier expansion of the action correspond to Graviton modes, while the scalar ones to the Radion field. Eventually, one obtains the following equation:

$$kr_c = \frac{1}{\pi}\ln\left(x_1\tilde{k}\frac{\overline{m}_{pl}}{m_{\mathrm{G}}}\right) \tag{1.43}$$

where kr_c should range between (10,12) to reproduce a phenomenologically acceptable condition, which is allowed when $\tilde{k} = k/m_{pl}$ lies within (0.01,1) and m_{G} between (100 GeV,2 TeV). In this context of `KK`-Gravitons originated from `WED`, there are two main sub-cases:

- **Bulk scenario** in which matter fields can propagate along the extra dimension.
- **RS1 scenario** in which matter fields are confined in the SM brane, also know as TeV brane.

1.5.1 Graviton production and decay at the LHC

In analogy with the Higgs boson production at hadron colliders, described in Section 1.3, KK-Gravitons are mainly produced via gluon fusion, while sub-leading contributions comes from VBF and associated production of an heavy graviton with a Z-boson.
In Figure 1.9, the predicted cross section for these production mechanisms is shown as a function of the Graviton mass, for both RS1 (left) and Bulk (right) models, fixing the ratio k/m_{pl} equal to 0.2 [46]. The ranking among the different production modes is similar to the one observed in the SM Higgs sector, where the ratio between gluon fusion and VBF cross sections ($\sigma_{ggh}/\sigma_{qqh}$) is model dependent. In fact, a factor ∼10 is expected for the Bulk Graviton production, while a factor ∼100 for the RS1 model.
Given the different phenomenology predicted by Bulk and RS1 scenarios, fixing the same $\tilde{k}$ value, the predicted cross section for two models are sizably different over the whole mass range, around 0.1 (100) *pb* at 1 TeV for Bulk (RS1) gravitons.

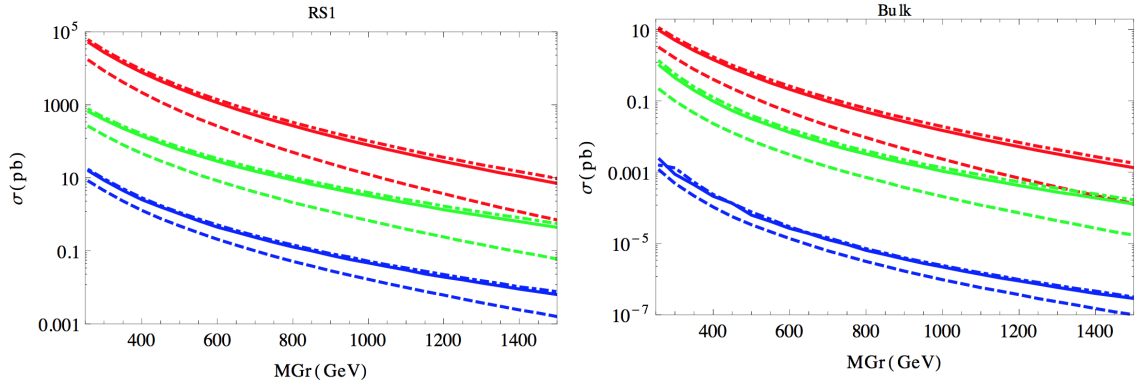

Figure 1.9: KK-Graviton production cross section with $k/m_{pl} = 0.2$, (left) RS1 scenario, (right) Bulk one. The red curves correspond to the gluon fusion production, the green ones to VBF and the blue ones to the associated production with a vector boson. The dot-dashed, continuous and dashed lines refer to $\sqrt{s} = 14$, 13 and 8 TeV, respectively.

Graviton branching fractions to SM particles, for masses above the $t\bar{t}$ pole, are substantially different between Bulk and RS1 scenarios. Bulk gravitons mainly decay into massive objects ($t\bar{t}$ ∼80%, WW ∼15%, ZZ ∼8% and hh ∼7%), while RS1 ones are mainly decaying into gluons (∼50%), WW (∼15%) and ZZ (∼8%). This implies that di-boson signatures (WW,ZZ) are the best sensitive topologies for discovering or excluding particles predicted by WED models.
Furthermore, fixing a common curvature parameter, the total width of RS1 gravitons is two order of magnitude larger than what is predicted in the Bulk scenario, which is less than 5 GeV up to m_G of 2 TeV. Therefore, for KK-Gravitons predicted in the Bulk model, the total width is usually small or negligible with respect to the experimental resolution (σ_m/m_G), which is typically around 10% for signatures involving jets and genuine missing transverse energy.
Finally, another relevant difference between RS1 and Bulk gravitons lies on the polarization state of the vector bosons produced by their decay. In fact, RS1 gravitons produce almost

pure transverse modes, while Bulk only longitudinal ones. A detailed description of dedicated searches for KK-Gravitons, decaying into di-boson final states, with the CMS experiment at the LHC, are reported in Chapter 4.

1.6 Vector boson scattering

Vector boson scattering and quartic gauge couplings are fundamental aspects of the SM, which remain mainly unexplored after the first years of LHC collisions. Evidence for the electroweak WW production has been reported by both ATLAS and CMS experiments, looking at 8 TeV data [47, 48]. The observation of the Higgs boson provides a natural candidate involved in the EWSB mechanism, strongly constraining BSM models like 2HDM, MSSM, NMSSM, composite Higgs, etc. Anyway, vector boson scattering still represents a model independent benchmark to precisely test of the EWSB sector, without assuming any specific model of new physics that could introduce deviations in the VBS spectrum with respect to the SM expectations.
With VBS one refers to purely electroweak processes in which, at LO in the perturbation theory, starting from two quarks in the initial state, a six parton final state is produced involving only electroweak vertexes (α_{ew}^6). Depending on the vector boson content of the final state (ZZ, $W^{\pm}W^{\mp}$, $W^{\pm}W^{\pm}$ or WZ), several diagrams, characterized by the same initial and final state, are involved:

- **Vector boson scattering diagrams** are divided into three independent logical steps: emission of a vector boson pair (W or Z) from the initial state partons, scattering of the emitted bosons and finally their decay into the final state particles. Examples are reported in Figure 1.10, which shows how quartic and triple gauge couplings are involved in these scattering interactions.

- **Non scattering diagrams** strongly interfere with the scattering ones, making not possible to disentangle scattering from non scattering processes. Examples are shown in Figure 1.11.

- **Electroweak $t\bar{t}$** and **single-top** production for which examples are shown in Figure 1.12.

- **Three vector boson production** for which examples are reported, as well, in Figure 1.12.

In conclusion, several contributions must be considered when looking for a complete LO estimate of α_{ew}^6 processes. The production times decay approximation is not suitable to obtain a correct description and the interference with diagrams of the order $O(\alpha_{ew}^4\alpha_s^2)$ leads, in some cases, to a sizable effect.

1.6.1 Longitudinal electroweak scattering

In absence of the SM Higgs boson, the amplitudes of the longitudinal electroweak scattering ($V_LV_L \to V_LV_L$) would increase as a function of $\sqrt{\hat{s}}$ as $\hat{s}/m_V^2$, violating unitarity [49].
The Higgs boson observed by the LHC experiments may restore the unitarity, although some scenarios of BSM physics predict enhancements in the cross sections through modifications of the Higgs sector or the presence of additional resonances [50, 51]. One possibility is represented

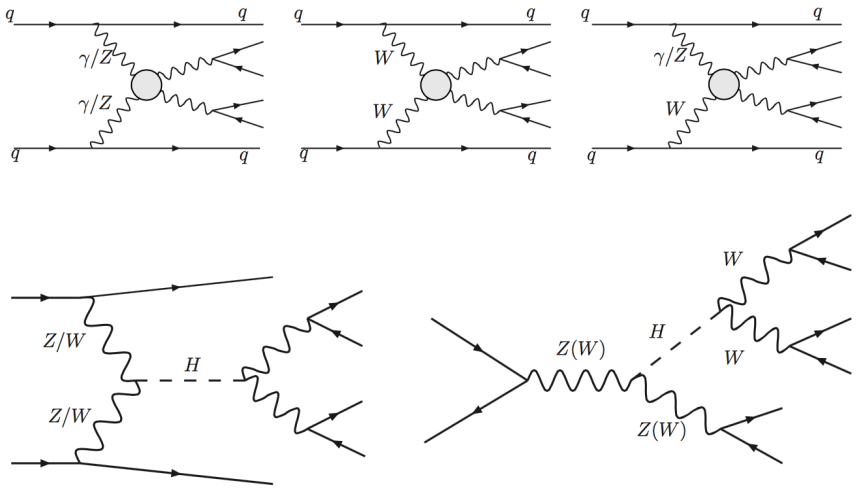

Figure 1.10: Examples of scattering diagrams contributing to α_{ew}^6 `VBS` process: (top) generic graphs, (bottom) `VBF` Higgs production and Higgs strahlung.

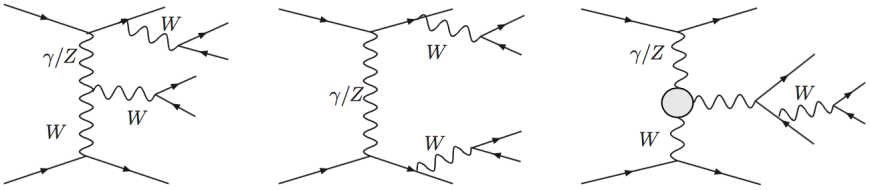

Figure 1.11: Examples of non-scattering and non-doubly vector boson production diagrams contributing to α_{ew}^6 `LO VBS` process.

by the electroweak singlet extension described previously in Section 1.4.
The longitudinal polarization 4-vector associated with a vector boson (V) is defined to be:

$$\epsilon_L^\mu(p) = \frac{p^\mu}{m_\mathrm{V}} + v^\mu(p) = \frac{p^\mu}{m_\mathrm{V}} - \frac{2m_\mathrm{V}}{2(p^0)^2}(p^0, -\vec{p}) \tag{1.44}$$

Considering as reference frame the center of mass of the incoming di-boson pair, where $\vec{p}_1 = -\vec{p}_2$, the previous equation becomes:

$$\epsilon_L^\mu(p_1) = \frac{p_1^\mu}{m_\mathrm{V}} - \frac{2m_\mathrm{V}}{s}p_2^\mu \quad \epsilon_L^\mu(p_2) = \frac{p_2^\mu}{m_\mathrm{V}} - \frac{2m_\mathrm{V}}{s}p_1^\mu \tag{1.45}$$

A similar relation holds also for the momenta of the outcoming vector bosons (k_1,k_2). Taking as example $\mathrm{W}^+(p_1)\mathrm{W}^-(p_2) \to \mathrm{W}^+(k_1)\mathrm{W}^-(k_2)$, the scattering Feynman diagrams with a quartic gauge coupling (4-point vertex), the s-channel and the t-channel exchange of a γ or a Z-boson or a Higgs boson are characterized by the following amplitudes:

$$\mathrm{M}_t^{\gamma+Z} = -\frac{g^2}{4m_\mathrm{V}^4}\left[(s-u)t - 3m_\mathrm{V}^2(s-u) + \frac{8m_\mathrm{V}^2}{s}u^2\right]$$

$$\mathrm{M}_s^{\gamma+Z} = -\frac{g^2}{4m_\mathrm{V}^4}\left[s(t-u) - 3m_\mathrm{V}^2(t-u)\right]$$

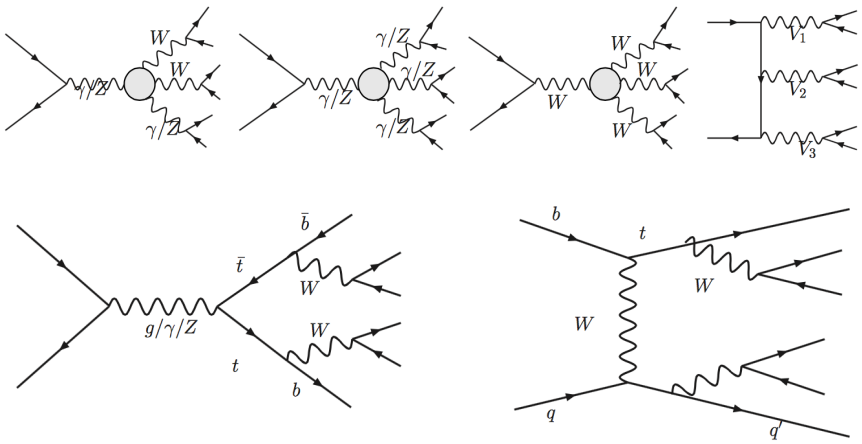

Figure 1.12: Examples of other electroweak processes contributing to α_{ew}^6 LO VBS: (top) three vector boson production graphs, (bottom) $t\bar{t}$ and single-top electroweak productions.

$$\mathrm{M}_4 = \frac{g^2}{4m_\mathrm{V}^4}\left[s^2 + 4st + t^2 - 4m_\mathrm{V}^2(s+t) - \frac{8m_\mathrm{V}^2}{s}ut\right]$$

$$\mathrm{M}^h = \frac{k_\mathrm{V}^2 g^2}{4m_\mathrm{V}^2}u$$

Summing up these different terms:

$$\mathrm{M}^{tot} = \left[\mathrm{M}_t^{\gamma+Z} + \mathrm{M}_s^{\gamma+Z} + \mathrm{M}_4\right] + \mathrm{M}^h = -\frac{g^2}{4m_\mathrm{V}^2}u + \frac{k_\mathrm{V}^2 g^2}{4m_\mathrm{V}^2}u \tag{1.46}$$

where, if the couplings to the Higgs boson are purely SM-like, the longitudinal scattering amplitude will be completely unitarized. On the other hand, if the $h \to VV$ coupling deviates from the SM prediction, the divergent term (s/m_V^2) becomes dominant after the Higgs pole, bringing to a distortion of the di-boson spectrum. It has been shown that the unitarity violation occurs at $\sqrt{s_{\mathrm{WW}}} = 1.7,\ 2.7,\ 3.8$ TeV for $k_\mathrm{V}^2 = 0.7,\ 0.9,\ 0.95$ respectively [52]. In Figure 1.13, the longitudinal scattering cross section for $\mathrm{W^+W^-} \to \mathrm{W^+W^-}$ and $\mathrm{W^+W^-} \to \mathrm{ZZ}$ is shown as a function of $\sqrt{s_{\mathrm{WW}}}$ for various k_V^2 values.
Due to the small electroweak production cross section $\mathrm{O}(\alpha_{ew}^6)$, the contamination from both di-boson pairs produced via diagrams of the order $\mathrm{O}(\alpha_{ew}^4\alpha_s^2)$ and from other SM processes with the same signature, the investigation of EWSB trough VBS will be one of the most challenging analyses of the future high luminosity LHC (HL-LHC) program [53].

1.6.2 Anomalous quartic gauge couplings

The presence of new resonances in the high mass regime are predicted by different BSM models. Even if their direct search would not be kinematically allowed at the LHC, a universal way to indirectly test them consists of precisely measure the vector boson self coupling interactions. Triple gauge interactions have been already studied at LEP [54, 55], Tevatron [56, 57], as well as with the collisions recorded by the LHC during Run-I [58, 59, 60], while first attempts of quartic gauge coupling measurements have just started.
These studies open a new opportunity in the search for BSM phenomenology, since new physics

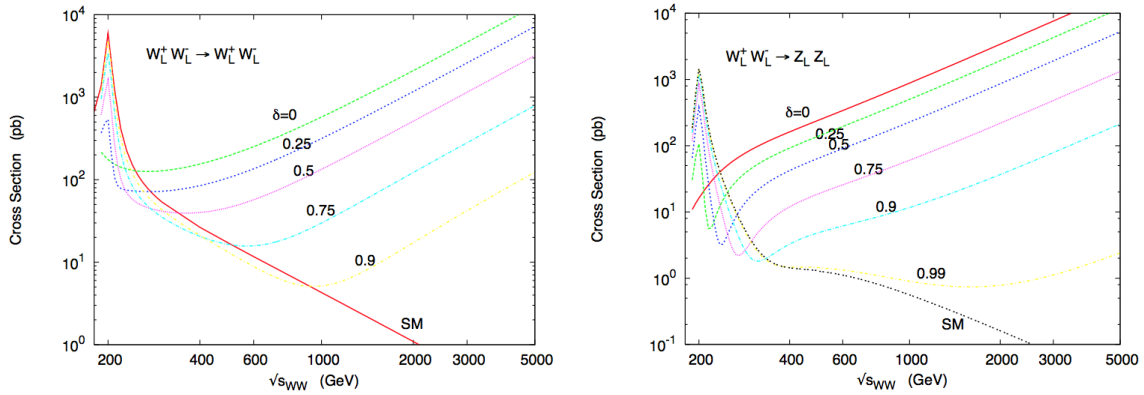

Figure 1.13: Scattering cross section for (left) $W_L^{\pm}W_L^{\mp} \to W_L^{\pm}W_L^{\mp}$ and (right) $W_L^{\pm}W_L^{\mp} \to Z_L Z_L$ as a function of $\sqrt{s_{WW}}$. Different values of δk_V^2, with respect to unity, are shown and a Higgs boson mass of 200 GeV is assumed.

effects can be parametrized through anomalies in the vector boson quartic couplings (aQGC), predicting an enhanced cross section with respect to the SM expectation [61]. This parametrization is performed in an effective field theory approximation, where only Lorentz invariant operators without derivative of gauge fields are considered, giving the following Lagrangian:

$$\mathcal{L}_{QGC}^{VVV'V'} = c_0^{VV'} T_0^{VV'} + c_1^{VV'} T_1^{VV'} \tag{1.47}$$

where all the possible operators are given by:

$$T_0^{WW} = g^{\alpha\beta} g^{\mu\nu} [W_\alpha^+ W_\beta^- W_\mu^+ W_\nu^-] \quad T_1^{WW} = g^{\alpha\beta} g^{\mu\nu} [W_\alpha^+ W_\beta^+ W_\mu^- W_\nu^-]$$

$$T_0^{WZ} = g^{\alpha\beta} g^{\mu\nu} [W_\alpha^+ Z_\beta W_\mu^- Z_\nu] \quad T_1^{WZ} = g^{\alpha\beta} g^{\mu\nu} [W_\alpha^+ W_\beta^- Z_\mu Z_\nu]$$

$$T_0^{ZZ} = T_2(ZZ) = g^{\alpha\beta} g^{\mu\nu} [Z_\alpha Z_\beta Z_\mu Z_\nu]$$

If the SM is thought as an effective low energy theory, which holds below a certain energy scale Λ, the coefficient c_0 and c_1 would become:

$$c_i^{VV'} = c_{i,SM}^{VV'} + g^2 \Delta c_i^{VV'} \tag{1.48}$$

where deviation from the SM case, $g^2 \Delta c_i^{VV'}$, are generated by higher order operators embedding new physics effects usually expected to be around the TeV scale. These additional terms should contain only the Higgs doublet (ϕ_h), $D_\mu \phi_h$ and the field strengths $W_{\mu\nu}^i$ and $B_{\mu\nu}^i$.
Dimension-six operators are better constrained by looking at anomalies in the triple gauge couplings, thus the lowest anomalous operators, considered to study aQGC models, are of dimension eight[6]. First bounds on the structure of quartic vector boson interactions, in the framework of dimension-eight effective field theory operators, have been set using 8 TeV data [48]. Prospects on aQGC limits obtained studying the VBS topology, in the context of HL-LHC program, will be described in Chapter 6.

[6] A complete picture of all the possible eight-dimensional operators considered in aQGC models can be found in the Appendix A of [61].

Chapter 2

The CMS detector at the LHC

Despite its success in describing the majority of the measurements performed at particle colliders so far, the SM leaves some unexplained phenomena, like the neutrino non-zero mass, the matter abundance with respect to anti-matter in the Universe and the dark matter indirect detection. Hadron collisions are one of the most effective ways to search for direct evidence of BSM physics, thanks to their large centre-of-mass energy and instantaneous luminosity. In this Chapter, starting from Section 2.1, the LHC project goals and the Cern accelerator system are illustrated. Then, in Section 2.2, a brief phenomenological summary of hadron collider physics is reported while, in Section 2.3, a general description of the CMS experiment and its sub-detectors is given. Finally, in Section 2.4, a description of offline physics object reconstruction algorithms is presented.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a two-ring superconducting proton-proton (pp) collider installed at the CERN laboratory, 100 meters underground, in the old 27 km long tunnel of LEP machine [62]. It is the most advanced High Energy Physics (HEP) project, both for the reachable center of mass energy ($\sqrt{s} = 14$ TeV) and for the design instantaneous luminosity ($\mathscr{L} = 1.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$). The large beam energy of 7 TeV, together with the radius of the machine, imposed to use a set of superconducting dipole magnets, with a inner magnetic field of about 8.3 T, to bend the beams. The Nb-Ti magnets work at a temperature of 1.9 K in a liquid helium bath; inside each of them there are two separate beam pipes for the two colliding beams. The beam is delivered to the LHC for its final acceleration by a multiple steps injection chain, where at the beginning protons are extracted from a bottle of hydrogen gas and accelerated, up to 50 MeV, by a linear machine (LINAC). Then, bunches of protons are prepared and accelerated, up to 26 GeV, in the Proton Synchrotron (PS) before being injected in the Super Proton Synchrotron (SPS), where they reach the energy of 450 GeV. Finally, they are injected into the LHC ring, as shown in Figure 2.1.
During the acceleration phase in the LHC ring, radio-frequency resonant cavities, operating at 400 MHz, give to each bunch a kick of about 0.5 MeV per turn. To reach large instantaneous luminosity, the beam is composed by approximately $3 \cdot 10^3$ bunches, each containing 10^{10}

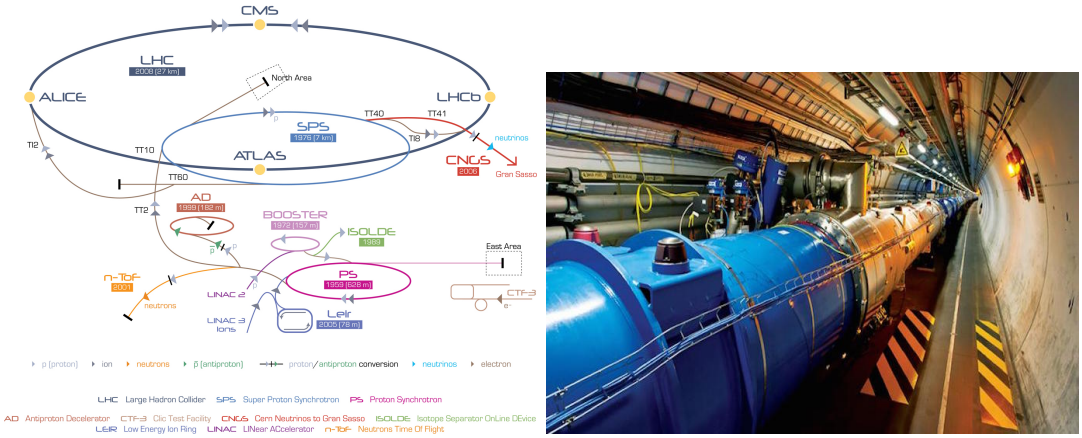

Figure 2.1: (Left) The CERN accelerating system is shown: from the LINAC to the LHC ring. (Right) A recent picture of the LHC magnets, taken inside the underground ring.

protons, separated in time by 25 ns. Beams collide in four interaction points along the LHC, where as many experiments are installed: CMS (Compact Muon Solenoid) [63] and ATLAS (A Toroidal LHC ApparatuS) [64] are general purpose experiments, ALICE (A Large Ion Colliding Experiment) [65] is devoted to heavy ion physics (Pb-Pb), while LHCb (the Large Hadron Collider beauty experiment) [66] is dedicated to flavor physics and CP violation measurements.

2.1.1 Physics goals of the LHC

The LHC already entered in the history of particle physics for the discovery of the Higgs boson and the first measurements of its properties, as reported in Section 1.3. Anyway, the LHC program is not only based on Higgs physics.
The LHC was designed to investigate hard scattering processes, characterized by an interaction energy between 100 GeV and 3 TeV. Within this range, the LHC experiments will search for new phenomena at the TeV energy scale, like new fundamental symmetries of nature (super-symmetry), the existence of additional massive quark and lepton generations, the existence of new heavy gauge bosons (Z′,W′), search for particles predicted in extra dimension models, technicolor, lepto-quarks, etc... Finally, thanks to the high production rate of W and Z-bosons, bottom and top quarks, the LHC experiments will provide precise electroweak measurements, as well as in QCD and B-physics, at an energy scale where the SM has never been tested before.

2.1.2 The LHC operation

The first pp collisions were delivered by the LHC in November 2009, at a center of mass energy of 0.9 TeV and successively at 2.36 TeV, for a total integrated luminosity of 10 μb^{-1} and 4 μb^{-1} respectively. In 2010, the LHC upgraded its colliding energy at $\sqrt{s} = 7$ TeV, providing 40 pb^{-1} of integrated luminosity. Starting from March 2011, the LHC maintained the same beam energy with a progressive increase of its instantaneous luminosity delivering, at the end of the year, 6.10 fb^{-1}.
In the year 2012, LHC operated at $\sqrt{s} = 8$ TeV, delivering an integrated luminosity of 20 fb^{-1} at the instantaneous luminosity of $7 \cdot 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, as shown in Figure 2.2. In addition, Figure 2.2 shows also the time evolution of the integrated luminosity for both 2010, 2011 and 2012.

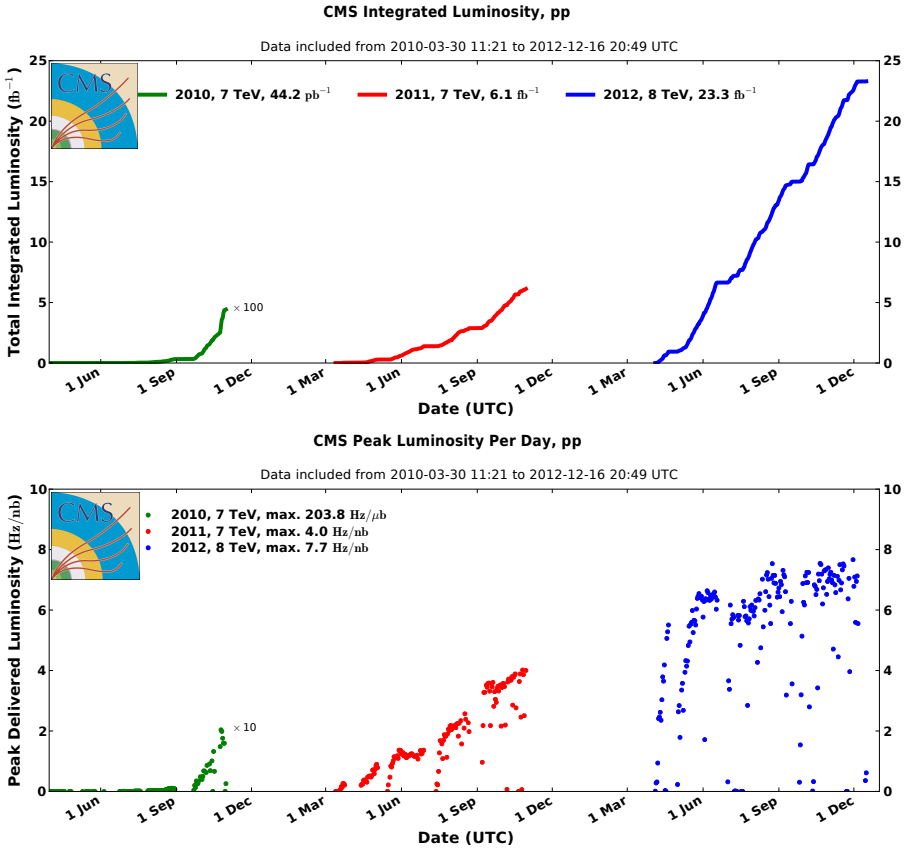

Figure 2.2: (Top) Total integrated luminosity versus time for 2010, 2011 and 2012 data taking. (Bottom) Evolution of the instantaneous luminosity from 2010 to 2012.

During 2013 and 2014, several maintenance interventions have been performed to upgrade the machine to reach a beam energy of 6.5 TeV. In the early spring 2015, the first collisions at $\sqrt{s} = 13$ TeV have been delivered. In Figure 2.3, the display of an event, recorder by CMS experiment during the first set of 13 TeV collisions, is reported.

2.2 Physics at hadron colliders

In a pp collider, the total cross section is in the range of 100 mb, which is consistent with a target area calculated taking the typical transverse size of a hadron of about 1 fm.
The total cross section grows slowly with energy, as shown in Figure 2.4, where the NLO predictions, expected at the LHC, are reported for the most common QCD and electroweak processes as function of the center of mass energy.
The inelastic cross section is mainly composed of events in which only low p_T hadrons are produced in the final state, usually called minimum bias interactions. The main purpose of hadron colliders is to look for events with large transferred momentum or, equivalently, short interaction distance, called hard scattering phenomena.
Cross sections for these events are calculated within the perturbative theory according to the QCD parton model approach. The incoming hadron is treated as a beam of point-like constituents, called partons. If the transferred momentum of the interaction is much larger than the binding energy within the hadron structure, the parton is effectively considered as a

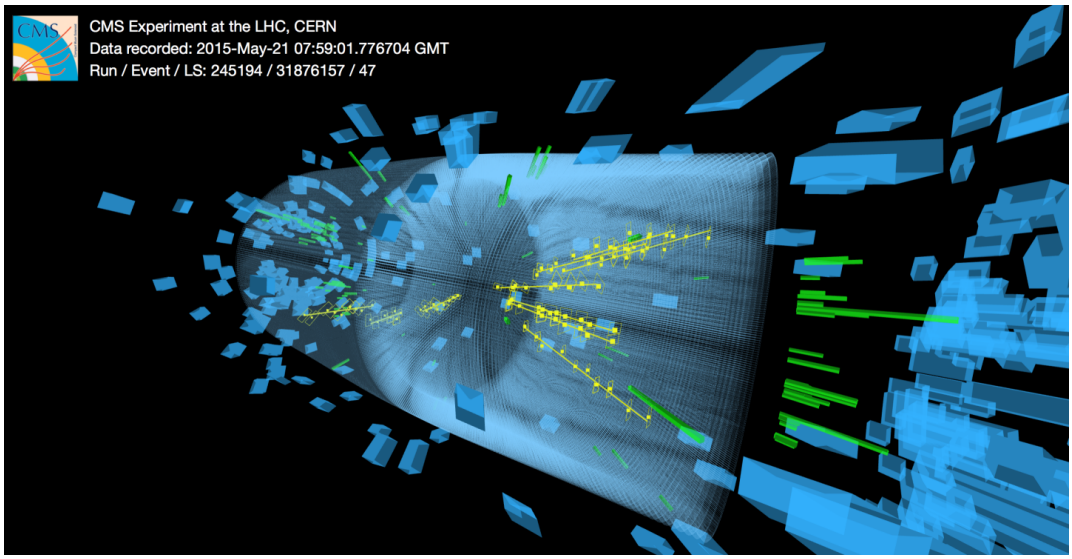

Figure 2.3: An event display recorded by CMS during the first run of unstable collisions, at $\sqrt{s} = 13$ TeV, happend on Thursday 21^{st} May 2015.

free particle. Thus, the parton model is based on two assumptions: describing only process at short distance, that is equivalent to high transverse momentum or large scattering angle, and decorrelating this hard process from the interactions of the hadron residuals.
The inclusive cross section for an hard scattering interaction is calculated through the QCD-improved parton model formula [67]:

$$\sigma_{\mathrm{H_1H_2}}(p_1, p_2) = \sum_{ij} \int dx_1 dx_2 \; f_i^{H_1}(x_1, \mu_f) f_j^{H_2}(x_2, \mu_f) \; \hat{\sigma}_{ij}(x_1 \cdot p_1, x_2 \cdot p_2, \mu_f, \mu_r) \tag{2.1}$$

where the main ingredients are:

- $\sigma_{\mathrm{H_1H_2}}(p_1, p_2)$ represents the hard scattering cross section, such as $t\bar{t}$, W or Z production.
- i and j are generic incoming parton indices, indicating quarks, or anti-quarks or gluons.
- $f_i^H(x, \mu_f)$ are called Parton Density Functions (PDF), which represents the probability to find, inside the hadron structure, a parton of type i with a fraction x of hadron total four-momentum. The PDF must be measured with collision data.
- $\hat{\sigma}_{ij}(x_1 \cdot p_1, x_2 \cdot p_2, \mu_f, \mu_r)$ is the short distance cross section, which is calculated in the perturbation theory framework at a given order.

Equation 2.1 depends on two arbitrary scales: the renormalization scale (μ_r), which scans the evolution of the strong coupling constant absorbing ultra-violet divergences with the renormalization procedure, and the factorization one (μ_f), that enters in the theory to rule collinear divergences from initial state emission. This factorization scale acts as a cut-off, where the finite contribution of the NLO partonic cross section, for $q^2 > \mu_f^2$ where q is the transferred momentum, remains in the partonic term while the divergent one, given by $q^2 < \mu_f^2$, is reabsorbed inside the hadron structure through the PDFs. Usually, the best scales choice consists in a common value close to the energy scale of the considered process, such as m_h for the

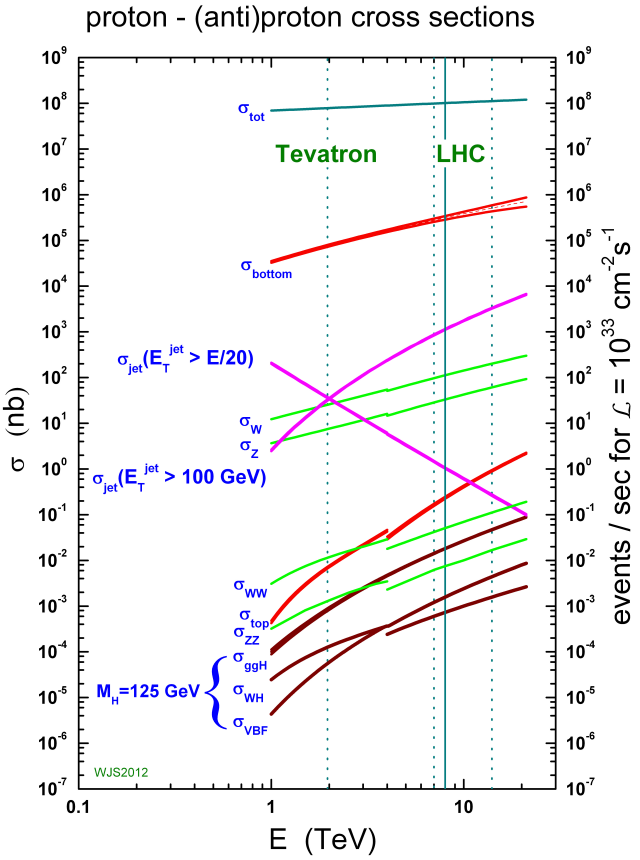

Figure 2.4: NLO cross sections expected at hadron colliders (pp and $p\bar{p}$) as function of $\sqrt{s}$. Dotted vertical lines mark the collision energies at 2 TeV (Tevatron), and at [7,10,13] TeV (LHC).

Higgs production or m_Z for Drell-Yan. The Scale choice represents an important source of uncertainty on the final cross section prediction, together with the approximated knowledge of the PDFs, whose uncertainties are estimated by the PDF for LHC working group [68].

2.2.1 Hadron collider kinematics

Differently from what happens in a e^+e^- machine where, both in the transverse and in the longitudinal plane, the center of mass system is at rest, in a hadron collider the center of mass is at rest only in the transverse one, since the incoming partons p_T are negligible while it is boosted along the beam direction (conventionally called z-axis) due to the unknown parton $\hat{p}_z = x \cdot p$. For this reason, it is useful to look for boost invariant kinematic variables, to better describe hadron collider kinematics.
A convenient set of observables is given by the transverse momentum p_T, the rapidity y and the azimuthal angle φ, which are defined as follows:

$$p_T = \sqrt{p_x^2 + p_y^2} \quad ; \quad y = \frac{1}{2} \log \frac{E + p_z}{E - p_z} \quad ; \quad p_x = p_T \cos\varphi \quad ; \quad p_x = p_T \sin\varphi \tag{2.2}$$

where p_T and φ are invariant under longitudinal boost, while the rapidity itself is not invariant, but the Δy meets this property. Experimentally, it is more convenient to use the

pseudorapidity (η) instead of the rapidity (y). The pseudorapidity is defined as the rapidity for a massless particle, i.e. when $E \approx |\vec{p}|$, which coincides with rapidity (y) only in the ultra-relativistic limit:

$$\eta = \frac{1}{2}\log\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z} = -\log[\tan(\vartheta/2)] \quad ; \quad |\vec{p}| = p_T \cosh\eta \quad ; \quad p_z = p_T \sinh\eta \tag{2.3}$$

where ϑ is the polar angle, defined between $\vec{p}$ and the beam axis.
Finally, in the (η, φ) plane, one can define a boost invariant metric commonly used to measure distances between objects:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\varphi^2} \tag{2.4}$$

2.3 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is a general purpose detector installed along the LHC ring. The detector design and layout were driven by the choice of the magnetic field configuration, fundamental for charged particles momentum measurement. The core of the experiment consists of a superconducting solenoid, 13 m long and with 5.9 m inner diameter, able to produce a magnetic field of about 3.8 T [69]. This allows to have high momentum resolution, at the order of few percent for muons with p_T up to 1 TeV. It is called compact because three main sub-detectors are placed inside the solenoid: the inner tracking system, which is the nearest detector to the interaction point, the electromagnetic calorimeter (ECAL) and the hadronic one (HCAL). The solenoid return field is used to saturate the external iron holding structure, where four layers of muon gaseous chambers are placed to track outgoing muons.
The CMS structure follows the typical design of collider experiment, where a cylindrical central section, called barrel, is closed by two endcaps. The whole structure, approximately 22 m long and 15 m high, is shown in Figure 2.5, while a transverse plane section is reported in Figure 2.6.
The adopted cartesian coordinate system has the origin placed in the nominal interaction point, coinciding with the geometrical detector center. The x and y axes define the transverse plane, while the z axis points along the beam direction. Therefore, the azimuthal angle φ is measured on the x-y plane, starting from the x axis, while the polar angle ϑ is measured from the z-axis and the radial distance r is referred to the beam line.

2.3.1 The tracking system

The CMS tracking system provides a precise and efficient reconstruction of charged particle tracks, together with a high resolution momentum measurement [70]. The tracker determines the three-momentum of each charged particle, measuring the curvature of its track in the transverse plane, thanks to the presence of a 3.8 T magnetic field parallel to the beam axis, induced by the superconductive solenoid. The trajectory deviation from the straight line propagation is measured by the sagitta:

$$s \approx \frac{0.3}{8}\frac{BL^2}{p_T}$$

In this way, once known the magnetic field B, the track lenght L and the sagitta s, the transverse momentum p_T of the track is measured and, as a direct consequence, also the tree-momentum $|\vec{p}| = p_T/\sin\vartheta$.

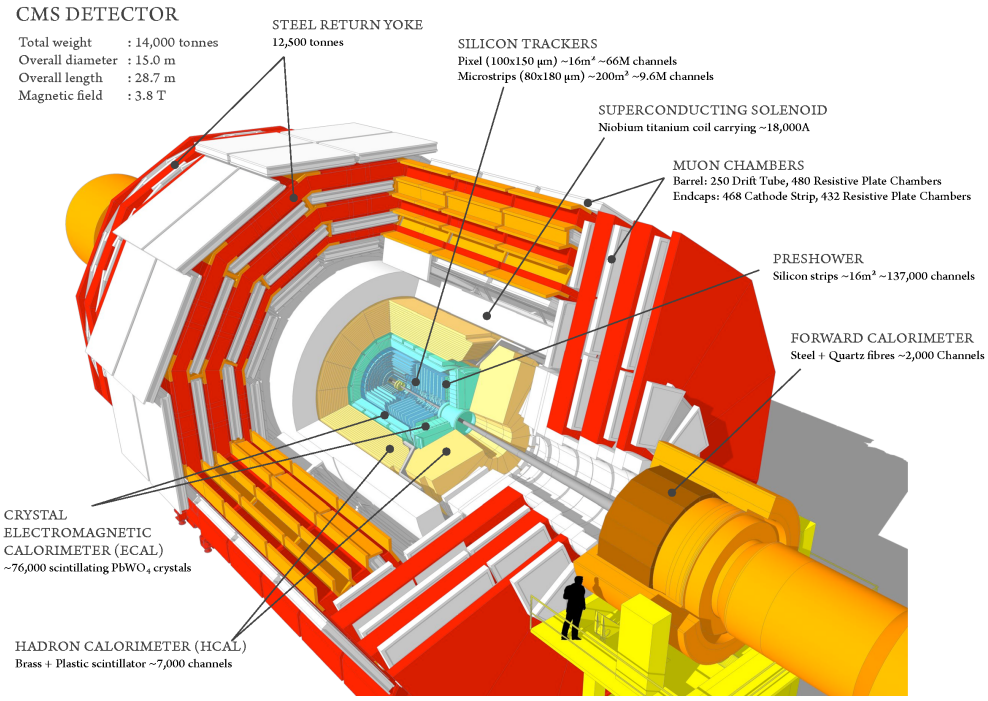

Figure 2.5: The CMS detector in a tridimensional view, where the various sub-detectors and their main features are also indicated.

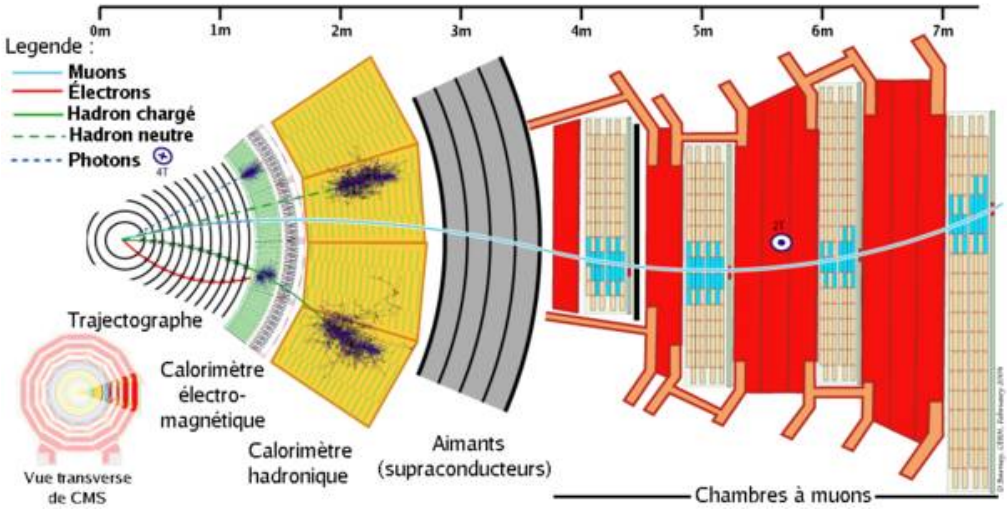

Figure 2.6: The CMS transverse plane: sub-detectors sequence and, as example, the interaction of different type of particles with the detector is also shown.

The transverse momentum resolution is mainly dependent on the geometric accuracy on the sagitta (σ_s), through:

$$\frac{\sigma_{p_T}}{p_T} \approx \frac{\sigma_s}{s} = \frac{8}{0.3BL^2} \cdot p_T \cdot \sigma_s \tag{2.5}$$

Thus, an optimal p_T resolution is provided by a high magnetic field (B) and long reconstructed track (L). The accuracy of this measurement improves with decreasing p_T of the particle.

The instantaneous luminosity, delivered by the LHC during the 8 TeV campaign, implies an average of about twenty *pp* interactions in the same bunch crossing (BX), therefore one of the main challenges is to operate in this harsh environment with competitive performances. This will be even more important during the 13 TeV data taking (Run-II), where around 40 pileup interactions are expected for each BX. Requirements on the granularity, speed and radiation hardness lead to a silicon based tracker, with the consequence of dealing with high power detector electronics, which requires an efficient cooling.
In CMS, the tracker is the closest sub-detector with respect to the beam line. It is a cylindrical system 5.8 m long, with an internal diameter of 6 cm and an external one of 2.2 m, composed by ten layer of silicon sensors. Its total volume is 24.4 m^3 and the running temperature is around -10°C. The CMS tracker is composed of two modules: the innermost pixel system and the strip one surrounding it. A global picture of the whole tracker system is shown in Figure 2.7. The pixel sub-detector consists of three barrel layers (BPix) and two endcap disks

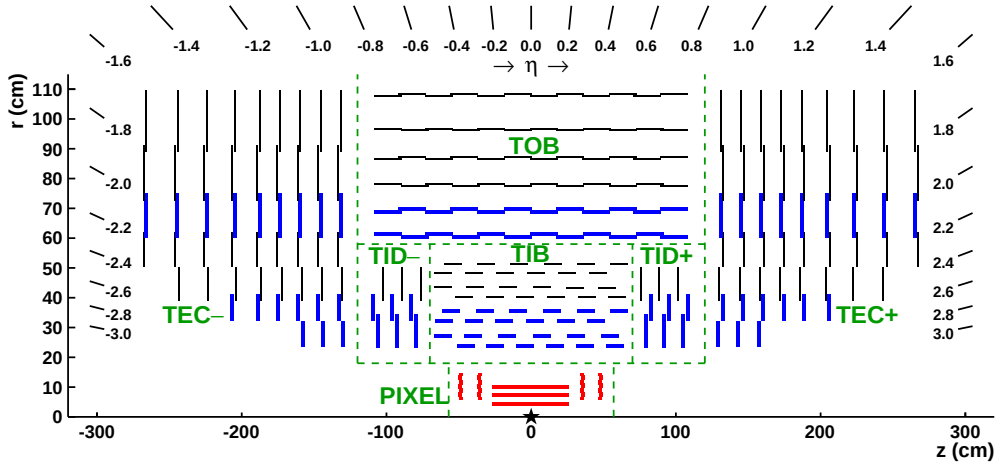

Figure 2.7: Schematic summary of the CMS tracker in the r-z plane. Green dashed lines are used to indicate each tracker subsystems. Strip modules that provide two-dimensional hits are shown by thin black lines, while those permitting the reconstruction of hit positions in three-dimension are shown by thick blue lines.

(FPix). BPix layers are located at radii of about 4.4 cm, 7.3 cm and 10.2 cm, while the FPix disks are placed at $|z| = \pm 34.5$ cm and $|z| = \pm 46.5$ cm.
The estimated position resolution on the single hit is of 10 μm for the (r, φ) coordinates and 15 μm for z in the barrel, while it is of 15 μm and 20 μm, respectively, in the endcaps. The pixel detector covers an area of about 1 m^2 with 66 million channels and the arrangement of barrel layers and endcap disks gives three hits over almost the full tracker coverage $|\eta| \leq 2.5$ [71]. The hit efficiency is substantially flat, close to 100% up to $|\eta| = 2.1$, then it degrades reaching 50% at $|\eta| \approx 2.5$.
Finally, the silicon strip detector is made of two concentric sets of layers in the barrel (TIB and TOB) and two blocks of forward disks in the endcaps, called TEC and TID, as reported in Figure 2.7. The strip tracker has a total of 9.3 million strips and 198 m^2 of active silicon area, making it the largest entirely silicon-based detector ever built. The single-point resolution is of about 30 μm in the r-φ plane and 300 μm in the z direction.
This layout is in conflict with the aim of minimizing the amount of material placed in front of the calorimetric system, to limit the probability of multiple scattering, nuclear scatter-

ing, bremsstrahlung and photon conversion, which introduce additional terms in the effective calorimeter resolution. The radiation length (X_0) is a characteristic of a material, related to the amount of energy lost by high energy charged particles that interact with it. It is defined as the mean length, usually expressed in $g \cdot cm^{-2}$, over which an electron reduces its energy by a factor 1/e by emitting bremsstrahlung radiation. The material budget, inside the tracker active volume, varies from 0.4 X_0 at $|\eta| \approx 0$ to 1 X_0 at $|\eta| \approx 2.5$, with a maximum of 1.8 X_0 in the transition region between the barrel and each endcap placed at $|\eta| \approx 1.5$. Its η-profile is shown in Figure 2.8. This constitutes one of the main source of uncertainty in the calorimetric measurements of electrons, that loose energy in the tracker volume by means of bremsstrahlung radiation, and photons, which may convert into e^+e^- pairs before reaching the calorimeter surface.

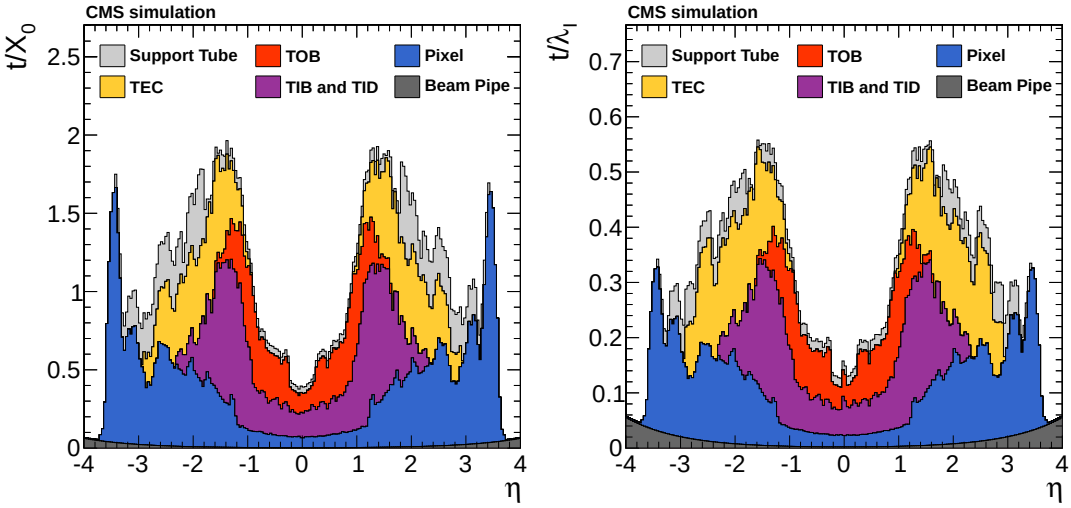

Figure 2.8: Total thickness (t) of the tracker material crossed by a particle, as a function of η, expressed in units of radiation length X_0 (left) and nuclear interaction length λ_I (rigth).

2.3.2 The Electromagnetic Calorimeter (ECAL)

The CMS Electromagnetic CALorimeter (ECAL) [72] is an hermetic and homogeneous system made of inorganic scintillating crystals, which provides precise electron and photon identification and energy measurements. This detector plays a central role in crucial physics searches, in particular in topologies with hard photons in the final state, like the search for the SM Higgs boson in the $h \to \gamma\gamma$ channel. ECAL is composed of 61200 lead tungstate crystals ($PbWO_4$), mounted in the central barrel part, closed by two endcaps including 7324 crystals each. These crystals are arranged in a quasi-pointing geometry, with their main axis oriented towards the centre of the detector. A small tilt (about 3°) prevents particles, produced at the nominal interaction point, from traversing the ECAL surface without crossing any crystal.

- **Ecal Barrel (EB)**: it covers the pseudorapidity range between $0 < |\eta| < 1.479$, where each crystal has a truncated pyramid shape with a front face cross section of $22 \times 22\,mm^2$ and 23 cm of length, that corresponds to 25.8 X_0. The single crystal coverage is $\Delta\eta \times \Delta\varphi = 0.00175 \times 0.00175$ and the barrel granularity is 360-fold in φ and (2×85)-fold in η. Crystals are grouped in arrays of 2×5, contained in a very thin 200 μm alveolar structure, forming a sub-module. Group of 40/50 sub-modules are then arranged in a module and

groups of four modules are assembled, with metallic cross plates, to form a bigger unity called super-module. Finally, EB is divided into 36 super-modules, each one with a coverage of 20° in φ. A global picture of EB is reported in Figure 2.9.

- **Ecal Endcaps (EE)**: it covers the pseudorapidity region between $1.479 < |\eta| < 3$ with identically shaped crystals, grouped in carbon-fiber structure of 5×5 elements, called super-crystals. The crystals have a cross section of $28.62 \times 28.62\,\mathrm{mm}^2$ and a length of 22 cm. Figure 2.9 shows the geometry adopted to arrange crystals in the endcaps.
- **Ecal Preshower (ES)**: it is placed in front of each endcap, improving π^0/e and π^0/γ discrimination, as well as electrons and photons position determination. The preshower is a sampling system with two different layers: two passive lead radiators, that initiate electromagnetic showers from incoming photons or electrons, and two active silicon strip planes, placed after each radiator, able to measure the deposited energy and the transverse shower profile. The total material budget corresponds to about $3\mathrm{X}_0$ over the covered region $1.65 < |\eta| < 2.6$.

A longitudinal view in the (r, z) plane, showing the geometric organization and the tilt of the crystal modules in EB and EE, is reported in Figure 2.10.

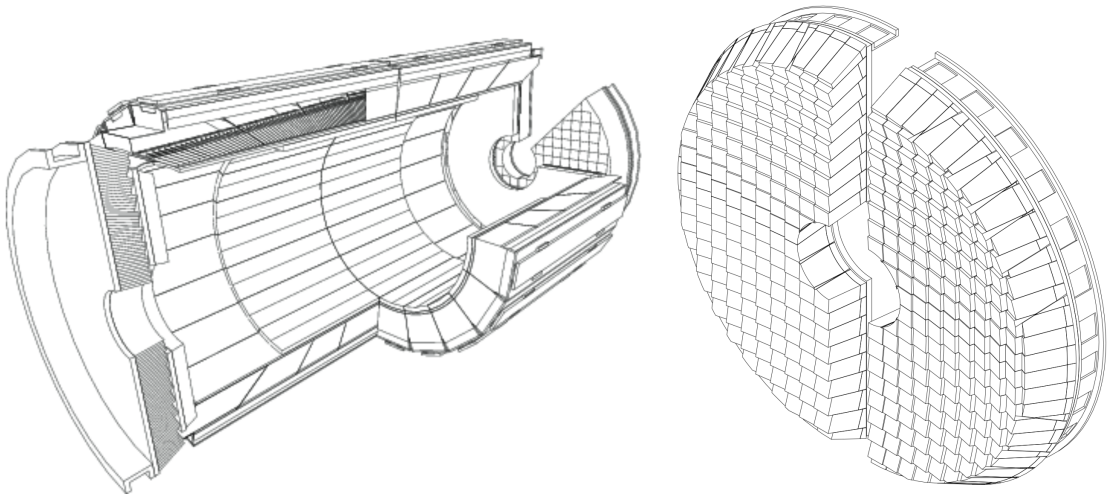

Figure 2.9: CMS ECAL geometry schema: (left) the ECAL barrel (EB) made of 36 super-modules: 18 in EB+ ($z > 0$) and 18 in EB− ($z < 0$). (Right) The ECAL endcap (EE) is divided in 4 Dees: 2 Dees in EE+ ($z > 0$) and 2 Dees in EE+ ($z < 0$).

Lead tungstate crystals were chosen as active detector because of their properties:

- They are dense ($\rho = 8.2$ g/cm^3) with a small radiation length ($\mathrm{X}_0 = 0.89$ cm), which allows for a compact calorimeter with short crystals (23 cm long).
- Small Molière radius ($r_\mathrm{M} = 2.2$ cm) that limits the transverse shower development within few crystals, which represents an important handle to limit pileup effects.
- A fast response to the excitation, in fact around 80% of the scintillation light is emitted within 25 ns, which complies with the fast LHC collision rate.
- They are radiation hard, thus able to withstand the harsh LHC environment.

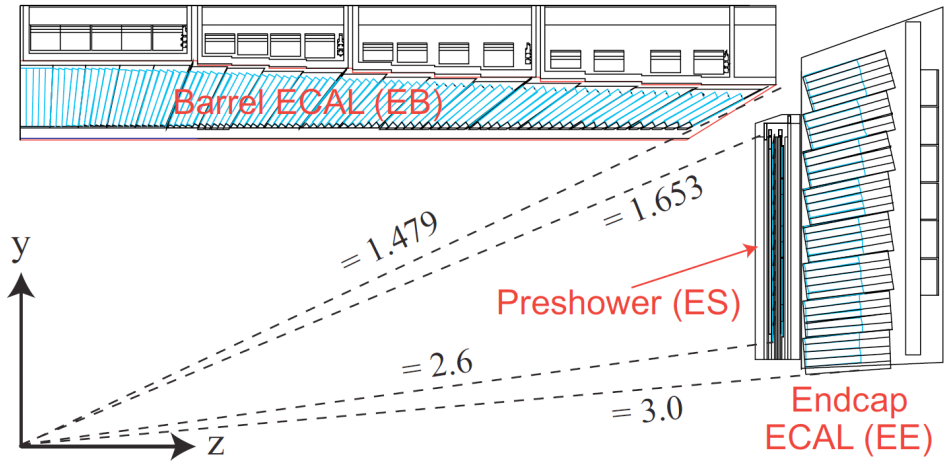

Figure 2.10: The crystal tilt in shown in a transverse view of ECAL. To produce a non-pointing geometry along η, crystal longitudinal axes are all inclined by 3° with respect to the line joining their front face centre to the interaction point.

The light yield of these crystals is low, around 50 photons per MeV, therefore high-gain photodetectors are necessary for the scintillation light readout. Avalanche Photo-Diodes (APD) are used in EB, while Vacuum Photo-Triodes (VPT) in EE. The APDs have an active area of $5{\times}5\,\mathrm{mm}^2$ and a pair of them is placed on the rear face of each crystal. Both crystals and APD responses are sensitive to temperature variations, thus an efficient cooling system is required to keep the temperature stable at 18°C, with oscillations smaller than 0.1°C.
In addition, a stable power supply system has been designed for the APDs, able to control the high voltage variations at the order of millivolts, minimizing effects on the energy resolution at 0.1% level. On the other hand, VPTs have a lower quantum efficiency and an internal gain. The ECAL electronic readout, called Front End (FE) system, amplifies, shapes and digitalizes the signal with an ADC stage every five crystals.
The global information coming from matrices of 5×5 crystals, called Trigger Towers (TT), are transmitted to the calorimeter trigger after each bunch crossing. In EB, trigger towers are classified into three categories, comparing the deposited energy with two chosen thresholds. A flexible usage of this classification, in terms of thresholds and suppression readout levels, allows to adapt the readout to the variation of the data taking conditions. High energetic trigger towers ($E_T > 5$ GeV) activate the full readout of all their crystals and those in the neighboring ones (225 crystals for the EB case). In the TT energy lies within $2.5 < E_T < 5$ GeV, only TT crystals are read without applying any energy suppression threshold, while low energy trigger towers, far from high energetic ones, are readout only if their energy exceed a zero suppression level, about 3σ above the expected noise.
The radiation, produced by the beam collisions, affects the crystal transparency via the creation of color centers that absorb and scatter the light, while it has almost no impact on the scintillation mechanism. This damage is dose rate dependent, so crystals loose transparency during LHC physics runs proportionally to the exposure time, then they recover in periods without radiation. To follow the damage curve and correct the energy measurement of each crystal during time, a dedicated laser monitoring system has been implemented, to keep the calorimeter performance stable in time. A laser light is injected into the crystals via optical

fibers, at two different wavelengths $\lambda_b = 440$ nm (blue light) and $\lambda_r = 530$ nm (green light). The blue light is near to the scintillation peak and it measures the transparency change due to radiation, while the green one is far from the typical scintillation wavelength and since its propagation is not affected by transparency losses, it is used as a reference benchmark.
The APD response to laser pulses is normalized and compared with a direct characterization of the source, made by a reference silicon PN. The ratio between the amplitudes APD(t) and PN(t) is equivalent to the crystal transparency loss. Laser pulses are sent to ECAL crystals during the 3.17 μs long LHC beam gaps. The time needed to scan the entire detector is of about 30 minutes.

2.3.3 The Hadron Calorimeter (HCAL)

The aim of the hadron calorimeter is the measurement of charged and neutral hadron energies, which produce shower of particles when they interact with the CMS detector [73]. In fact, when a hadron interacts with matter, an hadronic shower is developed mainly by a sequence of inelastic scatterings, producing a multi-particle final state.
To accomplish this goal, the detector must have a high granularity and a good hermeticity, covering the largest possible η range. The typical scale of the hadronic shower development is represented by the nuclear interaction length λ_I, which depends mainly on the material absorber in a sampling system. Since λ_I is usually larger than X_0, hadronic calorimeters are longer and placed behind the electromagnetic ones. The incoming hadron energy is measured through the energy sum of all the charged and the neutral particles inside the shower but, differently from an electromagnetic calorimeter, the energy resolution is typically much worse, due to several effects:

- Hadrons can decay via weak interactions, producing a neutrinos in the final state that don't interact with the detector.
- Part of the absorbed energy is used for inelastic nucleus splitting.
- Hadronic showers are mainly composed of pions ($\pi^{\pm}$, π^0) thus, thanks to $\pi^0 \to \gamma\gamma$ decays, a large electromagnetic signal is also produced inside hadronic showers for which hadronic systems have usually a poor response.

In CMS, HCAL consists of a sampling calorimeter divided into a barrel part (HB, $0 < |\eta| < 1.3$) and two endcaps (HE, $1.3 < |\eta| < 3$), with a transverse granularity of $\Delta\eta \times \Delta\varphi = 0.087 \times 0.087$ and $\Delta\eta \times \Delta\varphi = 0.17 \times 0.17$, respectively. HB is radially restricted between the outer radius of ECAL ($r = 1.77\,\text{m}$) and the inner extent of the solenoid ($r = 2.95\,\text{m}$). It is composted of active layers of plastic scintillators interspersed with brass absorbing ones. The light collection is performed by wavelength shifters, while the readout adopts a set of hybrid photodiodes (HPD). The total absorber thickness at $\eta \approx 0$ is 5.82 λ_I, then it increases with the polar angle ϑ as $1/\sin\vartheta$, resulting in 10.6 λ_I at the end of the barrel, while, in the endcaps, the total length is about 10 λ_I. In addition, there is also an outer module called HCAL Outer (HO), which is placed outside the magnet coil, to improve the shower containment in the barrel region $|\eta| < 1.26$.
The HCAL geometric coverage is extended from $|\eta| = 3$ to $|\eta| < 5.2$ by a forward calorimeter (HF), which is placed at 11.2 m from the interaction point. It is a radiation hard Cherenkov-base detector, made of quartz fibers embedded in bulky steel with a transverse granularity of

$\Delta\eta \times \Delta\varphi = 0.175 \times 0.175$. The light readout is performed by photomultipliers. The choice of this technology is related to its high radiation resistance, needed to survive in the very forward region. A longitudinal section of the whole detector is presented in Figure 2.11.

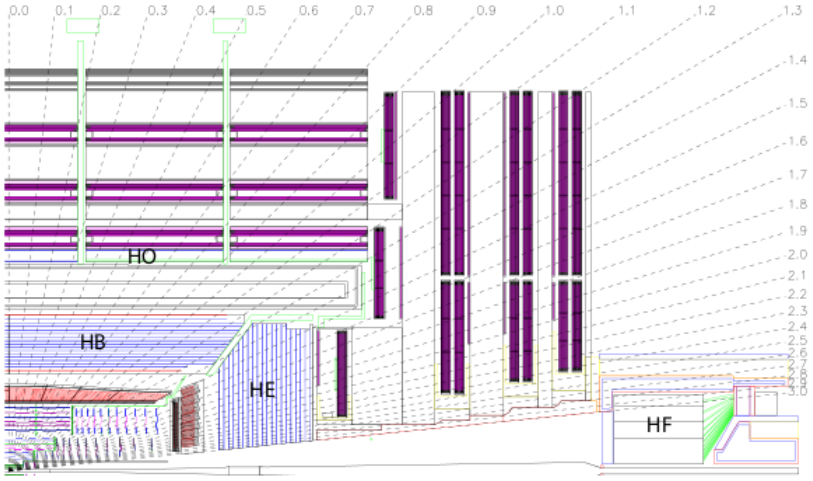

Figure 2.11: Longitudinal view in the (r, z) plane of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters.

2.3.4 The muon system

A dedicated muon detection system is placed outside the magnetic coil, which provides coverage for muon reconstruction up to $|\eta| \approx 2.4$ [74]. Three types of gaseous detectors are used to identify and reconstruct muons from a set of hits, according to different radiation environments along η and profiting from the strong magnetic field, granted by the solenoid return yoke. The barrel part ($|\eta| < 1.2$) consists of four layers, with radius between about 4 m and 7 m, in which drift tubes (DT) and resistive plate chambers (RPC) are installed. These modules are arranged so that a muon, produced by a collision, traverses at least three detection layers. On the contrary, in the forward region, cathode-strip chambers (CSC) and RPCs are mounted perpendicularly to the beam line. In Figure 2.12 a (r, z) view is reported to show the muon system components. The detector choice has been driven by trigger needs, excellent spacial resolution and the capability of operation in a dense radiation environment:

- DTs are used in the central barrel part, where the average particle fluence is low. In a DT chamber, many drift tubes are arranged in two super-layers with a central anode parallel to the beam line, providing a measurement of the (r,φ) coordinates. At the same time, an additional layer, with the anode perpendicular to the beam line, allows to measure the z one.

- CSCs are multi-wire proportional chambers in which the cathode plane is segmented into strips, running across wires. An avalanche collected by a wire induces a charge on several strips of the cathode plane. Therefore, they can measure two coordinates per unit, using both the information produced, by the same track, on the wires and on the segmented plane.

- A RPC consist of a thin double-gap chamber characterized by a poor spatial resolution with respect to DTs and CSCs but, thanks to a fast collection charge, they have an

optimal time resolution (< 1 ns). Therefore, they are used for trigger purposes and for an unambiguous identification of the bunch crossing time.

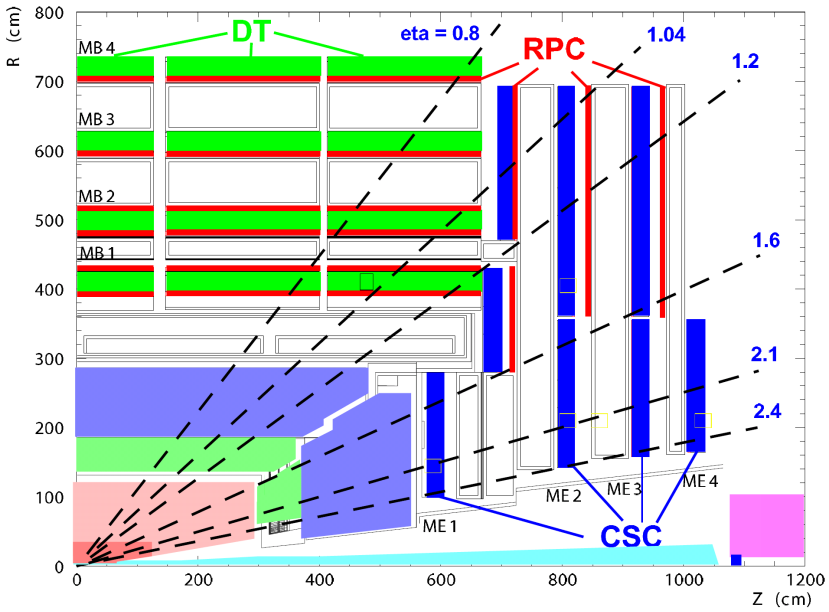

Figure 2.12: Layout of one quarter of the CMS muon system: RPC, CSC and DT positions are shown.

2.3.5 The trigger system

In a hadron collider, during each bunch crossing a measurable amount of energy is deposited in the detector. During the LHC Run-I, a collision rate of 2 GHz was reached, largely dominated by minimum bias events. Therefore, to collect only interesting data without saturating the limited storage capability, a trigger system was developed with the purpose of providing a large rate reduction factor on minimum bias events, maintaining a high efficiency on the interesting ones [75]. The total output rate is reduced at a level of 300 Hz by means of a two-layer system: a hardware Level-1 (L1) trigger and a High-Level Trigger (HLT), which is software based.

- **Level-1 Trigger**: after each bunch crossing, raw data are temporarily stored in pipelines, waiting for the trigger decision about taking or discarding events, which has to be taken within 3 μs. For each bunch crossing, the L1 decision time is too short to use the whole detector information, therefore this decision is based only on ECAL, HCAL and the muons system response. The calorimeter trigger is based on trigger towers, matrices of 5×5 crystals in ECAL, which match the granularity of the HCAL ones, while the muon trigger is performed separately for each muon detector. The result of muon and calorimeter triggers is then passed to a global one (GT) that takes the final decision.

- **HLT Trigger:** it is a fully software trigger system which reconstructs physics objects only in interesting regions of the detector. This leads to the development of three virtual trigger levels: the first level uses only information from the muon system and the calorimeters, the second adds tracks, while the last one is based on a full reconstruction, where events are selected according to their final state topology.

2.4 Physics object reconstruction

Raw data, extracted from interesting collisions inside CMS, need to be assembled to provide offline physics objects for the analyses (electrons, muons, jets .. etc).
In this Section, a short summary of different criteria used for the event reconstruction, after the HLT decision, is presented. In particular, in Section 2.4.1 muon reconstruction algorithms are exposed, while in Section 2.4.2 the electron ones. Finally, Section 2.4.4 and Section 2.4.5 are dedicated to jets and missing energy ($\not\!\!E_T$) definitions. In order to be consistent with the analysis results that will be discussed in the next Chapters, only the techniques adopted during Run-I are described. Only in a few cases, strategies used for the current data taking are discussed, as specified in the text.

2.4.1 Muon reconstruction

Muon are detected merging information from the tracking (tracker tracks) and the muon system (stand-alone muon tracks) measurements, following two possible approaches [76]:

- **Global muon reconstruction**: starting from stand-alone muon tracks, a matching with tracker ones is found comparing the parameters of the two tracks after a propagation on a common surface. Then, a global muon track is fitted using a Kalman-Filter algorithm [77]. This strategy is particularly useful to improve momentum resolution for high p_T muons ($p_T > 200$ GeV).

- **Tracker muon reconstruction**: all tracks with $p_T > 0.5$ GeV and $|\vec{p}| > 2.5$ GeV are considered as seeds, propagating them untill the muon system surface. If a muon segment is matched with the track, according to a local Δx Δy criterion, a tracker muon is formed. This strategy is more efficient at low momenta, in particular when $p_T < 5$ GeV.

These two collections are merged into a single one, providing an overall reconstruction efficiency of about 99%. In addition, a new class of muons, called particle-flow muons, is also defined, where further selection criteria are applied using constraints from all CMS sub-detectors. This helps to improve both reconstruction and identification of muons inside jets, also known as non-isolated muons. In addition, dedicated procedures for high p_T muon reconstruction and momentum assignment have been designed:

- **Tracker plus first muon station (TPFMS)** consists of a re-fit of the global track using only the first hit in the muon chambers, to reduce the sensitivity to secondary showers inside the muons system.

- **Picky track** is a re-fit of the global track using only muon chamber hits that are compatible, by means of a χ^2 criterion, with the extrapolated trajectory.

Eventually, to further improve momentum resolution, the Tune-P algorithm selects the best of these alternative measurements, on a muon-by-muon basis, according to the quality of each fit. This allows, looking at $Z \to \mu\mu$ events, to keep the muon momentum scale stable as a function of p_T, with a small residual η dependence. In fact, Figure 2.13 (left) shows the stability and the residual differences, between data and simulation, for the reconstructed Z-boson mass peak in $Z \to \mu\mu$ events as a function of η. In addition, in Figure 2.13 (right), the

muon momentum resolution is reported which varies from 1.5% in the central barrel to 5% in the outer endcap ($\eta = 2.1$).

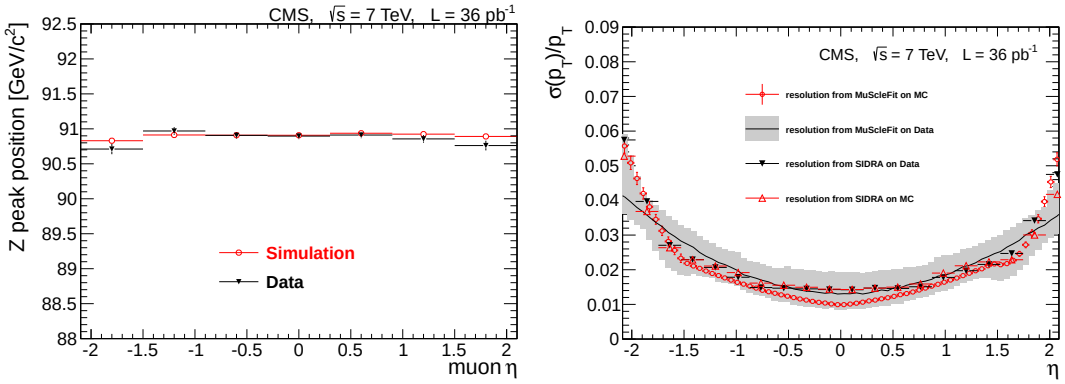

Figure 2.13: (Left) Comparison of the differences between the reconstructed Z-boson mass peak in data and in the simulation, as a function of the pseudorapidity. (Right) Transverse momentum resolution $\sigma(p_T)/p_T$ in data and simulation as measured in $Z \to \mu\mu$ events. Only statistical uncertainties are displayed.

2.4.2 Electron reconstruction

Electrons are reconstructed associating a charged track, fitted from hits in the tracking system, with a cluster of energy depositions in ECAL.

2.4.2.1 Electron clustering in ECAL

The electromagnetic shower, developed by an electron interacting with the ECAL detector, is well contained usually in a small number of crystals. In fact, at the test beam, almost 97% of a 120 GeV electron energy was deposited within a 5×5 matrix of crystals, centered around the impact position [78]. Unfortunately, relativistic electrons, produced in a hard scattering phenomena, interact with the tracker material emitting bremsstrahlung radiation, which distributes the total electron energy among a large number of crystals. Due to the track curvature in the transverse plane, electron energy is mainly spread along φ. Thus, to have a reliable energy estimation, it is necessary to collect bremsstrahlung photons together with the prompt shower energy. Two independent strategies have been developed to cluster energy in EB and EE [79]:

- The **Hybrid algorithm** exploits EB geometry, collecting crystals in a small η window but extended along φ. This algorithm starts from a single crystal with transverse energy grater than 1 GeV (seed), then arrays of 5×1 crystals in $\eta \times \varphi$ are grouped around it if their E_T is above 0.1 GeV. Contiguous arrays are then merged into basic clusters, where the most energetic crystal of the cluster must have a transverse energy larger than 0.35 GeV. Finally, all the basic clusters, belonging to the same $\eta \times \varphi$ macro-region, are gathered together in a super-cluster.

- **Multi 5x5 algorithm** exploits the EE geometry and starts from a seed crystal defined as the one with largest energy inside a 5×5 matrix, requiring $E_T^{\text{seed}} > 0.18$ GeV. Around each seed, starting from the one with the largest E_T, the energy is collected building

5×5 clusters that can partially overlap. The shared energy is assigned to each cluster according to their corresponding E_T and the minimum super-cluster energy threshold is 1 GeV. Eventually, ECAL super-clusters are matched with pre-shower ones, summing up the related energy measurements.

Figure 2.14 shows the effect of these clustering techniques, looking at $Z \to ee$ simulated events, with respect to consider the energy of a simple 5×5 matrix around a seed crystal. It is evident how the designed clustering strategy allows to better recover the generator level performance with respect to the 5×5 matrix.

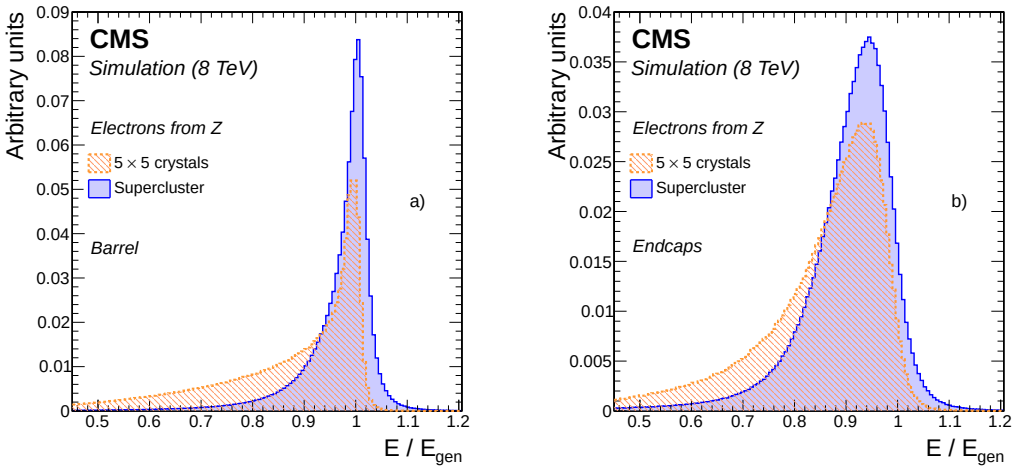

Figure 2.14: Ratio between reconstructed and generated electron energies produced by Z-boson decays, comparing the adopted clustering algorithms performance with the energy of 5×5 matrix: (left) EB, (right) EE.

2.4.2.2 Electron track reconstruction

Since large bremsstrahlung radiation can be emitted by an electron when crossing tracker material, a dedicated track reconstruction procedure, called Gaussian Sum Filter (GSF) algorithm, has been implemented. It is seeded by either the reconstructed super-cluster position in ECAL, or by fitted tracks pointing to an already existing super-cluster.
In the first case, an helix is back propagated from the ECAL surface to the pixel detector, considering both charge hypothesis and assuming an energy equal to the ECAL measurement. The found pairs or triplets of pixel hits are used as seeds for the GSF track fit. This strategy is usually more efficient for high p_T and isolated electrons. In contrast, tracker based seeding is particularly important for low p_T or non-isolated electrons. Tracker driven electron seeds are either Kalman tracks that well match, both geometrically and by compatibility with the measured transverse energy, ECAL super-clusters or tracks pointing to a super-cluster characterized by a bad quality χ^2, due to bremsstrahlung loss.
The selected seeds, which can be pairs or triplets of pixel hits, are used to initialize the electron track building. The bremsstrahlung emission probability is approximated by a superposition of Gaussians, to model a Bethe-Heitler function, which describes the differential cross section of QED soft photon emission. At each step, compatible hits are selected on the following available layer, by means of a navigation module taking into account multiple scattering and

bremsstrahlung probabilities across tracker material. The compatibility between predicted and measured hits is determined through an iterative χ^2 fit, performed at each evolution step using all the track candidate hits collected. At least of five total hits and a maximum χ^2 value are finally required to create a GSF track.

2.4.2.3 GSF electron candidates

Electron candidates are obtained from the association of a fitted GSF track and a super-cluster in ECAL. For ECAL driven electrons, the association is done by a simple geometrical ΔR matching between the super-cluster position, used to seed the track fitter, and the GSF track extrapolated to the ECAL surface. For tracker driven ones, a selection on a MVA discriminator is adopted [79].
The electron momentum is obtained by the combination of tracking and calorimeter information. A regression technique identifies, independently for EB-EE and high-low bremsstrahlung electrons, a weight (ω) used to combine super-cluster and track information as $p = \omega \cdot p_{\text{track}} + (1-\omega) \cdot E_{\text{SC}}$. The regression inputs are the super-cluster energy and its relative uncertainty, the track momentum and its relative uncertainty, the $E_{\text{SC}}/p_{\text{track}}$ ratio and its uncertainty.
Figure 2.15 shows the electron momentum resolution as a function of the generated p_T (left) or energy (right), comparing the super-cluster and the tracking estimate to the combined result. The improvement is sizable especially for low p_T electrons, in both EB and EE.
Figure 2.16 (left) shows the improvement on the reconstructed $h \to ZZ^* \to 4e$ invariant mass resolution, where the σ_{eff} stands for the RMS of the distribution, while σ_{CB} is the width extracted from a fit of the lineshape using a convolution between a Breit-Wigner core and a Crystal-Ball function. Finally, in Figure 2.16 (right) a comparison between the Z-boson mass peak, reconstructed both in data and in the simulation, is reported.

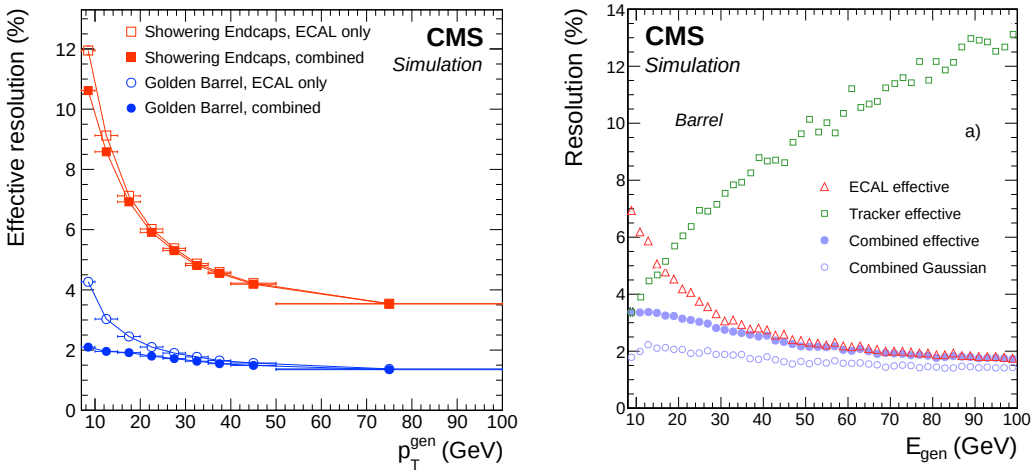

Figure 2.15: (Left) Effective electron momentum resolution after combining the E_{SC} and p_{track} estimates as a function of the generated p_T, where electrons are simulated with a uniform η and p_T spectrum. Golden stands for low bremsstrahlung electrons, selected requiring $E_{3\times3}/E_{\text{SC}} > 0.9$, while showering electrons are obtained inverting the previous condition. (Right) Another comparison between combined momentum assignment, tracker and ECAL measurements.

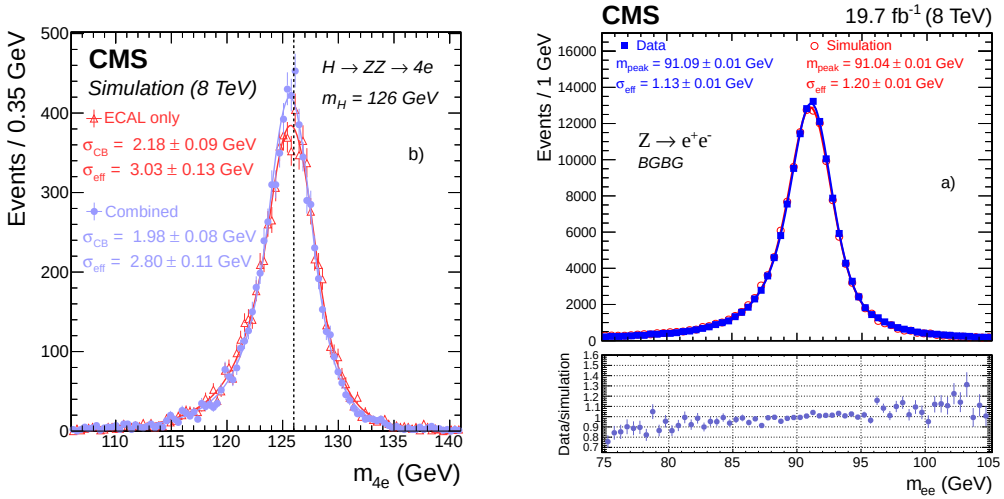

Figure 2.16: (Left) Reconstructed mass of the Higgs boson for $h \to ZZ^* \to 4e$ simulated events using either the corrected super-cluster energy (open triangles) or the electron momentum obtained combining E_{SC} and p_{track} estimates (solid dots). (Right) The di-electron invariant mass distribution from $Z \to ee$ events in data (solid squares) compared with the simulation (open circles) fitted with a convolution of a Breit–Wigner function and a Crystal-Ball one.

2.4.3 CMS Particle-Flow

The particle flow (PF) algorithm aims at reconstructing all the stable particles produced in an event, combining the information coming from all CMS sub-detectors to optimize particle identification, direction and energy determination. The PF starting elements are fitted tracks and calorimeter clusters, which must be delivered with high efficiency and low fake rate, requiring advanced tracking and clustering algorithms [80][1]. These elements are then connected to each other by a linking logic, to fully reconstruct single particles. This algorithm produces a set of blocks according to the following criteria: charged tracks are linked to calorimeter clusters extrapolating the last hit of the track to ECAL, at a depth equal to the longitudinal maximum of a shower profile, and to HCAL at a depth equal to one interaction length. All the clusters, crossed by this extrapolation, are linked to the track. Bremsstrahlung photons are linked to electron candidates extrapolating, up to the ECAL surface, the tangents to the track originated from the crossing points between the track and each tracker layer. If the tangent matches a cluster, this cluster linked to the track.
Calorimeter clusters are linked together if the position of a cluster in ECAL falls within the boundaries of a HCAL one. Eventually, a link among charged-particle tracks and muon segments is done looking at the χ^2 of the global track fit. After this, all these blocks are used to build particle candidates:

- **Particle-flow muons** are composed by all the global muons with a combined momentum measurement compatible, within $3\sigma_{track}$, with the tracker one, where σ_{track} is the p_T resolution as measured by the tracking system. Finally, all the blocks belonging to the muons are removed from the input list.

[1] Details of the iterative tracking procedure and the particle-flow calorimeter clustering, for both ECAL and HCAL (depth 1 and 2), are skipped since considered out of the scope of this overview on physics object.

- **Particle-flow electrons**: each track, belonging to a block, is submitted to a pre-identification stage, which recognizes possible electron candidates by looking for an ECAL cluster inside the block. These tracks are then re-fitted via the GSF algorithm, resulting in an electron candidate and the related blocks are removed before the next step.
- **Charged hadrons**: the remaining tracks are filtered, requiring σ_{p_T}/p_T to be smaller than σ_{E_T}/E_T expected for a charged hadron, giving rise to a particle-flow charged hadron. By default, tracker momentum and calorimetric energy are built under the pion mass hypothesis. If the energy of the linked calorimeter clusters is compatible with the track p_T, within uncertainties, the momentum estimate is obtained by a combination of these measurements through a χ^2 fit.
- **Neutral hadrons and photons**: in most cases, the calorimeter energy exceed the momentum one in a significant way. A PF photon or a PF neutral hadron is then added to the final particle collection depending whether the excess is located in ECAL or HCAL blocks. Clusters not linked to any track give rise to PF photon or neutral hadron candidates.

2.4.4 Jet reconstruction and calibration

Starting from a hard scattering process, quarks and gluons produced by the interaction have a defined color state, according to the QCD description of strong interactions. This represents an un-physical system since, in nature, quarks and gluons are confined in colourless particles called hadrons. At detector level, only the products of their fragmentation, simulated by the parton shower algorithms, can be observed. This exhibits as a jet of collimated particles which reflects, most of the time, the energy and the flight direction of the initial parton.
Therefore, a jet is a cluster of charged particle tracks and calorimetric energy deposits, in a defined local region of the detector. The properties of a jet depend on the algorithm used for its definition, and the clustering strategy has to provide finite and robust theoretical predictions. In particular, it must be collinear safe, i.e. jet clustering and properties shall be independent from collinear gluon splittings, and soft emission safe, i.e. jet clustering and properties shall not be sensitive to soft particles added to the final state.
In CMS, the anti-k_t clustering algorithm [81] is adopted for "standard" jet reconstruction, since it meets all the conditions previously exposed. Starting from all the reconstructed particles of a given event, provided by the PF algorithm, considering i and j as generic particle indices, two distances are defined:

$$d_i = p_{T,i}^{2p} \quad ; \quad d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad ; \quad \text{anti-}k_t \rightarrow p = 2$$

where ΔR_{ij}^2 is the distance between i and j defined by Equation 2.4, while R is a parameter that indicates the typical size of a jet in the η-φ plane (jet radius). The anti-k_t algorithm works as follows:

- Fixing i as proto-jet, if there is another particle j, with $i \neq j$, such that $d_{ij} < d_i$, j is merged with i into a new proto-jet.
- If $\nexists$ a j particle such that $d_{ij} < d_i$, i is promoted as jet.
- The iterative procedure stops when all the particles are associated to a jet.

The high instantaneous luminosity at the LHC results in multiple pp collisions (pileup) taking place within a single bunch-crossing. These additional interactions, occurring within the same bunch-crossing of the hard scattering, produce additional tracks in the tracker and energy deposits in the calorimeters. This contribution is usually called in-time pileup.
Moreover, collisions occurring in previous and subsequent bunch crossings also contribute to the calorimetric energy measurement, due to an overlap between electronic signals, as a consequence of a finite signal decay time in the calorimeters. This term is called out-of-time pileup. The out-of-time contribution is mitigated at the level of signal processing, while the in-time one is partially removed using tracking information. In fact, once a criterion is chosen to identify the primary vertex of an event, i.e the one associated to the hard scattering, charged tracks not pointing to this vertex are subtracted before running the jet clustering algorithm. This procedure is called charged hadron subtraction (CHS) and represents the reference benchmark for Run-I jet reconstruction. In addition to CHS, other techniques have been tested to reduced both pileup jet occupancy and residual effect of pileup particles in jets produced by the hard scattering [82, 83].
Unfortunately, there are many possible sources of residual biases in the jet energy reconstruction, such as out of cone radiation, non-linear calorimeter response, detector segmentation, electronic noise, physics noise (pileup and underling event), quark-gluon difference in the fragmentation etc... For these reasons, jet energy scale and resolution corrections become crucial to obtain an unbiased jet energy response [84]:

- **Pileup correction:** it is applied to subtract residual contributions from neutral pileup particles, overlapping inside the jet cone, and residual electronic noise.
  The electronic noise contribution is measured after zero-suppression acquisition, running on zero-bias events, while the pileup one is measured in simulated events, comparing the reconstructed kinematics of the same events generated with and without pileup contaminations. This correction is usually parametrized as a function of the pileup energy density (ρ), the jet area (A) (defined in [85]), jet p_T and η.
  In Figure 2.17 (left), the pileup offset, defined as the mean value of the difference between the p_T of a reconstructed CHS jet in events with and without pileup contamination, is shown as a function of the matched generator level jet p_T and for different bins of the mean number (μ) of the additional generated pileup interactions[2]. Furthermore, always in Figure 2.17, the pileup correction factor for a $\langle\mu\rangle = 20$ is shown as a function of the jet p_T (middle) and η (right).

- **True response:** it aims at correcting for non-uniformities in the different CMS subdetectors, equalizing the reconstructed jet response to the generated one using simulated events only. Response correction is both η and p_T dependent:

$$C_{truth}(p_T, \eta) = \frac{\langle p_T^{reco} \rangle}{\langle p_T^{gen} \rangle}(p_T^{reco}, \eta^{reco}) \tag{2.6}$$

  The simulated true response corrections are summarized in Figure 2.18.

- **Residual corrections**: residual scale factors, between data and simulation, are determined after correcting jets for pileup offset and true response. At first, an additional

[2] In this context, generator level jets are obtained clustering all the stable particles produced by the showering of the hard scattering, before any particle-detector interaction and pileup merging.

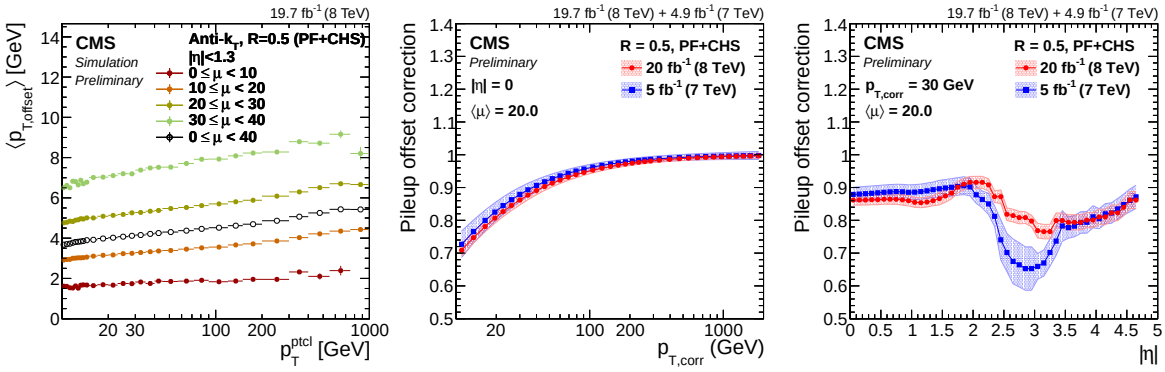

Figure 2.17: (Left) Simulated true offset $\langle p_T^{\mathrm{wPU}} - p_T^{\mathrm{noPU}} \rangle$ as a function of the generator jet p_T for $|\eta| < 1.3$. The pileup offset correction, with systematic uncertainty band, estimated for the average 2012 condition of $\langle \mu \rangle = 20$: (middle) PF-CHS $R = 0.5$ jets at $|\eta| = 0$ versus p_T, (right) same jets at $p_T = 30$ GeV as a function of η.

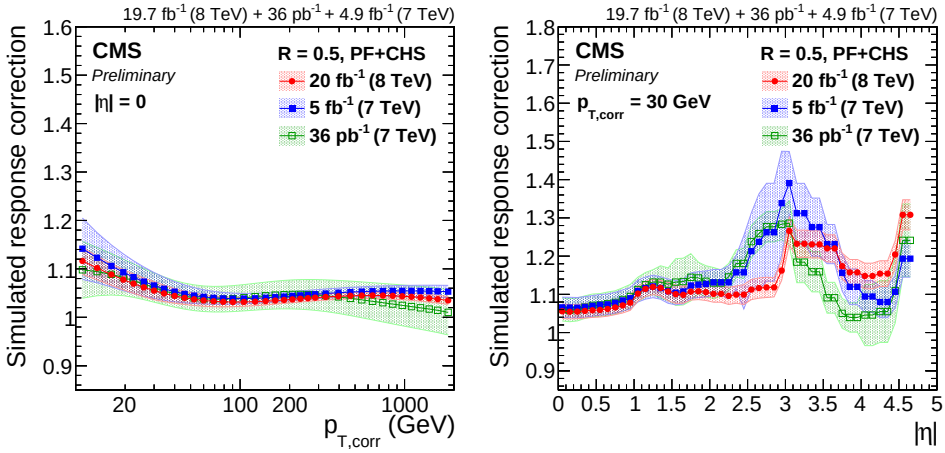

Figure 2.18: True response correction derived from simulated events of QCD interactions for PF CHS jets $R = 0.5$, compared to corrections extracted from the 7 TeV data taking. The comparison is shown for central jets as a function of p_T (left), and for $p_T = 30$ GeV versus η (right).

correction for the non homogeneous response along η is derived from di-jet events, in which the p_T response of a probe jet, outside the barrel region, is equalized to the one in the reference tag region ($|\eta| < 1.3$) as a function of the average p_T of the di-jet system. The result of the p_T dependent fit is shown in Figure 2.19 (left).

Then, an absolute scale correction is derived inside the reference barrel region ($|\eta| < 1.3$). It relies on a comparison between the transverse momentum of a central jet with the one of a well reconstructed object, such as pairs of muons or electrons from a Z-boson decay or hard photons in γ+jets events. In Figure 2.19 (middle), data-to-MC scale factors for the absolute jet response are shown, as a function of jet p_T, for the events used in the analysis. Eventually, a simultaneous fit of these samples is used to extract the ultimate estimate of the absolute jet energy scale, taking into account systematic uncertainties as nuisance parameters. Results are shown in Figure 2.19 (right). The whole procedure, adopted to obtain calibrated jets in both data and simulation, is summarized by:

$$p_T^{\mathrm{cor}} = p_T^{\mathrm{raw}} \times C_{\mathrm{pup}}(p_T^{\mathrm{gen}}, \rho, A) \times C_{\mathrm{truth}}(\eta, C_{\mathrm{pup}} \cdot p_T^{\mathrm{raw}}) \times C_{\mathrm{res}}(\eta, C_{\mathrm{truth}} \cdot C_{\mathrm{pup}} \cdot p_T^{\mathrm{raw}}) \quad (2.7)$$

where C_{pup} stands for the pileup correction, C_{truth} is the inverse of the true response and C_{res} is the global residual correction, that should be applied only on data.

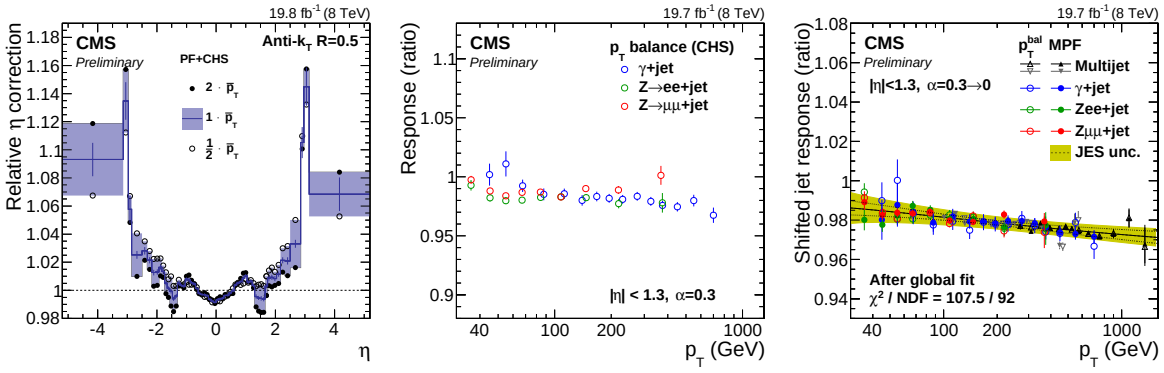

Figure 2.19: (Left) Relative η-dependent corrections, where in each η-bin the correction is fitted with a log-linear average p_T dependence : $p_0 + p_1 \cdot \log(p_T)$. (Middle) Comparison of the absolute jet response obtained from from $Z \to \mu\mu$, $Z \to ee$ and γ+jets events. (Right) Global fit of the data-to MC absolute scale corrections.

2.4.5 Missing transverse energy $\not\!\!E_T$

The missing transverse energy ($\not\!\!E_T$) is reconstructed through the PF algorithm, defined as the negative vector sum of the transverse energy of all PF candidates in a given event [86]:

$$\vec{\not\!\!E}_T^{\,\text{PF}} = -\sum_{\text{i}}^{\text{N}_{\text{PF}}} \vec{E}_T^{\,i} \tag{2.8}$$

The missing transverse energy is the common physics quantity, defined at hadron colliders, used as a signature for invisible particles like neutrinos. It is susceptible to detector effects, such as reconstruction inefficiency, geometric acceptance, calorimeter thresholds, non-linear responses and finite energy resolution. The PF $\not\!\!E_T$ bias is significantly reduced after jet energy calibration, as described in Section 2.4.4. This is achieved through the following formula:

$$\vec{\not\!\!E}_T^{\,\text{corr}} = \vec{\not\!\!E}_T^{\,\text{PF}} - \vec{\Delta}_{\text{jets}} = \vec{\not\!\!E}_T^{\,\text{PF}} - \sum_{\text{i}}^{\text{jets}} ([p_T^{\text{corr}}]_{\text{i}} - [p_T]_{\text{i}}) \tag{2.9}$$

where the sum is performed over all the corrected jets with $p_T > 10$ GeV. This propagation of jet energy calibration to the missing energy is also known as Type-I correction. In addition, other specific $\vec{\not\!\!E}_T$ corrections can be considered:

- **Pileup correction:** assuming an isotropic distribution of the particles produced by a pileup interaction, the pileup contribution to the missing energy measurement is expected to be close to zero. Therefore, the measurement of charged particles from pileup vertices ($\vec{v} = \sum_{\text{charged}}^{\text{PU}} \vec{p}_T$) should be well balanced by the neutral one. This constraint is used to induce the contamination of neutral pileup particles to the $\not\!\!E_T$:

$$\vec{\not\!\!E}_T^{\,\text{corr}} = \vec{\not\!\!E}_T^{\,\text{PF}} - \vec{\Delta}_{\text{PU}} = \vec{\not\!\!E}_T^{\,\text{PF}} - \sum_{\text{i}}^{\text{PU}} f(\vec{v})\vec{v} \tag{2.10}$$

where the correction function $f(\vec{v})$ is parametrized as $c_1(1+\mathrm{erf}(-c_2|\vec{v}|^{c_3}))$, where c_1, c_2 and c_3 are extracted from simulated minimum bias events.

- **φ-modulation**: although all the particles are distributed uniformly along φ, some asymmetry is observed in the PF candidate $\vec{p}_T$ sum, in both data and simulated events, due mainly to a not perfect detector alignment and to a shift of the nominal interaction point. A correction on x and y components of the missing energy has been designed to cure this effect, tuned as a function of the reconstructed number of vertices:

$$\not{E}_x^{\mathrm{corr}} = \not{E}_x - (c_{x_0} + c_{x_s} N_{\mathrm{vtx}}) \qquad \not{E}_y^{\mathrm{corr}} = \not{E}_y - (c_{y_0} + c_{y_s} N_{\mathrm{vtx}}) \tag{2.11}$$

Other more sophisticated missing energy determinations, in particular devoted to improve the resolution, like the `No-Pileup MET` or the `MVA-MET` [86], will not be discussed in this Section since they are not used later in the thesis.
In Figure 2.20, the PF missing energy distribution, obtained after applying the `Type-I`, pileup and φ corrections, is shown for selected $Z \to \mu\mu$ events. Since the presence of a well measured boson defines a unique axis ($\vec{q}_T$) in each event, the hadronic recoil ($\vec{u}_T$) is decomposed into an orthogonal ($u_\perp$) and a parallel ($u_{||}$) components. The scalar quantity $-\langle u_{||} \rangle / q_T$ is referred to as $\vec{\not{E}}_T$ response, while the width of $u_{||} + q_T$ and $u_\perp$ are used as estimates of the missing transverse energy resolution. The response curves ($-\langle u_{||} \rangle / q_T$), extracted from the data as a function of the vector boson boost q_T, are shown in Figure 2.20 (middle), where deviations from unity indicate a bias in the hadronic recoil determination, which is well recovered for $q_T > 40$ GeV. Finally, in Figure 2.20 (right), the distribution of $\sigma(u_{||})$ as a function of q_T is reported. The resolution linearly increases with the Z-boson boost, while data and simulation remain in good agreement in each control sample.

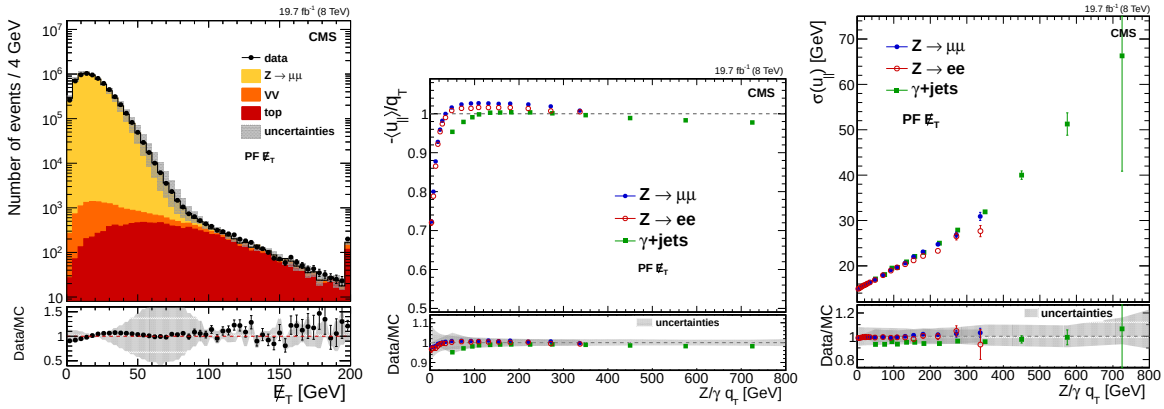

Figure 2.20: (Left) PF $|\vec{\not{E}}_T|$ distribution in $Z \to \mu\mu$ events where they grey band represents the systematic uncertainty on the simulation expectation. (Middle) Response curves for $\vec{\not{E}}_T$ in events with a Z-boson or a prompt photon (γ+jets). The upper frame shows the response in data, while the lower one shows the ratio between data and simulation. (Right) Parallel recoil resolution $\sigma(u_{||})$ related to the PF missing energy estimate. Again, the upper frame shows the resolution in data, while the lower one shows the ratio between data and simulation.

Chapter 3

Identification of highly boosted hadronically decaying vector bosons

In searches for new physics beyond the SM at high mass, which involve objects with large transverse momenta, it is crucial to distinguish hard prompt jets, originating from the decay products of heavy particles, from the ones obtained by the fragmentation of single quarks or gluons. In final state topologies with $W^{\pm}$, Z or Higgs bosons, to achieve high signal selection efficiency, it is common to look for hadronic or semi-leptonic signatures. When the boost becomes large, above order of $p_T > 200$ GeV, the hadronization products of a $V \to qq'$ decay overlap to each other inside a single clustered jet, where traditional analysis techniques, relying on resolved jets, are no longer applicable.
In this context, the analysis of jet substructures, i.e. the internal components of a jet, plays a key role to identify these objects, arising from heavy boson decays, against quark or gluons jets usually referred to as QCD jets[1]. In fact, the characteristic two-body decay of a high p_T vector boson result in a hard substructure, which is absent in typical high p_T QCD jets. These subtle differences are resolved more clearly by removing soft QCD radiation from jets. Such adaptive modification of the jet algorithm or selective removal of soft radiation during the clustering is generally referred to as jet grooming [87, 88]. On the other hand, at pp colliders, hadronic signatures are faced with the relatively poorer reconstruction of jets and large multi-jet background from QCD processes. Thus, an efficient identification of hadronic V-boson decays through jet substructure observables allows to profit from the substantially larger branching fraction of hadronic channels, providing significant gains in searches for new physics.
Measurements of substructure observables, related to identification of W-boson jets, have been already performed with 7 TeV data by both CMS [89, 90] and ATLAS [91, 92] collaboration.
This Chapter is organized as follows: in Section 3.1, a general overview on the algorithms adopted for V-jet identification is reported, then, in Section 3.2, performance studies on these jets, looking at 8 TeV data collected by the CMS experiment, are described [93].
Furthermore, in Section 3.3, the expected performance on V-jet tagging is investigated in light of the 13 TeV data taking (Run-II) of the LHC operations [83, 94].

[1] Since the masses of W and Z-bosons are rather close to each other and the experimental resolution on large cone jets is not optimal, these two resonances cannot be perfectly distinguished. Thus, it is common to refer to them as V-jets, where V indicates a W or a Z-boson.

Eventually, Section 3.4 describes the commissioning of the PileUp Per Particle Identification (PUPPI) algorithm performance with the first Run-II data, used to mitigate pileup contaminations on large cone jets and missing transverse energy.

3.1 Algorithms for V-jet identification

Since the mass of a vector boson (W,Z) is larger than the one of a typical quark-gluon (q/g) jet, the jet mass itself represents the primary observable to distinguish a heavy hadronic resonance from the hadronization products of a single parton. The bulk of the V-jet mass arises from the kinematics of the two jet cores, corresponding to the fragmentation of the two quarks produced in the decay, while the one of a QCD jet originates mostly from soft gluon radiation. Grooming methods are commonly adopted to improve the discrimination by removing the softer radiation, shifting the jet mass of QCD jets to smaller values, while maintaining the one for V-jets close to the expectation. Therefore, grooming algorithms change the jet constituent composition without affecting the overall jet production cross section, giving a infrared collinear (IRC) safe result. Starting from a large cone jet that contains the whole decay products of a vector boson, typically clustered with the anti-k_t (AK) or the Cambridge-Aachen (CA) [95] algorithm, the following grooming techniques have been considered:

- **Filtering**: the constituents of the input jet are reclustered with the CA algorithm, using a smaller distance parameter (R_{filt}) and keeping only the first N_{filt} leading subjets[2] [87].

- **Trimming**: the input jet is reclustered via the k_T algorithm [81], adopting a smaller radius R_{trim} and keeping all the hard subjets passing a transverse momentum selection, defined as: $p_T^{subjet}/p_T^{jet} > p_T^{frac}$ [96].

- **Pruning**: it is based on a CA reclustering of the input jet constituents, in which the iterative sequence is modified to remove soft and wide angle protojets (single particles or group of particles already combined in previous steps) [97]. At each stage of the CA algorithm, two quantities are computed:

$$z = \frac{\min(p_T^i, p_T^j)}{p_T^i + p_T^j} \quad ; \quad \Delta R_{ij} = \sqrt{\Delta\eta_{ij}^2 + \Delta\varphi_{ij}^2} \tag{3.1}$$

  The protojet with lower p_T is ignored if the hardness (z) is below a chosen threshold (z_{cut}) or if it is sufficiently distant from the harder one ($\Delta R_{ij} > \alpha \cdot m/p_T$), where m and p_T are, respectively, the mass and the transverse momentum of the original jet.

- **Soft drop**: as any other grooming method, soft drop declustering removes wide-angle soft radiation from a jet, to mitigate the contaminations from initial state radiation (ISR), underlying event (UE) and multiple hadron scattering (pileup), according to the following selection [98]:

$$\frac{\min(p_T^i, p_T^j)}{p_T^i + p_T^j} > z_{cut} \times \left(\frac{\Delta R_{ij}}{R}\right)^{\beta} \tag{3.2}$$

[2] Subjet stands for a protojet obtained by undoing the jet clustering sequence, following the inverted jet history.

where i and j are two protojets, R is the input jet radius, while z_{cut} and β, respectively called hardness threshold and angular exponent, rule the behavior of the algorithm. Differently from other groomers, soft drop removes protojets during the declustering stage, following the inverse clustering history of the input jet. When $\beta > 0$, soft drop removes soft radiation while maintaining a fraction of the soft-collinear one, acting as a groomer. In contrast, when $\beta < 0$, it removes both soft and collinear contributions behaving as a "tagger", vetoing jets that do not have two well separated hard prongs. The boundary point, when $\beta = 0$, is called modified mass drop tagger (`MMDT`) [99], where soft drop behaves like a tagger at any fixed-order in the α_s expansion, giving a "Sudakov safe" groomer [100].

Besides groomed masses, additional discrimination between V-bosons and `QCD` jets is obtained from a more extensive use of jet substructure, considering jet shape based observables:

- **Mass drop** (μ) is defined as the ratio of the highest mass pruned subjet (m_1), obtained by undoing the last iteration of the `CA` clustering after pruning, and the total pruned jet mass (m_{pruned}) [87]. Since a V-jet is formed by merging the showers of two decay quarks, the mass of each subjet is much smaller than the whole jet one. In contrast, a massive `QCD` jet is formed through continuous soft radiation, thus the subjet with larger mass contains the bulk of the jet constituents, pushing the mass drop close to unity. The separation power of this observable depends mainly on the ability to resolve subjets inside a large cone one, thus is observed to be strongly dependent on the boost of the system.

- **N-subjetttiness** (τ_{N}) is a generalized jet shape observable, computed under the assumption that a jet is composed by N subjets. It is the p_T-weighted ΔR distance between each jet constituent and its nearest subjet axis [101, 102], defined as:

$$\tau_{\text{N}} = \frac{1}{d_0} \sum_k p_{T,k} \cdot \min(\Delta R_{1,k}, \Delta R_{2,k} ... \Delta R_{N,k}) \tag{3.3}$$

  where d_0 is a normalization factor, defined as $d_0 = \sum_k p_{T,k} \cdot R_0$, while R_0 is the original jet distance parameter. The τ_{N} observable has a small value if the jet is consistent with having N or fewer subjets, as almost every jet constituent will be close in ΔR to its own true subjet. Assuming N subjets inside the jet, their related axes are obtained by re-clustering the input jet with the exclusive k_T algorithm [103], reversing the last N clustering steps. The axes choice is usually optimized to minimize the N-subjettiness value, increasing the related separation power. To discriminate between a V-jet, which is expected to have two prong subjets, and a `QCD` one, usually consistent with a single subjet, the ratio τ_2/τ_1 is particularly useful as it tends to maximize the separation.

- **Q-jet volatility** (Γ_{Qjet}) is a statistical measure of a sample of similar jet clustering sequences [104]. A jet is defined by its own cluster sequence, which is topologically equivalent to a recombination tree. During the clustering stage, the protojet recombination scheme is randomized and the pruning algorithm is run for each jet tree. In this way, a statistical set of trees for each jet is obtained and used to compute a jet mass distribution. The continuous soft radiation that forms massive `QCD` jets is more susceptible to

fluctuations, since a small deviation in the soft radiation can result in a very different order of assembling the jet. In contrast, V-jets are characterized by two strong jet cores and small perturbations usually yield nearly identical clustering sequences. Therefore, a large volatility is a characteristic of QCD jets and can be used to distinguish them from signal V-jets.
The algorithm works as follow: fixing a chosen number of alternative recombinations (N_{tree}), at each step of the clustering a weight ω_{ij} is assigned to each constituent pair according to:

$$\omega_{ij} = \exp\Big(-\alpha \cdot \frac{d_{ij} - d_{min}}{d_{min}} \Big) \tag{3.4}$$

where $d_{ij} = \Delta R_{ij}^2$, while d_{min} is its minimum over all pairs involved in this procedure and α controls the level of randomness. A random number is used to select the constituent pair to be merged, according to their probabilities $\omega_{ij} / \sum_{ij} \omega_{ij}$. Finally, the Q-jet volatility is defined by the ratio between the variance and the mean value of the pruned mass distribution obtained from the N_{tree} recombinations of each jet:

$$\Gamma_{\text{Qjet}} = \frac{\sigma(m_{\text{pruned}})}{\langle m_{\text{pruned}} \rangle} \tag{3.5}$$

- **Energy correlation functions** are defined in [105] and are usually combined into dimensionless double ratios (C_2^β) as:

$$C_2^\beta = \frac{\sum_{i<j<k} \; p_{T,i} \; p_{T,j} \; p_{T,k} \; (R_{ij} R_{ik} R_{jk})^\beta \; \sum_i p_{T,i}}{[\, \sum_{i<j} \; p_{T,i} \; p_{T,j} \; R_{ij}^\beta \,]^2} \tag{3.6}$$

  where i, j and k run over all the jet constituents, satisfying $i < j < k$. Similarly to τ_2/τ_1, the numerator quantifies how likely a jet is composed of two subjets, while the denominator gives a probability for being composed of a single one.

- **Jet charge** is a measure of the electric charge of the parton that originates the jet [106], originally introduced for flavor tagging of neutral B-mesons. It is defined as the p_T-weighted average charge of the jet constituents:

$$Q^k = \frac{\sum_i q_i \cdot p_{T,i}^k}{p_{T,\text{jet}}^k} \tag{3.7}$$

  where i refers to all particles in a jet. The parameter k is used to enhance the discrimination among quarks, gluons and V-jets.

- **Quark-Gluon Likelihood** (QGL) has been introduced by the CMS collaboration in [107] and it is a likelihood-based discriminator designed to distinguish between jets originating from quarks and gluons. It is applied on both groomed jets and their leading subjets.

- **Planar flow**, with $R = 0.5$, and **trimmed grooming sensitivity** have also been considered in the following performance studies. Planar flow characterizes the geometric distribution of the energy deposition from a jet, where QCD jets are expected to be more isotropic than V-jets. Trimmed grooming sensitivity is defined as the ratio between the trimmed and the original jet mass, when the trimming algorithm is applied to a V-jet candidate.

3.2 V-tagging performance with 8 TeV data

To precisely investigate V-jets identification performances, a number of different event topologies have been considered, providing inputs for a large class of physics analyses. In particular, semi-leptonic $t\bar{t}$ events ($t\bar{t} \to \bar{b}\mathrm{W}^- \, b\mathrm{W}^+ \to \ell^{\pm}\nu + \text{jets}$) provide a relative pure sample of W-jets, which is used to compare the V-tagging efficiency, mass scale and resolution between data and simulation. In contrast, the semi-leptonic W+jets topology and the inclusive di-jet one are used as source of high p_T QCD jets, to study the fake rate model in both data and simulated events. These are also useful benchmarks for searches where the SM W+jets and multi-jet productions are leading backgrounds. The W+jets sample allows to access, with high statistics, the low boosted regime, while the di-jet one is used to study q/g jets with a transverse momentum up to the TeV scale. In Figure 3.1, the fraction of the leading p_T jets, originating from the hadronization of single quarks or gluons, is reported for W+jets events (left) and di-jet (right), as a function of transverse momentum. The W+jets control region is largely dominated by light flavor quark jets, while multi-jet composition switches from gluon to quark enriched going moving from low to high jet p_T regime.

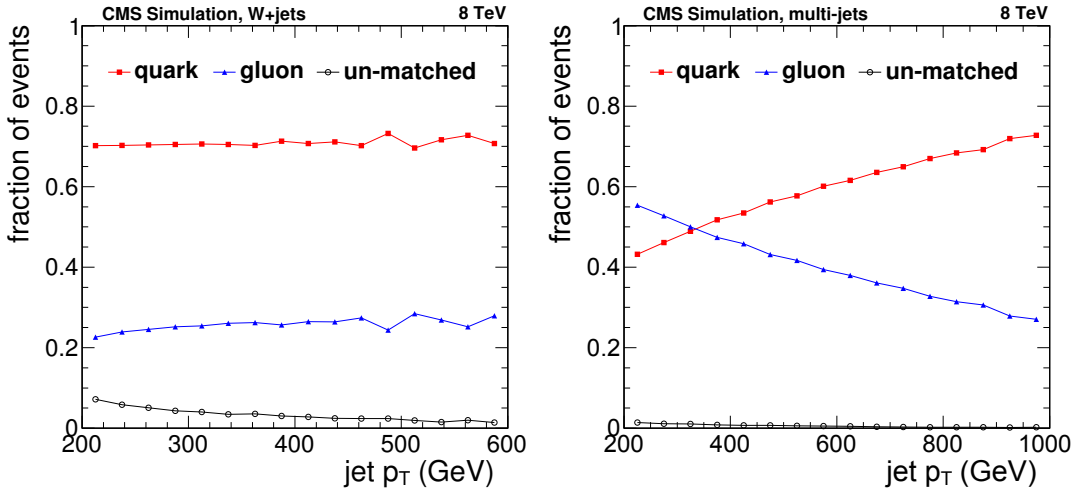

Figure 3.1: Leading p_T jet fraction in W+jets events (left) and multi-jet (right), irradiated by a quark or a gluon as a function of the jet p_T. In a small fraction of events, the leading jet is not matched to a parton produced by the hard scattering within a cone of $\Delta R < 0.3$.

3.2.1 Data and simulated events

Data were collected with the CMS detector from pp collision at a center-of-mass energy of $\sqrt{s} = 8$ TeV, thanks to a set of single lepton and jet activity triggers accumulating an integrated luminosity of 19.7 ± 0.5 fb^{-1} [108]. The selected single lepton triggers require the presence of at least one identified lepton with large transverse momentum, $p_T > 40$ (80) GeV for μ (e) final state, while multi-jet triggers are based on requirements on the transverse momentum scalar sum (H_{T}) of all the reconstructed jets in the event and on the invariant mass of the two leading ones (m_{jj}).
As a source of simulated V-jets, a generic heavy resonance decaying into a pair of longitudinally polarized W-bosons is considered. These events are simulated considering either a

Bulk graviton, predicted in WED models, or a SM-like high mass Higgs boson. The former are generated via JHUGen 3.1.8 [109], while the latter by POWHEG 1.0 [110, 111], both interfaced with Pythia 6 [112] with Tune Z2* [113] for parton showering and underlying event simulation. In addition, to study the effect of vector boson polarization on jet substructures, a BSM model with SM-like Higgs boson couplings is compared with a pseudoscalar resonance, which originates only transversely polarized bosons in the decay.
The background is modeled considering QCD multi-jet, inclusive semi-leptonic W+jets, Drell-Yan pairs ($Z/\gamma^* \to \ell\ell$), $t\bar{t}$, single-top and di-boson (WW/WZ/ZZ) productions. To estimate the agreement between different simulation models and the data, three alternative multi-jet samples are considered: MadGraph v5.1.3.30 [114] showered with Pythia 6, Herwig++ [114] with Tune 23 and Pythia 8 [115] with Tune 4C [116]. MadGraph and Pythia samples use CTEQ61L [117] parton densities, while Herwig++ adopts MRST2001 [118].
Moreover, two different W+jets samples are generated via MadGraph, interfaced with Pythia 6, and Herwig++. A similar choice is adopted also for the $t\bar{t}$ generation, where a POWHEG+Pythia 6 sample, using CT10 as PDF [119], is compared to MC@NLO [120] evolved with Herwig++, using CTEQ6M PDF [117]. Finally, single-top quark production is simulated with POWHEG, while the di-boson samples are obtained from Pythia 6.
A summary of the background samples used in this study and the related production cross sections are reported in Table 3.1. Generated events are then processed through a GEANT4 [121] based simulation of the CMS detector, adding an average of 22 pileup interactions, to match the additional particle production observed in data.

Process	Generator	Cross section (pb)
QCD	MadGraph + Pythia 6	–
QCD	Herwig++	–
QCD	Pythia 8	–
W+jets ($p_T > 100$ GeV)	MadGraph + Pythia 6	228.9
W+jets ($p_T > 180$ GeV)	MadGraph + Pythia 6	26.4
W+jets ($p_T > 100$ GeV)	Herwig++	237.2
DY+jets ($m_{\ell\ll} > 50$ GeV	MadGraph + Pythia 6	3503
$t\bar{t}$	POWHEG + Pythia 6	245.8 (NNLO)
$t\bar{t}$	MC@NLO + Herwig++	245.8 (NNLO)
WW	Pythia 6	54.8
WZ	Pythia 6	33.2
ZZ	Pythia 6	8.1
single-top: s-channel	POWHEG + Pythia 6	87.1
single-top: t-channel	POWHEG + Pythia 6	5.5
single-top: tW-channel	POWHEG + Pythia 6	22.4

Table 3.1: Summary of the simulated samples considered in the analysis, where no cross section is reported for the background samples that will be normalized to data.

3.2.2 Event selection

Di-jet and W+jets topologies are selected to investigate background QCD jets inside the kinematic regime typically considered in searches for new phenomena at high mass, while the semi-leptonic $t\bar{t}$ signature represents the only pure source of boosted W-jets in data. In these topologies, one W-jet candidate is selected for each event as the highest p_T jet, reconstructed via the CA algorithm ($R = 0.8$) implemented in the FastJet package [122], requiring $|\eta| < 2.4$ so that the jet core falls within the tracker acceptance.
A large distance parameter (R) increases the efficiency to reconstruct a poorly boosted W-boson as a single jet, since the average distance between its decay products is inversely proportional to the boson p_T, i.e. $\Delta R(qq') = 2m_{\mathrm{V}}/p_T$.
The chosen value of R provides high reconstruction efficiency for W-bosons characterized by a small boost, ensuring that no performance is lost in the transition between a resolved analysis, in which two independent jets are associated to a $\mathrm{V} \to qq'$ decay, and its reconstruction as a single large cone jet at high transverse momentum. In addition, $R = 0.8$ allows a to collect a sufficiently large sample of W-jets in $t\bar{t}$ data for validating the V-tagging strategy. In fact, if the R value is too large, the b-jet produced by $t \to \mathrm{W}b$ decay tends to merge with the W-boson one, spoiling the typical substructure properties of a boosted V-jet.
Figure 3.2 compares, as a function of the vector boson p_T, the reconstruction efficiency of a hadronic W-boson from two resolved $R = 0.5$ jets to the one obtained from CA8 clustering. When the W-boson p_T is larger than 200 GeV, the CA8 choice becomes more efficient, exceeding 90% efficiency after 400 GeV.

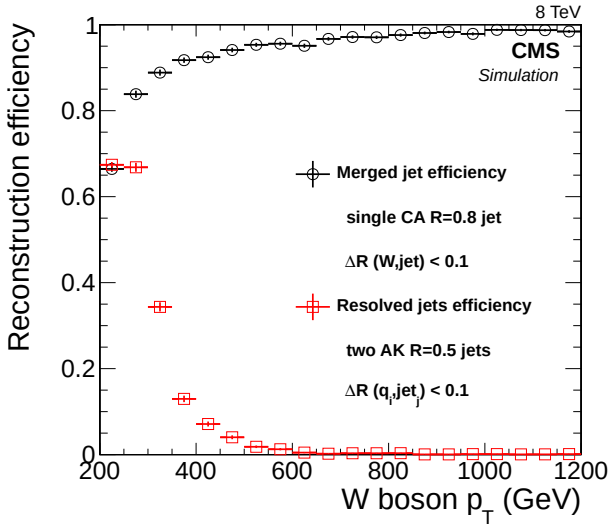

Figure 3.2: Efficiency to reconstruct a CA8 jet matched within $\Delta R < 0.1$ with a generated W-boson (black points), and the efficiency to reconstruct two AK5 jets within $\Delta R < 0.1$ of the generated quarks (red ones), obtained from hadronic decays of longitudinally polarized W-bosons, as a function of the transverse momentum.

To select pure samples of signal (V-jets) and background (QCD) jets, a set of specific requirements are applied in these three phase spaces:

- **Dijet selections:** events are initially selected requiring two AK5 jets with $p_T > 30$ GeV and $|\eta| < 2.4$, with a separation of $|\Delta\eta_{jj}| < 1.3$ and an invariant mass $m_{jj} > 890$ GeV,

following the selection criteria typically applied in fully hadronic di-boson searches for BSM physics [123]. In this set of di-jet events, W-tagging studies are performed adopting the leading CA8 jet as W-jet candidate, considering two independent p_T bins: $400 < p_T^{\text{CA8}} < 600$ GeV and $1.1 < p_T^{\text{CA8}} < 1.4$ TeV.

- **W+jets selections**: the goal is to isolate events in which a boosted leptonically decaying W-boson recoils against a high p_T jet. On the lepton side, the event should contain one muon with $p_T > 50$ GeV and $|\eta| < 2.1$ or one electron with $p_T > 90$ GeV and $|\eta| < 2.5$, reconstructed with the prescriptions given in Section 2.4.1 and Section 2.4.2, respectively. Muons are identified according to the high-p_T selection [76], which requires that each reconstructed muon must be both a tracker and a global muon, with at least one pixel hit in the fitted track, at least one matched muon station, more than eight valid tracker hits, a transverse impact parameter ($|d_{xy}|$), with respect to the selected primary vertex, smaller than 2 mm and a longitudinal one ($|d_z|$) smaller than 5 mm. Finally, the muon candidate shall be also isolated from charged hadron activity in the detector, requiring the relative sum of tracks p_T, within a cone of $\Delta R = 0.3$ around the muon track, to be:

$$\frac{I_{tk}}{p_T(\mu)} = \frac{\sum_{i \in \text{track}} p_T^i}{p_T(\mu)} < 0.1 \tag{3.8}$$

At the same time, each reconstructed GSF electron is identified by means of the HEEP criterion [124], which requires more than two hits in the fitted track, a super-cluster with E_T grater than 35 GeV, both η and φ separation between the GSF track and the super-cluster position smaller than 0.005 and 0.006 respectively, a ratio between HCAL and ECAL energies (H/E) smaller than 5%, $|d_{xy}| < 0.2$ (0.5) mm for EB (EE), to have either $E_{2\times5}/E_{5\times5} > 0.94$ or $E_{1\times5}/E_{5\times5} > 0.83$ [3], $I_{tk} < 5$ GeV and :

$$[I_{em} + I_{had} - \Delta R \times \rho] < \begin{cases} \text{EB} : & 2 \text{ GeV} + 0.003 \cdot E_T \\ \text{EE} : & 2.5 \text{ GeV} \end{cases} \tag{3.9}$$

where, I_{em} and I_{had} are the electromagnetic and the hadronic E_T sums inside the isolation cone, except for the contribution associated to the electron candidate.
Events containing additional loosely identified muons or electrons, defined relaxing isolation requirements, with $p_T > 20$ GeV are rejected to improve the purity of the W+jets sample against contaminations from Z/γ^*+jets events.
The missing transverse energy is required to be larger than 50 or 80 GeV for the muon and electron channel, respectively, to suppress the contamination from QCD multi-jet background where a jet is mis-identified as a lepton. Furthermore, the transverse momentum of the leptonically decaying vector boson, given by the sum of the lepton $\vec{p}_T$ and the missing energy $\vec{\not{E}}_T$, and of the highest p_T CA8 jet (W-jet candidate) are required to be larger than 200 GeV.
Additional criteria are applied to ensure that the W-boson and the CA8 jet are mostly back-to-back in the transverse plane: the ΔR between the lepton and the CA8 jet must be larger than $\pi/2$, the azimuthal distance $\Delta\varphi$ between $\not{E}_T$ and the W-jet must be greater than two radians, as well as the distance between the V-boson and the CA8 jet.

[3] These quantities refer to the energy sum of all the crystal within a matrix $n \times m$ around the super-cluster seed position.

A cutoff on the additional jet activity is also applied vetoing the presence of b-tagged AK5 jets, lying outside the CA8 jet cone ($\Delta R_{j,\mathrm{CA8}} > 0.8$) and identified using the CSV algorithm [125], according to a selection that provides about 70% efficiency on simulated b-jets and less than 1% of mis-tag rate. Eventually, only events with a moderate boosted W-jet candidate, $250 < p_T^{\mathrm{CA8}} < 450$ GeV, are considered.

- **$\mathrm{t\bar{t}}$ selections**: the same selections used for the W+jets topology are adopted, except for the b-veto condition which is inverted requiring at least one b-jet with $p_T > 30$ GeV. In addition, if two CA8 jets are found in the same event with $p_T > 200$ GeV, the one with highest mass is kept as the W-jet candidate and no back-to-back topology cuts are applied.

3.2.3 Substructure performance in the simulation

To understand the performance of jet substructure algorithms with the CMS detector, two dominant effects on boosted jet observables are investigated: particles from multiple overlapping pileup interactions, detector efficiency and resolution in reconstructing jet constituents. To prove the separation power of the substructure observables listed in Section 3.1, BSM resonance masses, which are decaying into pairs of boosted W-bosons, are chosen to obtain a falling shape for the W-jet candidate p_T distribution in the different analysis p_T bins, similar to one expected for QCD background jets in W+jets and multi-jet events.
Therefore, when the jet p_T is within 250-350 GeV, a SM-like Higgs boson with a mass of 600 GeV is considered, while, when p_T^{CA8} is in the range between 400 and 600 GeV in the di-jet topology, a Bulk graviton with a mass of 1 TeV is adopted.
In Figure 3.3, the pruned jet mass distributions for W-jets and QCD ones are shown comparing the generator level results with the full CMS simulation, as a function of different pileup scenarios. The convolution of both pileup and detector effects results in a broadening of the W-boson mass peak and a slight bias to higher values. The pileup dependence, comparing two scenarios with averages of 12 and 22 pileup interactions, is small thanks to the effect of the pruning algorithm. Thus, to discriminate boosted W-bosons from QCD jets background, one can select events falling in a defined window around the W-boson mass peak, as done in already published searches [126, 127]. In this study, the chosen jet mass window is fixed to range from 60 to 100 GeV. Finally, looking at the W+jets topology, there are few signal events with a pruned mass close to zero. They correspond to cases in which the WW system is decaying fully leptonically, where one lepton is reconstructed, together with UE and pileup, as a jet with few constituents failing the isolation requirement, having therefore a very low mass value since composed by almost a single high p_T object.
Figure 3.4 (left) shows the mass drop variable in the W+jets topology. As expected, it peaks at lower values for V-jets since it represents the ratio of the pruned subjet mass coming from a quark to the much larger W-boson one. In addition, Figure 3.4 (right) shows the mass drop shape considering only the events falling in the pruned mass window previously defined. The discrimination power of the mass drop is strongly reduced due to a high correlation between these two observables, particularly for QCD jets. Figure 3.5 shows the τ_2/τ_1 variable, in the W+jets topology, before (left) and after (right) the W-boson pruned mass selection. The separation power is evident and the overall performance looks better than the mass drop one, with a less pronounced correlation with the pruned mass. Furthermore, Figure 3.6 shows the

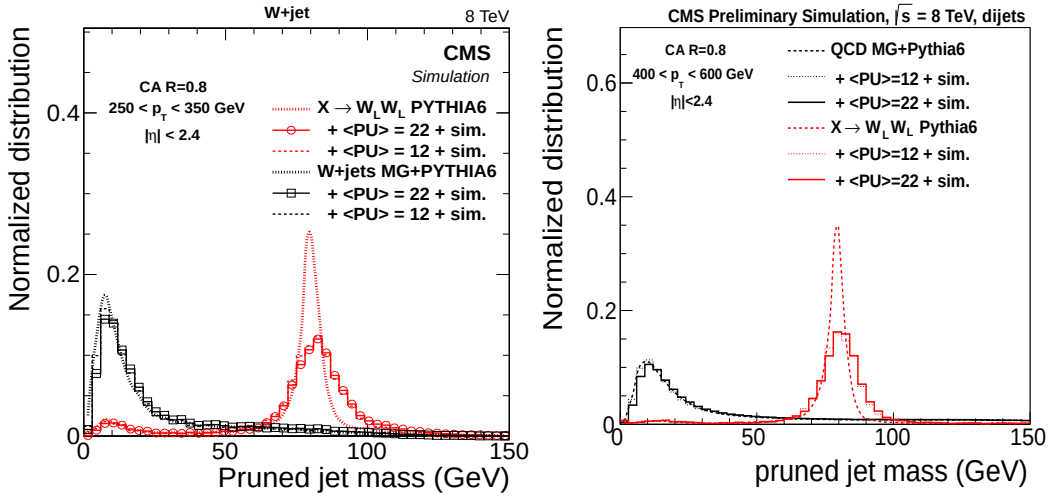

Figure 3.3: Pruned mass distribution in simulated samples of highly boosted longitudinally polarized W-bosons and QCD jets, expected in the W+jets (left) and QCD multi-jet (right) topologies. Thick dashed lines represent the generator level predictions without pileup interactions and CMS detector simulation. The continuous histograms are the expected distributions after full CMS simulation, in a pileup scenario corresponding to an average number of 12 and 22 interactions, respectively.

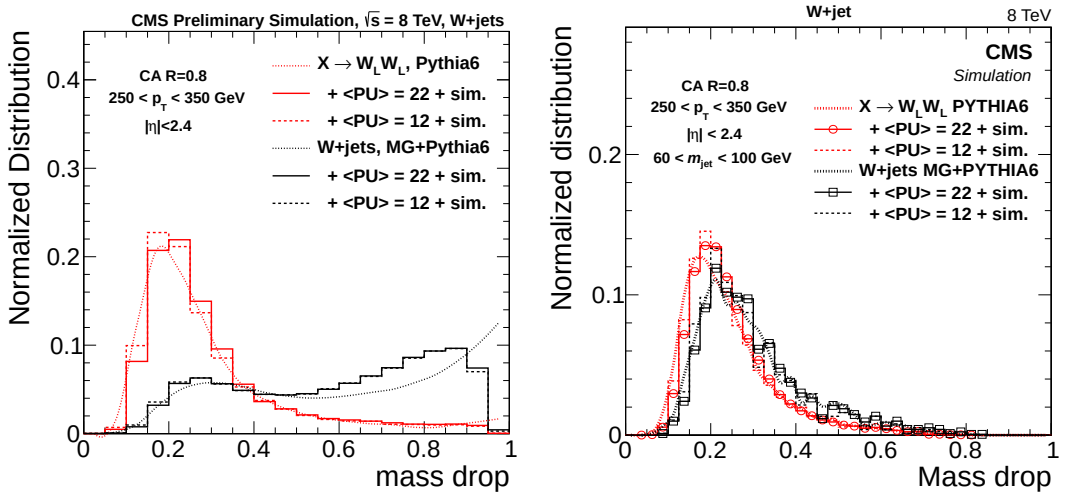

Figure 3.4: The mass drop distribution in simulated samples of boosted W-bosons and QCD jets before (left) and after (right) a selection on the pruned jet mass. Thick dashed lines represent the generator predictions without pileup interactions and CMS simulation. The continuous histograms are the expected distributions after full CMS simulation, in a pileup scenario corresponding to an average number of 12 and 22 interactions, respectively.

double ratio of energy correlation functions $C_2^{\beta=1.7}$. While the discrimination is visible at generator level, it is much reduced after pileup and detector simulation.
Eventually, Figure 3.7 (left) shows the jet charge distribution for positively and negatively charged W-boson jets, as well as the one for background QCD jets. Both pileup and detector resolution effects are quite limited on this observable, since it is defined only using charged hadron tracks from the leading vertex in the event. To assess the separation power between

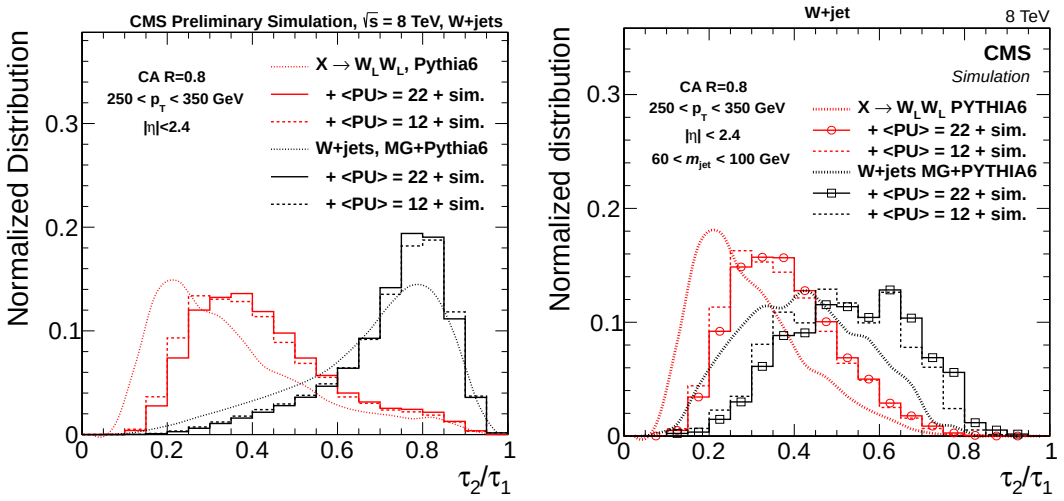

Figure 3.5: The τ_2/τ_1 observable in simulated samples of boosted W-bosons and QCD jets before (left) and after (right) a selection on the pruned jet mass, $60 < m_{\text{pruned}} < 100$ GeV. Thick dashed lines represent the generator predictions without pileup interactions and CMS simulation. The continuous histograms are the expected distributions after full CMS simulation, in pileup scenario corresponding to an average number of 12 and 22 interactions, respectively.

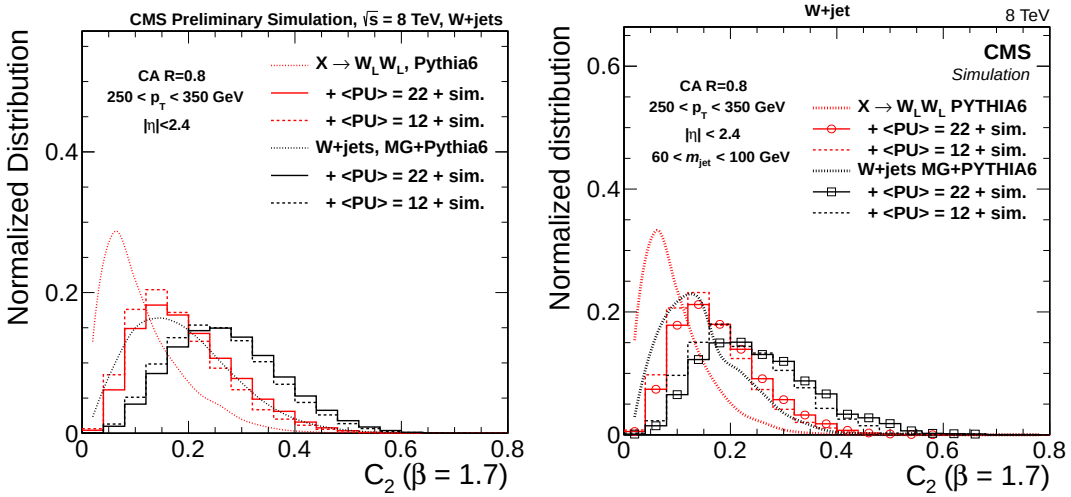

Figure 3.6: The C_2^β distribution in simulated samples of boosted W-bosons, from a heavy BSM resonance decay, and QCD jets in leptonic W+jets events before (left) and after (right) a selection on the pruned jet mass. Thick dashed lines represent the generator predictions without pileup interactions and CMS simulation. The continuous histograms are the expected distributions after full CMS simulation, in pileup scenario corresponding to an average number of 12 and 22 interactions, respectively.

W^+ and W^- via the jet charge, a hypothesis test has been performed, where the two jet charge distributions are first fitted with a Crystal-Ball function, then these are used as templates to generate pseudo-experiments (PEs). For each PE, a χ^2_+ (χ^2_-) is calculated with respect to the positive (negative) template and the $\Delta\chi^2_\pm$ is used as a test statistics. The hypothesis separation, estimated according to Gaussian quantiles, is reported in Figure 3.7 (right) as a function of the average number of events generated in each toy experiment. The separation between

two W-boson charge hypotheses is around $3\,\sigma$ for $k = 0.3$ and about 20 events per PE. Similar results are also obtained in the higher p_T bins, exploring the di-jet signature.

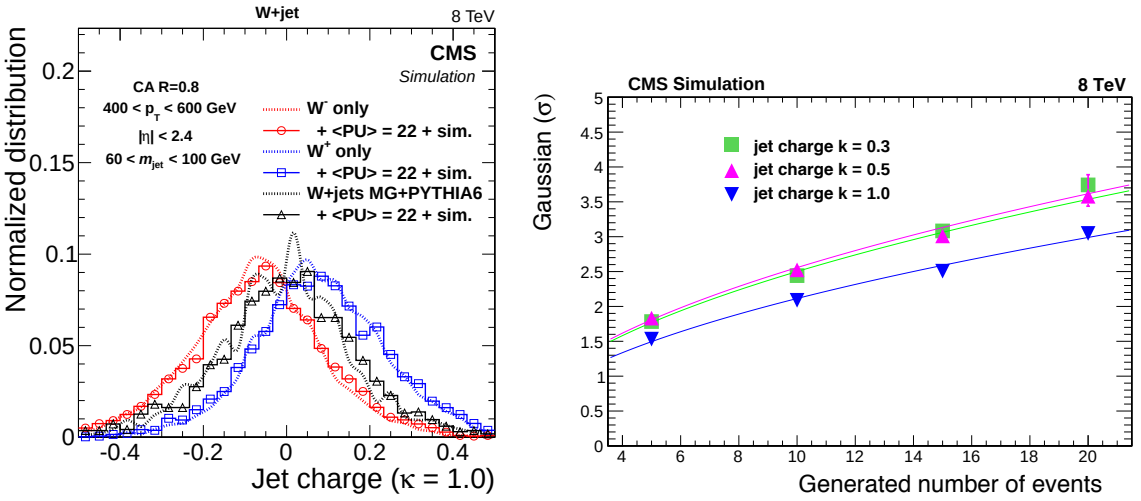

Figure 3.7: (Left) Jet charge distribution in simulated samples of boosted W-bosons, produced by a heavy resonance decay, and QCD jets in leptonic W+jets events for $k = 1.0$, after a selection on the pruned jet mass ($60 < m_{\text{pruned}} < 100$ GeV). (Right) Separation, in terms of Gaussian σ, between W^+ and W^- hypotheses, for different values of k, as function of the average number of events per PE.

3.2.4 Algorithm performances

The performance of several substructure observables used to identify V-jets are compared with the aim of establishing which provides the best signal-to-background discrimination. Since the groomed masses are the best discriminants, as shown in Section 3.2.3, the performance of other observables are tested only for jets with $60 < m_{\text{pruned}} < 100$ GeV. The adopted figure of merit is the background rejection efficiency as a function of the signal one, also known as "receiver operating characteristic" (ROC) curve.

Figure 3.8 shows the observable performances in the W+jets final state, where the W-jet candidate p_T is required to be in the range between 250-350 GeV. The performance of τ_2/τ_1, pruned τ_2/τ_1, exclusive-k_T τ_2/τ_1, Γ_{Qjet}, C_2^{β}, mass drop and jet charge are compared. For the jet charge, a positively charged lepton is required in the event, therefore the separation between negatively charged W-jets and QCD ones is displayed. The best performing variable is τ_2/τ_1 up to a signal efficiency of 75%, then Γ_{Qjet} is slightly better, while pruned and τ_2/τ_1 without axes optimization are worse. The least performing variables are the mass drop, $C_2^{\beta=1.7}$, and the jet charge.

In addition to the individual observable discrimination power, a multivariate combination of them is performed using the TMVA package [128]. This combination is performed with both a naive Bayes classifier and a Multilayer Perceptron (MLP) neural network discriminant. Additional variables with respect to those reported in Section 3.2.3 are considered, trying to increase the final discrimination power. The variables used in both discriminants are the mass drop, Γ_{Qjet}, τ_2/τ_1, C_2^{β}, the planar flow ($R = 0.5$), the number of jet constituents, ΔR between the two leading pruned subjets, trimmed sensitivity and the number of reconstructed vertices in the event. The MVA methods are trained using as signal sample a SM-like Higgs resonance, decaying into a pair of longitudinally polarized W-bosons, while the background

sample consists of the leading jet in W+jets events generated with MadGraph. Events are equally split between the training and the test sample to compute the ROC curve. Comparing the MVA results with the performance of τ_2/τ_1, about 15% improvement, at a fixed signal efficiency of 50%, is obtained. This relative small gain is limited by the large correlation between τ_2/τ_1, which is the most sensitive variable over a large range of efficiencies, and several of the other substructure observables, as reported for both signal and background events in Figure 3.9.

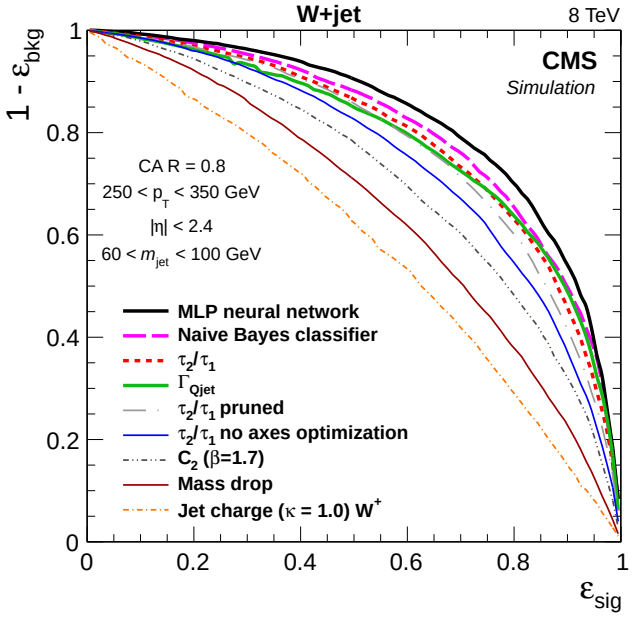

Figure 3.8: Performance of several discriminants in the background-signal efficiency plane for the lower jet p_T bin of 250–350 GeV in the W+jets topology. The signal efficiencies and mis-tagging rates of the various discriminants are estimated on samples of W-jets and QCD jets that satisfy a pruned jet mass selection of $60 < m_{\text{pruned}} < 100$ GeV.

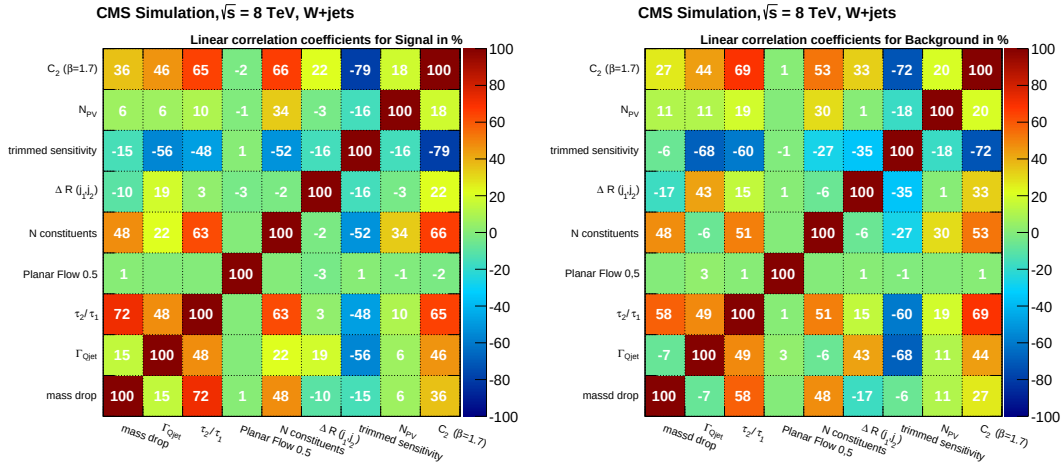

Figure 3.9: Linear correlations for the input variables considered in the MVA training: (left) correlation matrix evaluated on simulated signal jets (W-jets), (right) correlation matrix for QCD jets in a W+jets sample.

Therefore, the baseline W-tagger is built on a simple rectangular selection on pruned jet mass and τ_2/τ_1, requiring $60 < m_{pruned} < 100$ GeV and $\tau_2/\tau_1 < 0.5$, which represents the optimal value for a reference significance estimator defined as $\varepsilon_S/\sqrt{\varepsilon_B}$, where ε_S and ε_B are the signal and the background selection efficiency, respectively.
The simulated p_T and pileup dependence of the defined W-tagger has also been studied. In particular, Figure 3.10 (left) shows the efficiency of this tagger on simulated signal W-jets. At low transverse momentum, the efficiency increases thanks to the better containment, into a single CA8 jet, of the showers produced by the fragmentation of the W-boson decay quarks. Above 600 GeV, the efficiency starts to decrease, since at larger p_T the PF reconstruction degrades in resolving the jet substructure, so that the pruning algorithm removes a too large fraction of the jet mass. Furthermore, also the efficiency of the τ_2/τ_1 selection drops at high p_T, loosing around 20% signal efficiency at 1 TeV with respect to the maximum, placed for a jet p_T around 600 GeV. In Figure 3.10 (right), the signal efficiency of the pruned mass selection is reported as a function of the number of reconstructed vertices, which decreases by 6% moving from 5 to 30 reconstructed vertices, whereas the additional N-subjettiness selection efficiency drops by 12% over the same range. However, the mistagging rate of the background also decreases with pileup when the same selections are applied, yielding a similar discrimination ($\varepsilon_S/\varepsilon_B$).

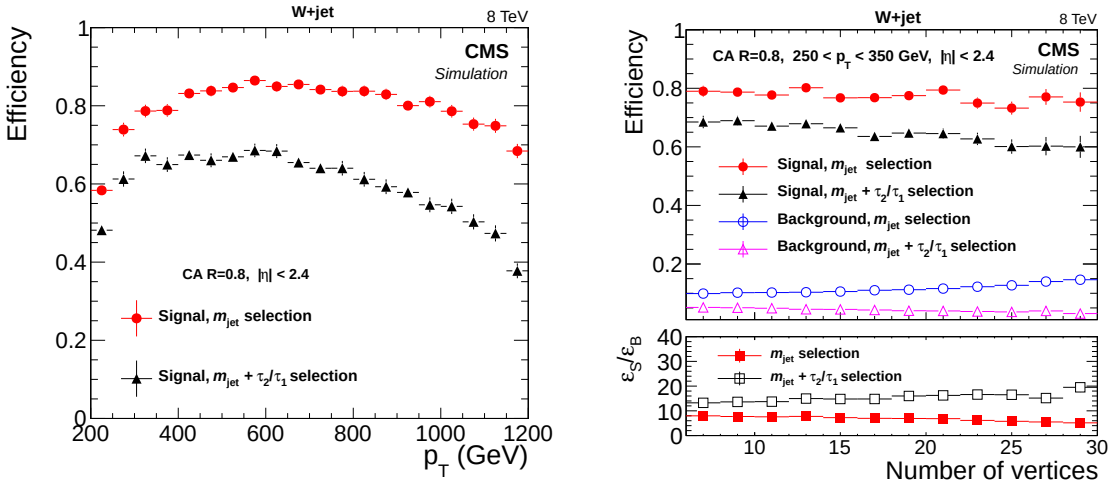

Figure 3.10: Efficiency of the pruned mass selection and its combination with $\tau_2/\tau_1 < 0.5$ on a signal sample of W-jet as a function of p_T (left) and the number of reconstructed vertices (right). In the latter case, also the mistagging rate for QCD jets, estimated from the W+jets background sample, is shown. The error bars represent the statistical uncertainty in the MC simulation.

3.2.5 Vector boson polarization

Jet substructure observables are particularly sensitive to the polarization of the decayed vector boson, which can be used as additional information to detect possible new physics scenarios, like searches for new resonances decaying into vector boson pairs, W-boson helicity at large $t\bar{t}$ masses or VBS measurements. To study the V-jet tagger potential in this field, X $\rightarrow$ WW samples are generated fixing the W-bosons to be either purely longitudinally or transversely polarized, by means of the JHUGen generator. The key observable is represented by the helicity angle of the W $\rightarrow q\bar{q}'$ system ($\cos\vartheta_j$), defined as the polar angle between the decay quark and

the W-boson flight direction in the W-boson rest frame. At parton level, where quarks are treated as final state particles, the helicity angle of the hadronic W-boson is computed and shown in Figure 3.11 (left), separately for transverse and longitudinal modes. After showering, hadronization, pileup, detector simulation and reconstruction, the same quantity is computed and reported in Figure 3.11 (right), considering the two leading subjets after pruning. Since after reconstruction it is no longer possible to distinguish between quarks and anti-quarks, i.e. to have an unambiguous subjet charge assignment, the absolute value of $\cos\vartheta_j$ is reported.
This shows how a transverse polarized vector boson decays with subjets more aligned with its direction of flight. The depletion of events at $|\cos\vartheta_j|$ around 1 is due to acceptance effects. When ϑ_j is close to zero, the two subjets overlap, preventing their reconstruction. On the other hand, when ϑ_j is around π, one subjet becomes much softer than the other one, causing a drop in the identification efficiency.
In fact, after the pruned jet mass selection, the ratio between the trailing and the leading pruned subjet transverse momenta $p_{T,2}/p_{T,1}$ is shown in Figure 3.12, for W-jets belonging to two different p_T regimes, $250 < p_T < 350$ GeV on the left and $400 < p_T < 600$ GeV on the right. Pruned subjets from transverse polarized W-bosons are less balanced in p_T with respect to the longitudinal case.

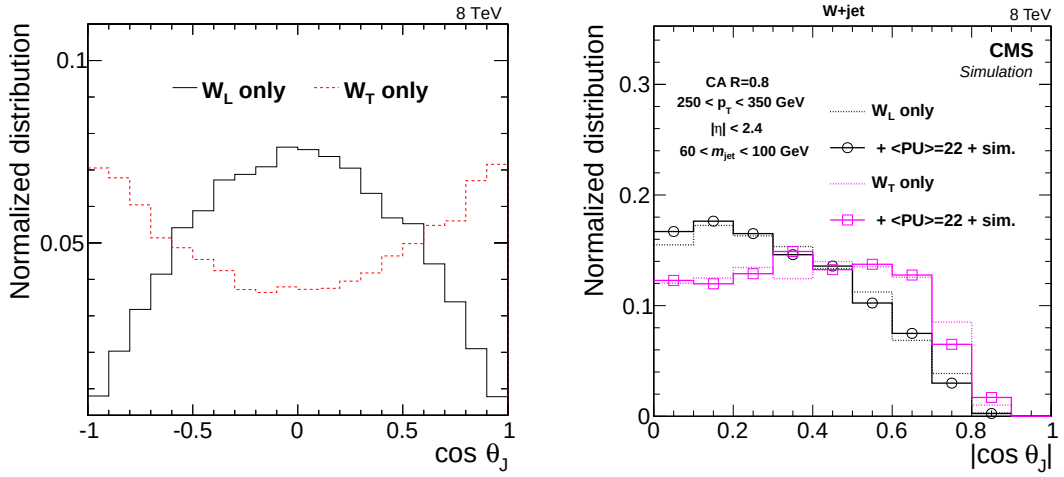

Figure 3.11: (Left) Generator level $\cos\vartheta_j$ distributions for purely longitudinally and transversely polarized W-bosons. (Right) Pruned subjet angular distributions, after requiring the pruned jet mass in the range $60 < m_{\text{pruned}} < 100$ GeV, for pure $\mathrm{W_L}$ or $\mathrm{W_T}$ and for jets with $250 < p_T < 350$ GeV.

Finally, to test if the polarization of hadronic W-bosons can be used as a discriminant in the data analyses at the LHC, the experimental resolution on $\cos\vartheta_j$ is measured through the difference between the one reconstructed from subjets and the one from parton level quarks. The experimental angular resolution is measured to be around 10 mrad in the laboratory frame, which translates into a resolution of about 65 mrad in the W-boson rest frame, which remains relatively constant over a large W-jet p_T range.

3.2.6 Systematics effects on substructure performances

In this section, different systematic effects are analyzed in terms of changes in the ROC response, looking at the di-jet topology for jets with $400 < p_T < 600$ GeV, where the Herwig++ sample is used to describe QCD jets, since it models the pruned jet mass in data better than

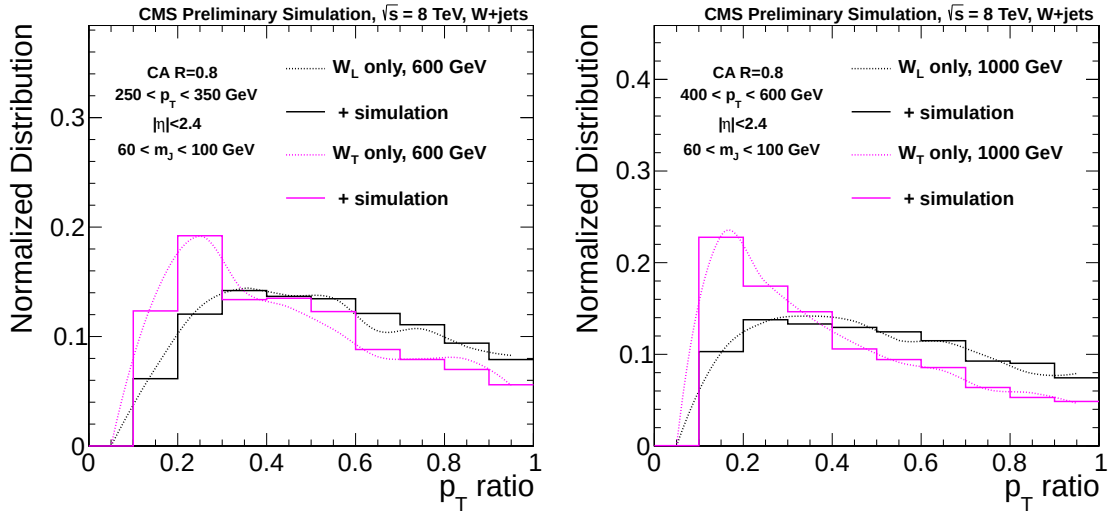

Figure 3.12: Pruned subjets p_T ratio distribution, obtained after the pruned jet mass selection, for pure W_L and W_T samples: (left) W-jets with $250 < p_T < 350$ GeV, (right) W-jets with $400 < p_T < 600$ GeV.

Pythia 6. Figure 3.13 reports a set of τ_2/τ_1 ROC curves, showing the impact of different systematic sources on the previously defined V-jet tagger. Signal efficiency and background rejection of the pruned mass requirement only, $60 < m_{\text{pruned}} < 100$ GeV, is indicated by filled circles, while its combination with $\tau_2/\tau_1 < 0.5$ by filled rectangles.
In Figure 3.13 (left), the pileup and boost dependence of the V-tagger are displayed. Moving from an average of 12 to 22 pileup interactions, no changes in the ROC response are noticed, which means that pruning and N-subjettiness are quite insensitive to pileup up to 30 reconstructed vertices. The signal efficiency versus background rejection can be kept constant adapting the working point as a function of the pileup condition, as indicated by the difference between the green and the red filled rectangles. In addition, also the generator level prediction, without pileup and detector reconstruction effects, is compared with the full CMS simulation with pileup, showing a small degradation in performance although τ_2/τ_1 shape is significantly shifted, for both real W-jets and QCD ones, to higher values by these contributions.
Furthermore, in Figure 3.13 (left), the ROC curve for jets with p_T between 0.8 and 1.2 TeV, coming from a BSM resonance decaying to WW with a rest mass of 2 TeV, is almost indistinguishable from the one derived inside the lower p_T range 400-600 GeV, except for both the position of the pruned jet mass and τ_2/τ_1 working points, which are placed at lower signal efficiency. Consequently, a fixed working point will degrade the signal efficiency with increasing jet p_T, but, since the ROC lines are overlapped, the same performance can be achieved moving the selection values. In particular, comparing the pruned mass and the N-subjettines distributions as predicted at generator level in different p_T bins, one notes that substructure performances are almost stable as a function of the boost, as indicated in Figure 3.14.
The performance loss seen at detector level is related to the degradation of the CMS particle-flow resolution in resolving jet substructure at high p_T. This brings the pruning algorithm to reject a too large fraction of particles inside a W-jet, leading to events with pruned jet mass in the range between 40-60 GeV, as shown in Figure 3.14 (left). This effect is also visible for the N-subjettiness τ_2/τ_1.
Figure 3.13 (right) compares signal-to-background discrimination of the W-tagger for pure

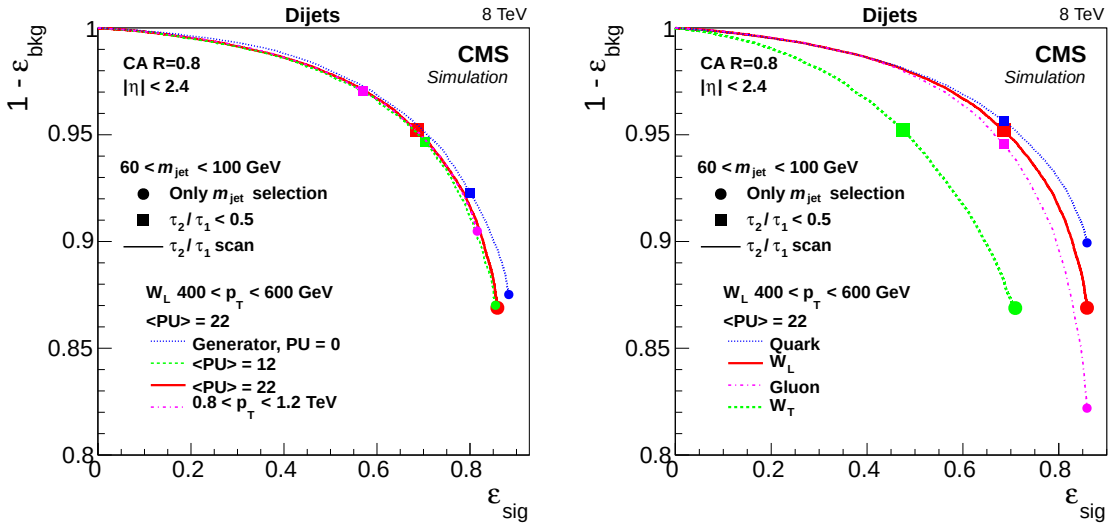

Figure 3.13: Systematic effects on the performance of the pruned jet mass and τ_2/τ_1 in the high jet p_T bin of 400–600 GeV. The performance of the pruned jet mass selection $60 < m_{\text{pruned}} < 100$ GeV, in the various scenarios, is indicated as a filled circle, while its combination with $\tau_2/\tau_1 < 0.5$ is indicated as a filled rectangle. The lines correspond to the ROC curves obtained with a selection on the N-subjettiness in addition to pruned jet mass requirement. The solid red line corresponds to the standard scenario, with an average of 22 pileup interactions and longitudinally polarized W bosons (W_L).

W_L and W_T signal samples. The pruned jet mass selection is less efficient, around 15%, on transversely polarized W-bosons, which is in agreement with both Figure 3.11 and Figure 3.12, where W-jets with $|\cos\vartheta_j| \approx 1$ are removed by the pruned mass selection due to either a large p_T unbalance between the generated quarks or a small angular separation between subjets. In addition, the ΔR seperation between pruned subjets for pure W_L bosons is smaller than in the transverse case, thus it is more likely for a low boosted W_T boson to be identified as a single CA8 jet. The N-subjettiness power is also slightly degraded for the transverse case by a smaller degree than the jet mass.
Finally, also the QCD background composition has an impact on the W-tagging performance. In fact, gluons tend to produce jets with larger masses which are less suppressed by pruning with respect to light quark jets, as shown in Figure 3.13 (right). On the contrary, τ_2/τ_1 rejects more gluon jets, since their shape is usually more similar to a single substructure than the one for quarks. In the end, a similar performance is achieved through the adopted W-tagging strategy for both gluon and quark initialized jets. In Figure 3.15, the pruned jet mass and the τ_2/τ_1 distributions, obtained from a W+jets sample showered with Pythia6, are reported separating gluon jets from light flavour quark ones.

3.2.7 Comparison of data and simulation

The comparison between data and simulation for the two observables adopted in the V-jet tagging definition, namely τ_2/τ_1 and the pruned mass, is of fundamental importance to assess the reliability of the simulated events used in the analysis, as well as to introduce calibration factors when needed. This comparison is performed in three different control regions: inclusive di-jet and W+jets for q/g jets, while semi-leptonic $t\bar{t}$ for signal W-jets. W+jets and di-jet events are compared in two different jet p_T bins of 250–350 GeV and 400–600 GeV respectively,

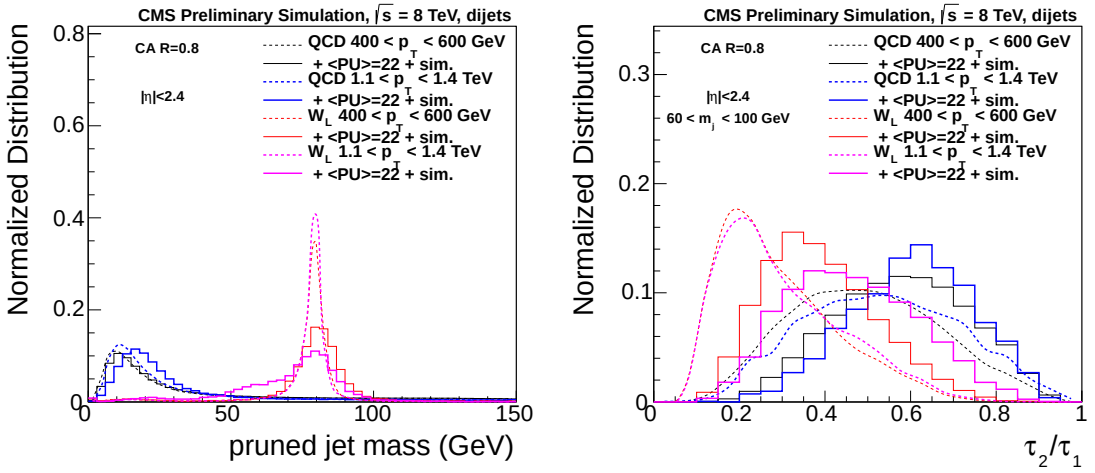

Figure 3.14: Comparison of substructure observables for both signal W-jets and background QCD jets, exploring the di-jet topology, in two jet p_T bins: $400 < p_T < 600$ GeV and $1.1 < p_T < 1.4$ TeV. (Left) pruned jet mass distribution, (Right) N-subjettiness τ_2/τ_1.

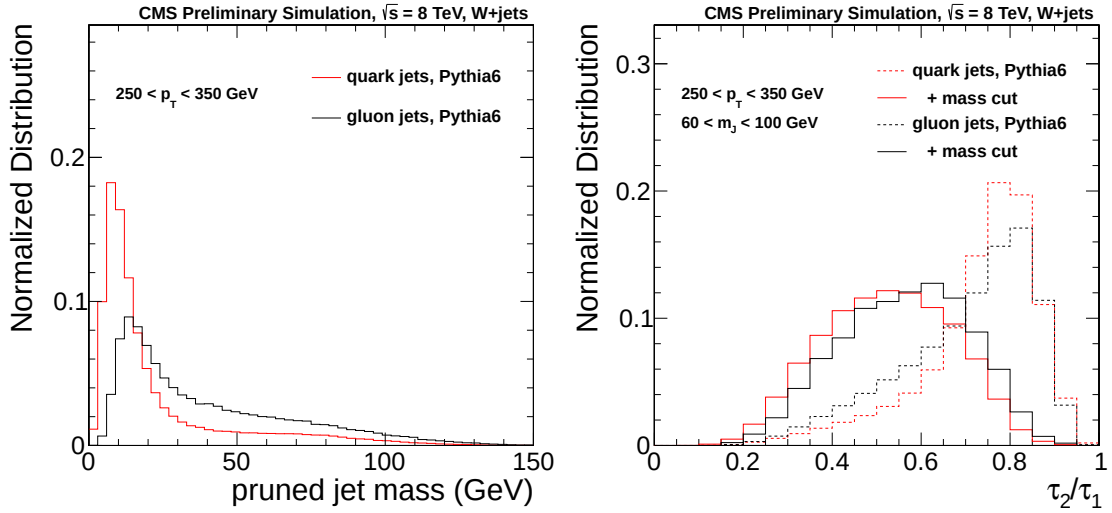

Figure 3.15: Pruned jet mass and τ_2/τ_1 distributions, obtained after cutting on the pruned jet mass, for quark and gluon jet simulated with Pythia 6.

while, in the $t\bar{t}$ sample, looking at W-jet candidates with $p_T > 200$ GeV. Starting from Figure 3.16, pruned jet mass and τ_2/τ_1 distributions are shown in the W+jets topology, which clearly corresponds to a QCD jets enriched region. In Figure 3.17, the same distributions are reported for the di-jet sample. In the W+jets control region, Herwig++ agrees better with data with respect to Pythia 6, while, in the di-jet one, the best substructure modelling is provided by Pythia 8.

To test the description of W-jets, a control sample of pure hadronic W-bosons in data is obtained from high p_T semi-leptonic $t\bar{t}$ events. Applying the $t\bar{t}$ event selection described in Section 3.2.2, Figure 3.18 shows the pruned jet mass and the N-subjettiness τ_2/τ_1 for W-jet candidates in this region. Both systematic and statistical uncertainties are included in the error band, where systematics are coming from the theoretical uncertainties on di-boson, single-top

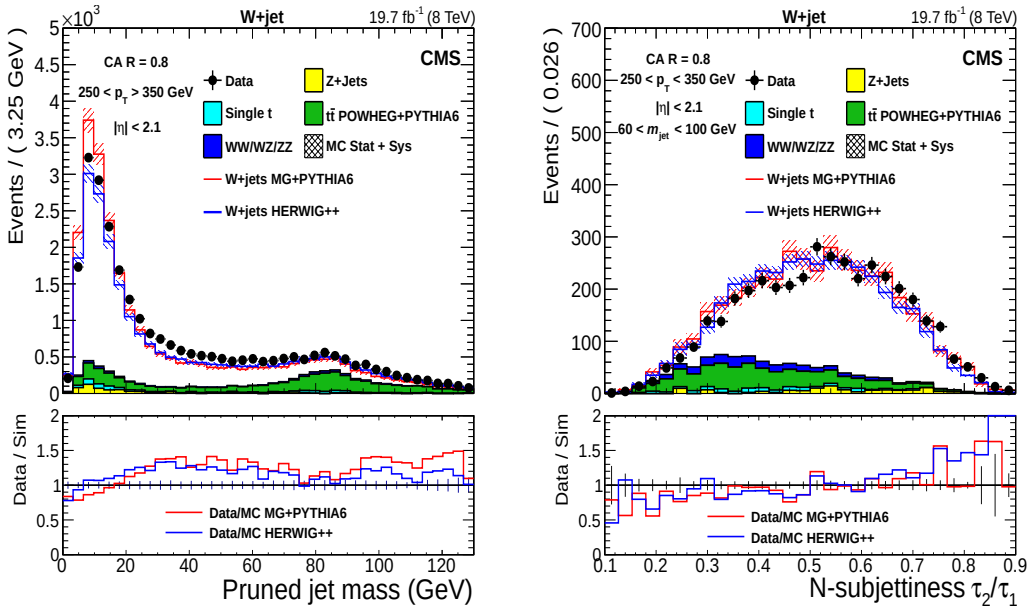

Figure 3.16: Pruned jet mass (left) and N-subjettiness τ_2/τ_1 (right) distributions in data and simulation for W+jets events. Below each figure the relative deviations, between data and MC, are reported.

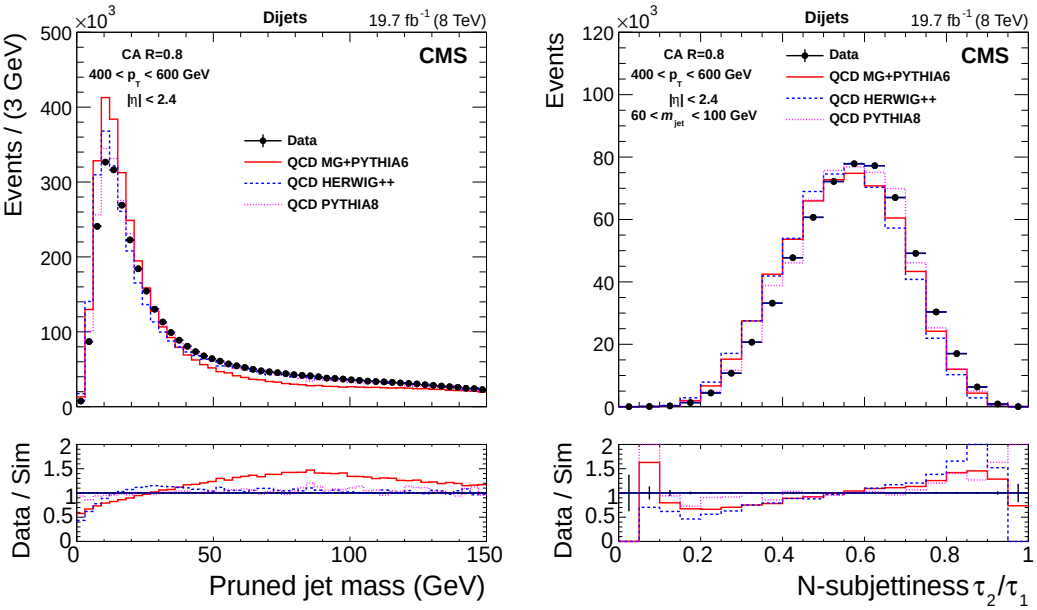

Figure 3.17: Pruned jet mass (left) and N-subjettiness τ_2/τ_1 (right) distributions in data and simulation for di-jet events. Below each figure the relative deviations, between data and MC, are reported.

and $t\bar{t}$ production cross sections, when evaluated in a boosted regime. The agreement between data and simulation shows discrepancies at the level of about 20%, where POWHEG + Pythia6 provides a slightly better description with respect to MC@NLO + Herwig++.
Finally, the jet charge distribution for W-jets is reported in Figure 3.19, always considering events in the semi-leptonic $t\bar{t}$ control region. By selecting a negatively or positively charged

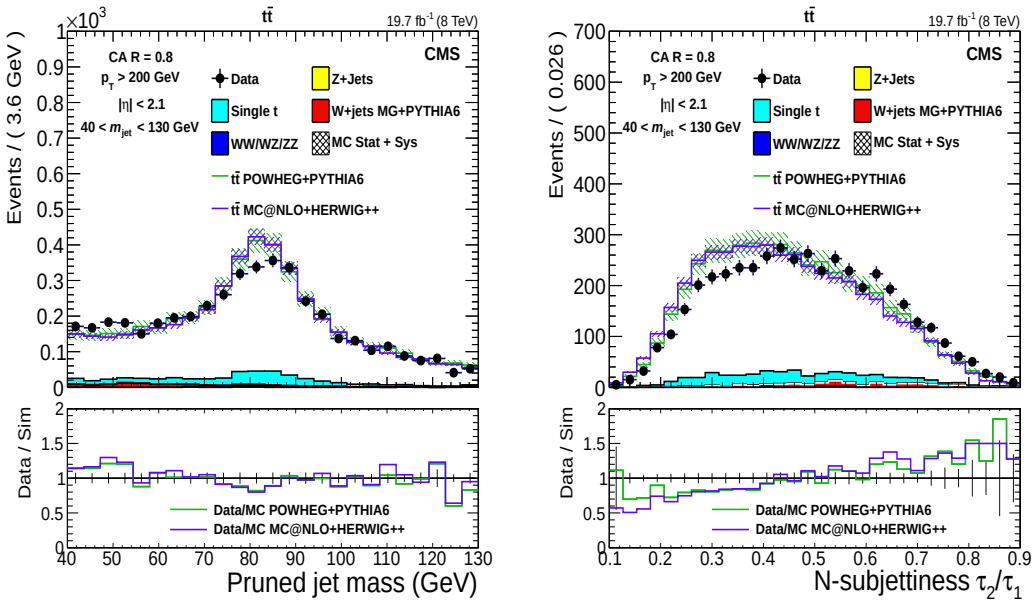

Figure 3.18: Pruned jet mass and τ_2/τ_1 distributions inside the semi-leptonic $t\bar{t}$ control sample, considering only muon + jets final state. Below each figure the relative deviations, between data and simulations, are reported.

lepton, one can effectively tag a W^+ or a W^- jet. As expected from simulation based studies in Section 3.2.4, charged V-jets cannot be distinguished on an event-by-event basis.

3.2.8 Mis-tagging rate measurement

The di-jet sample is used to measure, over a wide p_T range, the rate of QCD jets passing the W-tagging selection, also known as mis-tag rate. The measurement performed in data is compared with both Herwig++, Pythia6 and Pythia8 simulations. In Figure 3.20, the fraction of QCD jets passing the pruned mass requirement, $60 < m_{\text{pruned}} < 100$ GeV, as well as the simultaneous m_{pruned} and τ_2/τ_1 selections are shown as a function of the jet transverse momentum and the number of reconstructed primary vertices. The mistag rate of the pruned mass criterion is better reproduced by both Herwig++ and Pythia8, while Pythia6 underestimates it. In addition, when both pruned mass and N-subjettiness requirements are applied, the fake rate in data is better modelled by the Pythia8 simulation. The pruned mass fake rate is stable, around 1% level, as a function of pileup, while it drops when the τ_2/τ_1 selection is applied, as explained in Section 3.2.3.

3.2.9 Data-to-MC scale factors and jet mass resolution

The $t\bar{t}$ control region is used to extract data-to-simulation correction factors for the W-jet tagging efficiency, as well as to measure possible miss-modelling of W-jet mass scale and resolution. These corrections are derived using as a benchmark POWHEG plus Pythia6 $t\bar{t}$ simulation, as it has been adopted as reference sample in the analysis described in Chapter 4 and in Chapter 5. Since one is concerned only with efficiency, mass scale and resolution for pure W-jets, background contamination must be identified and properly subtracted.
Minor background sources, coming from W+jets, single-top and di-boson productions are

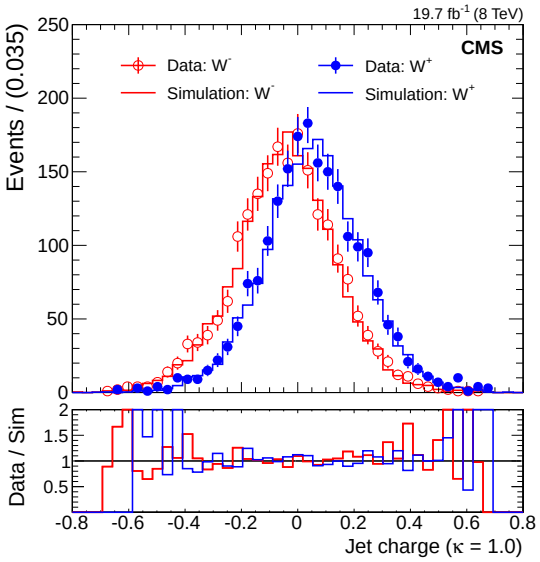

Figure 3.19: Jet charge distributions in the $t\bar{t}$ control sample for W^+ and W^- jets, selected by requiring negatively and positively charge leptons, respectively. Simulated distributions reflect the sum of $t\bar{t}$ (POWHEG interfaced with Pythia 6) and all other background processes. Below each figure the relative deviations between data and simulation is reported.

estimated from the simulation. Once these contributions are subtracted, the remaining $t\bar{t}$ events are not representing a pure set of W-jets. In fact, there is still the presence of a combinatorial background mainly due to additional jets that can be wrongly selected as W-jet candidates or, when the top quark is highly boosted, the b-jet closest to the hadronic W-boson is partially clustered inside the CA8 cone, changing the jet shape. For this reason, two additional requirements are imposed:

- To reduce the effect of partial merging between the b-jet and W-jet one inside the CA8 cone, the pruned mass window is shifted up by 5 GeV, to be $65 < m_{\text{pruned}} < 105$ GeV. This size of this shift has been estimated looking at $t\bar{t}$ simulated events.

- To reduce the combinatorial contribution related to a wrong W-jet selection, the angular distance ΔR between the W-jet candidate and the closest b-tagged AK5 jet, identified through the CSV algorithm, is required to be less than 2.0. This allows to exclude the events in which the b-jet closest to the leptonically decaying W-boson or additional q/g jets have both high transverse momentum and pruned mass close to W-boson one. This selection reduces the scale factor uncertainty by around 20%, enhancing the purity of the selected W-jet sample.

To extract the pure W-jet selection efficiency, the mass scale and resolution, a simultaneous unbinned fit to the pruned mass for events passing or failing the V-tagger selection is used. To have a correct model for the pruned mass shape of real W-jets and the combinatorial background, each simulated hadronic W-boson candidate is matched to the corresponding generated one, within a cone $\Delta R < 0.3$. Matched jets represent real W-jets, while unmatched ones are used as a proxy for the combinatorial background. Furthermore, since the N-subjettiness

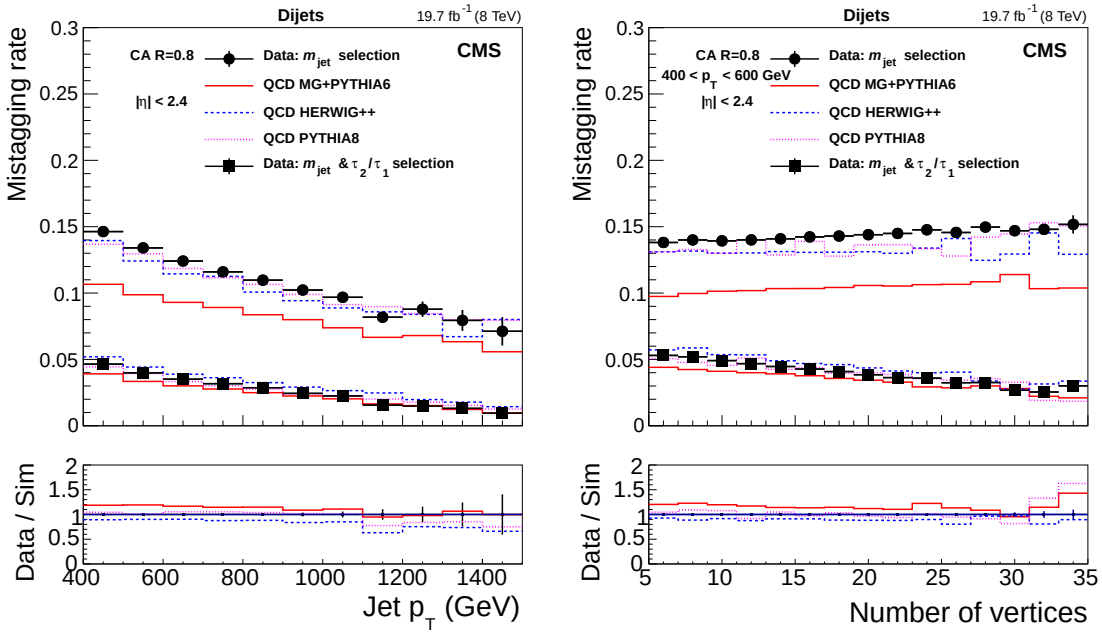

Figure 3.20: Fraction of jets passing the m_{pruned} and the τ_2/τ_1 selections in di-jet data and simulation as a function of the W-jet candidate jet p_T (left) and the number of reconstructed vertices (right). The data over simulation ratio is shown only for the $m_{\text{pruned}} + \tau_2/\tau_1$ tagger.

criterion, $\tau_2/\tau_1 < 0.5$, is not fully efficient in selecting real W-jets, independent matched and un-matched shapes are extracted for both events passing or failing the τ_2/τ_1 selection. These generator level distributions are parametrized according to the following analytical functions:

- Distributions of the pruned mass for un-matched jets are fitted with:

$$f_{\text{un-match}}(m_{\text{pruned}}) = \frac{1 + \text{Erf}((m_{\text{pruned}} - a)/b)}{2} \cdot e^{c \cdot m_{\text{pruned}}} \tag{3.10}$$

- The pruned mass for jets matched with generator level W-bosons, failing the N-subjettiness selection, is fitted with combination of a Gaussian function and a polynomial background:

$$f^{\text{fail}}_{\text{match}}(m_{\text{pruned}}) = \text{Gaus}(m_{\text{pruned}}, \overline{m}, \sigma) + (a + b \cdot m_{\text{pruned}} + c \cdot m^2_{\text{pruned}}) \tag{3.11}$$

- Events passing τ_2/τ_1 selection and matched with simulated W-bosons are parametrized with a double Gaussian:

$$f^{\text{pass}}_{\text{match}}(m_{\text{pruned}}) = \text{Gaus}(m_{\text{pruned}}, \overline{m}, \sigma) + \text{Gaus}(m_{\text{pruned}}, \overline{m}_1, \sigma_1) \tag{3.12}$$

Figure 3.21 shows the results of the fits to simulated events for matched (upper line) and un-matched (bottom line) events that pass (left) or fail (right) the N-subjettiness selection. Finally, splitting the selected semi-leptonic $t\bar{t}$ events according to the lepton flavor (μ+jets or e+jets), a simultaneous fit to the pruned jet mass, for events that are passing or failing the τ_2/τ_1 criterion, is performed by means of the product of the following likelihoods:

$$\mathscr{L}_{\text{pass}} = e^{-(N_{\text{W}} \cdot \varepsilon_{\text{W}} + N_{\text{comb}} + \sum_j^{bkg} N_j)} \prod_i^{N_{\text{pass}}} \left[N_{\text{W}} \cdot \varepsilon_{\text{W}} \cdot f^{\text{pass}}_{\text{match}}(m) + N_{\text{comb}} \cdot f^{\text{pass}}_{\text{un-match}}(m) + \sum_j^{bkg} N_j \cdot f^j_{\text{pass}} \right]$$

$$\mathscr{L}_{\text{fail}} = e^{-(N_{\text{W}}\cdot(1-\varepsilon_{\text{W}})+N'_{\text{comb}}+\sum_j^{bkg} N'_j)} \prod_i^{N_{\text{fail}}} \left[N_{\text{W}}\cdot(1-\varepsilon_{\text{W}})\cdot f^{\text{fail}}_{\text{match}}(m) + N'_{\text{comb}}\cdot f^{\text{fail}}_{\text{un-match}}(m) + \sum_j^{bkg} N'_j \cdot f^{j}_{\text{fail}} \right]$$

where m is the pruned jet mass, N_{W} stands for the number of real W-jets in the $t\bar{t}$ sample, ε_{W} is the W-tagging efficiency, j loops on the minor backgrounds (di-boson, single-top and W+jets), N_{comb} and N'_{comb} are the number of the selected combinatoric jets. In addition, N_{W}, ε_{W} and the Gaussian parameters ($\overline{m}$, σ), used to model the resonant part of the pruned mass shape, correlate events passing or failing the τ_2/τ_1 selection, as well as the two lepton flavor categories. Eventually, the W-tagging scale factor is defined as:

$$\text{SF}_{\text{W}} = \frac{\varepsilon_{\text{data}}}{\varepsilon_{\text{MC}}} \tag{3.13}$$

Results of this simultaneous fit are displayed in Figure 3.22, where top (bottom) plots show the fit in the muon (electron) category, while left (right) column refers to events passing (failing) the N-subjettiness requirement. A nice agreement between data and simulation is shown for events which fulfill the W-tagging requirements, while the sample for which $\tau_2/\tau_1 > 0.5$ is

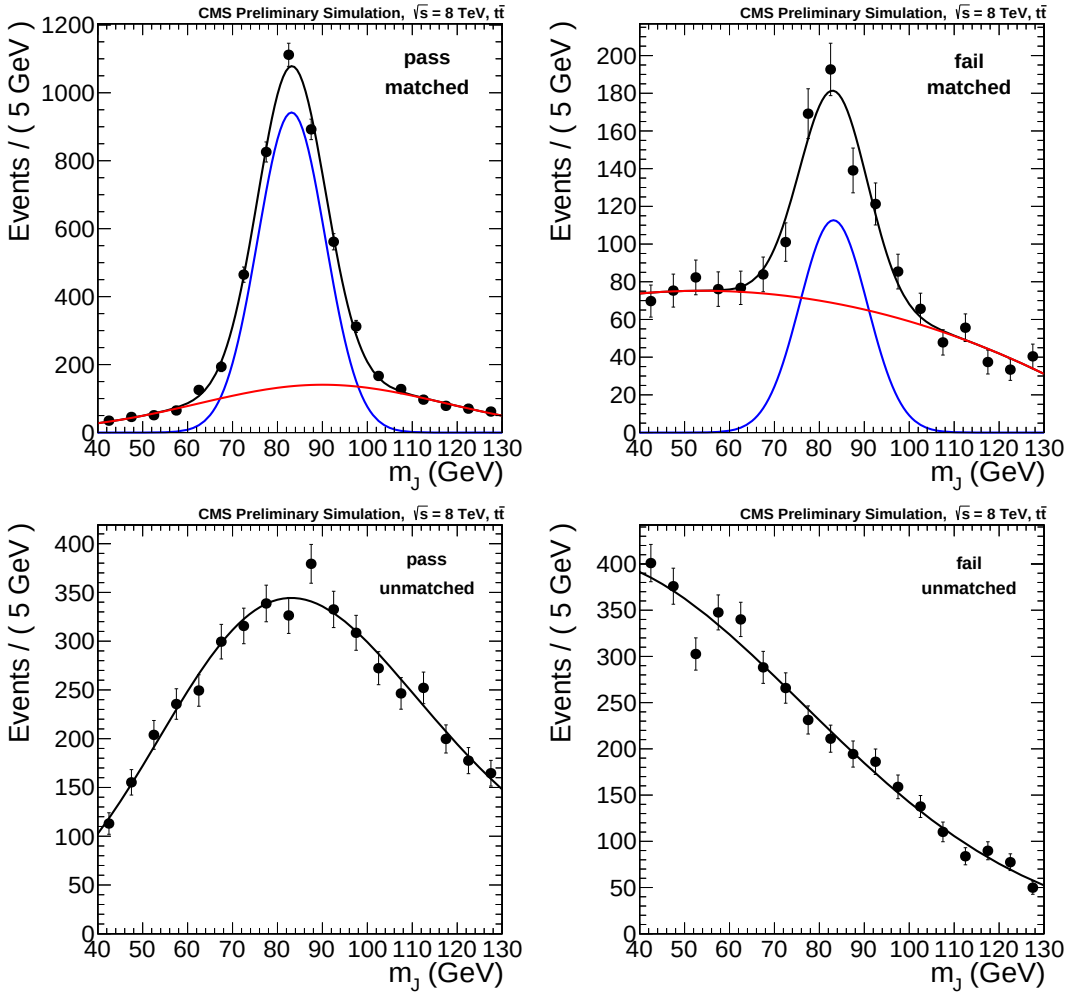

Figure 3.21: Matched (top) and un-matched (bottom) W-jet candidates, with respect to the generated hadronic W-boson, for simulated $t\bar{t}$ events passing (left) and failing (right) the selection $\tau_2/\tau_1 < 0.5$.

not well modeled by the MC simulation, in particular when the W-jet candidate is not coming from a boosted W-boson but from a quark or a gluon jet. The scale factor, as defined in Equation 3.13, is measured to be 0.93 ± 0.06, where the uncertainty is purely statical from the fit. The p_T dependence of this scale factor has been studied, taking into account the limited statistics of events surviving the analysis selections. Considering two p_T bins of 200-265 GeV and of 265-600 GeV, the measured scale factors are 1.00 ± 0.09 and 0.92 ± 0.10, respectively. Thus, no significant dependance, as a function of the jet transverse momentum, is observed.
Corrections for the pruned jet mass scale and resolution are derived comparing the Gaussian mean $\overline{m}$ and width σ, used to model the W-boson peak in both passing and failing events, between the values extracted from a fit to the data and the ones coming from the simulation. The resulting parameters are reported in Table 3.2. Mass scale and resolution are larger in data than in the MC, therefore the scale must be shifted by $(1.7 \pm 0.6)\%$ and the resolution must be enlarged by $(11 \pm 9)\%$ in the simulation.

Parameter	Data	Simulation	Data/Simulation
$\langle m \rangle$	84.1 ± 0.4 GeV	82.7 ± 0.3 GeV	1.017 ± 0.006
σ	8.4 ± 0.6 GeV	7.6 ± 0.4 GeV	1.11 ± 0.09

Table 3.2: Summary of the fitted W-boson mass peak model parameters in semi-leptonic $t\bar{t}$ events.

In searches exploring signatures with high p_T V-jets, it is common to recover the V-tagging signal inefficiency introducing a low purity category, grouping events with a lower N-subjettiness value $0.5 < \tau_2/\tau_1 < 0.75$. This sample is usually statistically combined with the so called high purity one ($\tau_2/\tau_1 < 0.5$), enhancing the analysis performance.
In this way, events are classified into three categories depending on the W-jet candidate N-subjettiness value: high purity, low purity and failing events ($\tau_2/\tau_1 > 0.75$). Since for probability conservation $\varepsilon_{\mathrm{HP}} + \varepsilon_{\mathrm{LP}} + \varepsilon_{\mathrm{fail}} = 1$, the scale factor for low purity category is defined as:

$$\mathrm{SF}_{\mathrm{LP}} = \frac{1 - \varepsilon_{\mathrm{HP}}^{\mathrm{data}} - \varepsilon_{\mathrm{fail}}^{\mathrm{data}}}{1 - \varepsilon_{\mathrm{HP}}^{\mathrm{MC}} - \varepsilon_{\mathrm{fail}}^{\mathrm{MC}}} \tag{3.14}$$

where $\varepsilon_{\mathrm{fail}}^{\mathrm{data}}$ and $\varepsilon_{\mathrm{fail}}^{\mathrm{MC}}$ are obtained fitting the event sample with $\tau_2/\tau_1 > 0.75$ with a likelihood given by the superposition of two exponential shapes, one for failing matched W-jet candidates and another one for the un-matched ones. Finally, the scale factor for this low purity category is measured to be 1.28 ± 0.30, when muon and electron channels are simultaneously fitted.

3.2.10 Systematic uncertainties

Several systematic uncertainties affecting the scale factor estimation are considered: parton shower modeling of boosted V-jets, PDF used in the simulation, vector boson polarization, pileup simulation, contamination from nearby jets, jet mass scale and resolution, jet energy scale and resolution, as well as less dominant effects like uncertainties in lepton identification, b-tagging and $\not{E}_T$.
The joint effect of parton shower modeling and parton densities is measured by the difference between W-jet efficiencies evaluated from POWHEG, interfaced with Pythia 6, and a sample generated with MC@NLO interfaced with Herwig++. Furthermore, the effect related to the underlying event simulation is estimated comparing the predictions from three alternative

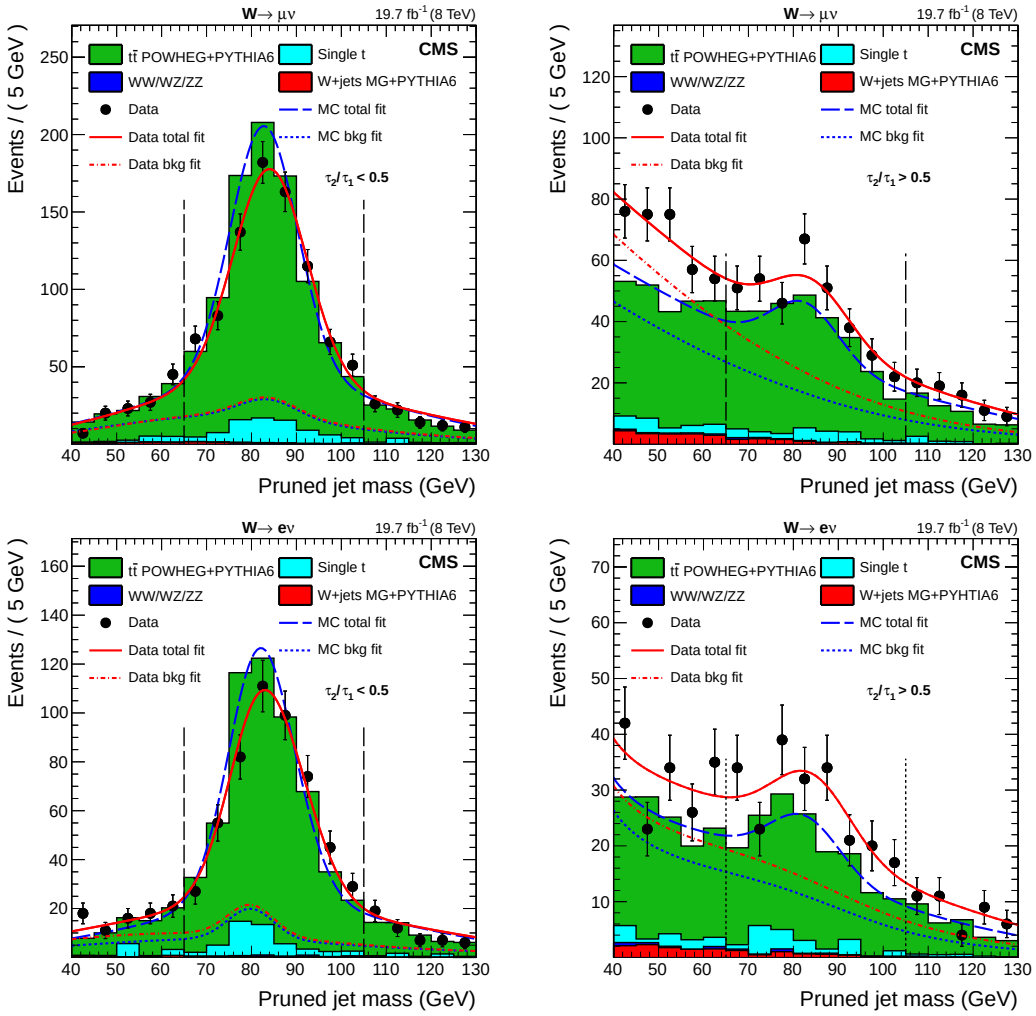

Figure 3.22: Pruned jet mass distribution in the $t\bar{t}$ control sample that pass (left column) or fail (right column) the $\tau_2/\tau_1 < 0.5$ selection for muon (top line) and electron (bottom line) channels. The result of the fit to data and simulation are shown, respectively, by the solid and long-dashed line, while the background components of the fit are shown as dashed-dotted and short-dashed line.

`Pythia6` tunes: `Z2*`, `AMBT1` [129] and `AMBT2` [130], taking the maximal observed difference as an estimate of this contribution. The vector boson polarization has an important impact on the W-tagging performance, that has to be taken into account when the $t\bar{t}$ scale factor is propagated to specific analysis final states. W-bosons polarization in $t\bar{t}$ has been measured by `CMS` [131], with a precision of 4%. Although the agreement was found to be better respect to the quoted precision, this number is used as an upper bound and is included in the final W-tagging uncertainty.

The effect of the pruned jet mass scale and resolution are evaluated by shifting or smearing each simulated W-jet candidate, using the uncertainty extracted from the fitted W-boson peak mean $\overline{m}$ and resolution σ, which are listed in Table 3.2. The effect on the efficiency is small, since the W-boson mass peak is well contained inside the pruned mass window $65 < m_{\text{pruned}} < 105$ GeV, adopted for the baseline W-tagger definition. Similarly, the jet energy scale and resolution are changed according to their p_T and η dependent uncertainties [84].

The contribution from pileup is determined by changing the minimum bias cross section within its measured uncertainty of about 6% [132]. The scale factors for lepton and b-jet identification, used to cover discrepancies between data and simulation, are also changed within their uncertainties to estimate the effect on the W-tagger.
Finally, the impact of nearby jets on the scale factor is estimated by comparing the central prediction to an alternative evaluation, performed with an explicit requirement on the angular distance between the closest AK5 jet and the W-jet candidate of $\Delta R > 1.3$.
Results are summarized in Table 3.3, where the dominant systematic effect comes from modeling of the parton shower and PDF, measured to be around 6%. Assuming that systematic sources are un-correlated each other, their quadratic sum leads to a total systematic uncertainty of 7.6%, which is comparable to the statistical one of about 6.4%.

Source	Effect on the scale factor
Parton shower + PDF	6.0%
Underlying event	$< 0.5\%$
W-polarization	2.0%
Pileup	1.8%
Nearby jets	2.4%
Jet mass scale	$< 0.5\%$
Jet mass resolution	1.6%
Jet energy scale	2%
Jet energy resolution	0.9%
Lepton ID	$< 0.5\%$
b-tagging	$< 0.5\%$
$\not{E}_T$	$< 0.5\%$
Total systematic	7.6%
Statistical	6.4%
Total	9.9%

Table 3.3: Summary of systematic uncertainties on the W-jet identification efficiency scale factor as measured in the $t\bar{t}$ control sample.

3.2.11 Scale factor at high p_T

The V-tagging scale factors aim to correct the effect of mis-modelings of jet substructure variables on the estimated signal efficiency in simulated events. These are mainly due to imperfections in the parton shower simulation, as it represents the most relevant systematic source in the scale factor computation. The scale factors are extracted from a $t\bar{t}$ control sample that, because of limited statistics in the dataset and W-jet/b-jet merging for boosted top quarks, covers only a limited W-jet transverse momentum range. Therefore, it is not possible to control whether the quality of the parton shower simulation is stable over a large range of jet transverse momenta, useful for searches of new physics with jets in the TeV range.
The baseline analysis uses Pythia6 as benchmark, considering Herwig++ as alternative reference. One starts from the same parton-level generation of Bulk graviton events produced by the JHUGen generator, which are considered as source of W-jets, scanning a wide range of jet p_T by changing the input mass hypothesis. These events are then showered and hadronized

by the two aforementioned showering models. Differences between the two shower approaches are visible especially in the description of the τ_2/τ_1 distribution, where Herwig++ shows a shift to higher values. To assess the p_T-dependence of the V-tagging selection, the ratio of the efficiencies derived from these alternative models is studied as a function of the resonance mass, which is highly correlated with the p_T of the W-jet candidate. This trend must be consistent within the uncertainty already measured on the V-tagging scale factors at low p_T, i.e. 8% for the high purity selection and 30% for the low purity one.
In the high purity category, no clear dependence versus the resonance mass is observed, while in the low purity one a trend is observed, but its size is still of the order of the measured systematic uncertainty. Numeric values are reported in Table 3.4. Based on this observation, any further systematic has to be applied when high p_T simulated V-jet are considered.

m_G (TeV)	Category	ε_s(Pythia 6)	ε_s(Herwig++)	Ratio (Herwig++/Pythia 6)
1.0	HP	40.9%	39.9%	0.976 ± 0.013
1.8	HP	33.9%	33.9%	0.981 ± 0.020
2.5	HP	24.2%	23.4%	0.964 ± 0.025
1.0	LP	10.8%	15.3%	1.415 ± 0.039
1.8	LP	12.5%	15.8%	1.260 ± 0.044
2.5	LP	13.6%	15.7%	1.154 ± 0.040

Table 3.4: V-tagging efficiencies for Bulk graviton events hadronized with Pythia 6 or Herwig++, calculated summing up the electron and muon channels. sSatistical only uncertainties are considered with binomial error propagation.

3.3 V-tagging performance in view of LHC Run-II

In this Section, the expected performances of V-tagging techniques are studied on a set of dedicated simulations at $\sqrt{s} = 13$ TeV, with a pileup scenario foreseen during the LHC Run-II [83, 94]. The increase of beam intensities, producing possibly up to fifty interactions per bunch crossing, will generate conditions that adversely affect the performance of substructure quantities. Moreover, the increase in center of mass collision energy, from 8 to 13 TeV, will raise the production rate and energies of highly boosted objects. Therefore, the joint effect of higher pileup and higher boost represents a challenge for the jet substructure algorithms.
To investigate V-tagging strategies in this new set of conditions, the multi-jet topology is considered as a benchmark. For each event, the leading PF-CHS jets with $p_T > 300$ GeV, clustered via anti-k_t $R = 0.8$, is considered as V-jet candidate. Background jets (q/g jets) are taken from QCD multi-jet events generated via Pythia 8, while signal W-jets are produced by RS gravitons decaying into WW pairs.
In contrast with what was done with 8 TeV data, where a single set of grooming parameters, suggested by phenomenological studies, was adopted, several algorithms are compared as a function of different possible configurations, as reported in Table 3.5. Groomed masses are corrected for pileup contamination through a four-vector "safe" subtraction [133], which is an extension of area subtraction including a correction for jet masses [85, 134]:

$$p^{\mu}_{\text{sub}} = p^{\mu} - \rho \cdot A^{\mu} - \rho_m \cdot A^{\mu}_m \tag{3.15}$$

Grooming algorithm	Parameters
Trimming [96]	$R_{trim} = 0.2$, $p_T^{frac} = 0.05$ $R_{trim} = 0.2$, $p_T^{frac} = 0.03$ $R_{trim} = 0.1$, $p_T^{frac} = 0.03$ $R_{trim} = 0.3$, $p_T^{frac} = 0.03$
Pruning [97]	$z_{cut} = 0.10$, $D_{cut} = 0.50$ $z_{cut} = 0.05$, $D_{\mathrm{cut}} = 0.50$ $z_{cut} = 0.10$, $D_{\mathrm{cut}} = 0.75$ $z_{cut} = 0.05$, $D_{\mathrm{cut}} = 0.75$
Soft drop [98]/MMDT [99]	$z_{cut} = 0.1$,$\beta = 0$ $z_{cut} = 0.1$, $\beta = 1$ $z_{cut} = 0.1$, $\beta = 2$

Table 3.5: Summary of grooming parameters considered for trimming, pruning and soft drop algorithms.

where ρ and ρ_m are measures of the average pileup density. In particular, ρ_m is used to account for the non-zero hadron masses of particle-flow inputs. Since CHS is applied, a special care is required to estimate ρ_m from neutrals candidates only when jets reconstructed in the central region $|\eta| < 2.5$ are corrected, while using the full η acceptance to correct forward jets. In the cases of soft drop, modified mass drop tagger and trimming, the four-vector subtraction corrects both jet p_T and mass at each step in the clustering algorithm, while for pruning, the correction is applied to the final jets using the pruned jet area. Different figures of merit are used to judge each algorithm and the possible configurations:

- **Pileup sensitivity**: it is used to evaluate the robustness of each algorithm setup, looking at the performance stability as a function of the amount of pileup.
- **Mass response**: it measures the accuracy of the algorithm in reconstructing the jet mass with respect to the generator level prediction (m_{gen}), matching the reconstructed jet with the generated one within $\Delta R < 0.3$.
- **W-jet mass resolution**: the ability to reconstruct, with high resolution, the W-boson mass peak is an indirect measurement of the separation power between signal and background jets.

The goal consists in finding the configuration which allows to push QCD jets to lower mass values and, at the same time, to preserve a stable performance and resolution for signal jets as a function of pileup. In Figure 3.23, the jet mass distribution, obtained for each groomer, is shown for high p_T QCD jets as a function of parameter configurations listed in Table 3.5. The best performing setups are soft drop ($\beta = 0$), pruning ($z_{cut} = 0.1$, $D_{cut} = 0.5$) and trimming ($R_{trim} = 0.1$, $p_T^{frac} = 0.03$), since they tend to move the bulk of the groomed mass distribution to smaller values.
Moreover, in Figure 3.24, to investigate the pileup dependence of each algorithm, the average jet mass for QCD jets is shown as a function of the number of reconstructed vertices ($\mathrm{N_{PV}}$). For trimming and soft drop, the jet mass is perfectly stable versus pileup thanks to the four vector subtraction, while pruning shows, for all the considered setups, a worsening in the

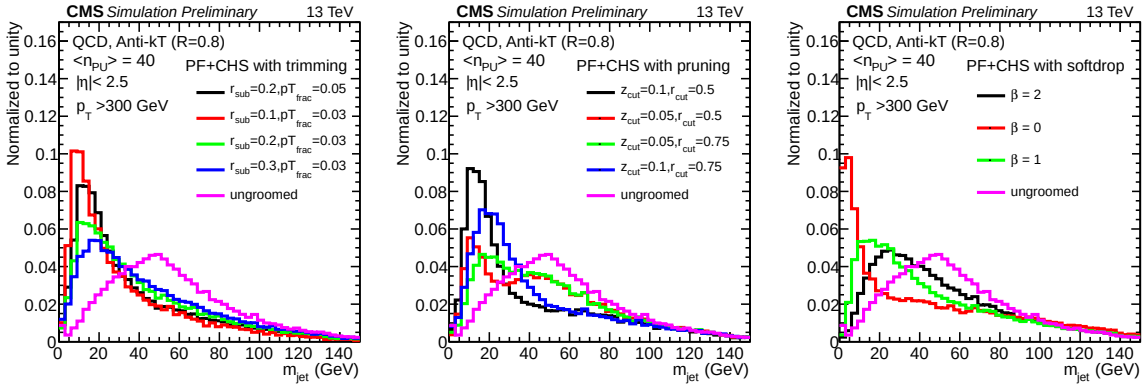

Figure 3.23: Groomed QCD jet mass distributions obtained applying trimming (left), pruning (middle) and soft drop (right) as a function of the parameter scan reported in Table 3.5.

performance in a high pileup environment. Simultaneously, the effect of each groomer on W-jets is evaluated by looking at the RMS of the mass response distribution as a function of N_{PV}. Trimming shows a slight improvement in the mass resolution with respect to the other techniques, as reported in Figure 3.25.
Finally, to quantify the expected separation power between signal and background jets, the W-jet mass scale is extracted, noting that trimming shows an offset to negative values, while pruning and soft drop preserve, within the experimental resolution, the generated W-boson lineshape. In addition, trimming is a fixed cone algorithm, which means that its parameters have to be carefully adapted as a function of the considered boosted regime. For these reasons, soft drop ($\beta = 0$) and pruning ($z_{cut} = 0.1$, $D_{cut} = 0.5$) still represent the favorite groomers for future V-tagging strategies.

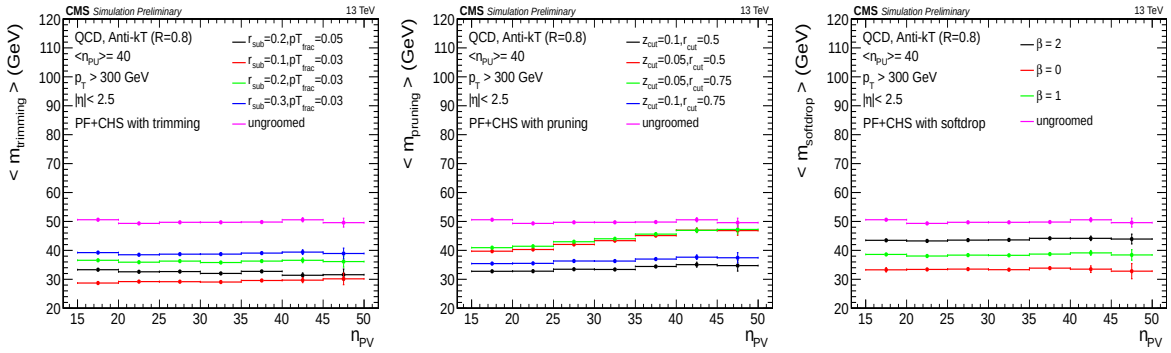

Figure 3.24: Pileup dependence of the average jet mass for PF+CHS jets for several grooming algorithms and parameters, evaluated in a QCD jet enriched region. From left to right: trimming, pruning and soft drop.

3.3.1 V-tagging performance vs pileup

Considering only jets in a moderate p_T regime 475-600 GeV, to select a phase space where signal and background events have almost the same falling jet p_T spectrum, the performance of groomers and jet substrcture observables are compared in two pileup conditions: a low pileup region, requiring $N_{PU} < 40$ with an average pileup of 31 interactions, and a higher one, where $N_{PU} > 40$ which has a mean of 47 pileup events. The "out-of-time" contribution, which

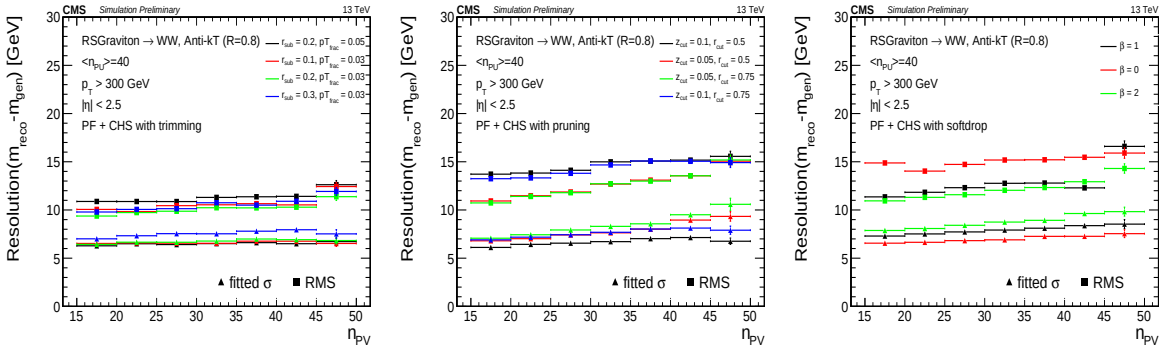

Figure 3.25: Pileup dependence for the W-jet mass resolution, evaluated only for reconstructed jets matched to the generated W-bosons, for various parameters of the grooming algorithms. Both the RMS and the σ, extracted from a Gaussian fit to the mass response distribution, are reported. From left to right: trimming, pruning and soft drop.

occurs when the integration time of signals from calorimeters is longer than the time difference between two consecutive bunch crossings, is expected to be less important the in-time one, thus it has not benn addressed by these studies, where no out-of-time contribution has been simulated.

In addition to the observables listed in Section 3.1, constituent subtraction is also considered, which involves a local subtraction of pileup particles at the level of jet constituents [135]. Figure 3.26 shows the ROC curves (ε_{sig} vs ε_{bkg}) for different groomed mass observables in both the low ($\mathrm{N_{PU}} < 40$, left) and high ($\mathrm{N_{PU}} > 40$, right) pileup regions. The performance of raw jet mass, trimming ($R_{trim} = 0.1$, $p_T^{frac} = 0.03$), pruning ($z_{cut} = 0.1$, $D_{cut} = 0.5$), constituent subtraction and soft drop ($\beta = \{0, 1, 2\}$) are compared. While, in Figure 3.27, the performance of τ_2/τ_1, τ_2, τ_1, $C_2(\beta = \{1, 1.5, 2\})$, Γ_{Qjet} and QGL discriminator, applied on the pruned jet and on the trailing pruned subjet, are reported. For both cases, no significant degradation in the performance is found at larger pileup.

Considering a signal efficiency benchmark of $\varepsilon_{sig} = 50\%$, the best performing mass variables are $m^{\beta=0,1}_{\mathrm{softdrop}}$ and the pruned mass, while the trimmed mass is known to be less powerful as one extends to higher p_T, since it relies on a fixed cone subtraction. In addition, looking to the various substructure variables, τ_2/τ_1 is clearly the best performing one.

To study the performance of a tagger which employs a combination of different observables, an MVA based on a Boosted Decision Tree (BDT) is adopted. The working point that maximizes the background rejection, for a signal efficiency of 50%, is taken as a benchmark for performance comparisons and the score Z, defined as the inverse of the background efficiency, is used as figure of merit. To determine the relevant minimal set of variables that performs at best and the corresponding pileup dependence, a training is developed for all the possible pairs and triplets of observables, where for triplets only all the possible pairs combined with the pruned mass are considered, in each of the two pileup bins. To calculate the uncertainty on the metric Z, two sources must be accounted for: the first is due to statistical fluctuations in the background efficiency evaluation, while the second originates from the uncertainty in the BDT training, due to a limited event sample, which is assumed to be of the order of the former one.

Figure 3.28 (left) shows the score Z for all BDTs trained on pairs of different observables in the low pileup bin. The best discrimination is achieved by combining jet mass variables, like

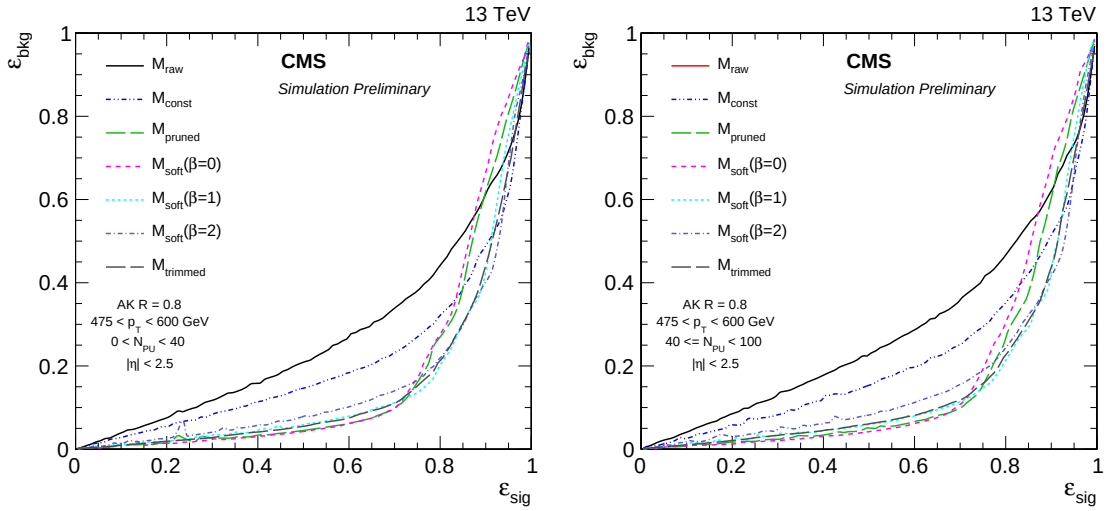

Figure 3.26: Comparison of the expected performance for various mass observables in the 475-600 GeV jet p_T bin, obtained considering simulated RS gravitons decaying into WW pairs as W-jet source. (Left) Comparison in the low pileup region $\mathrm{N_{PU}} < 40$, (right) high pileup one $\mathrm{N_{PU}} > 40$.

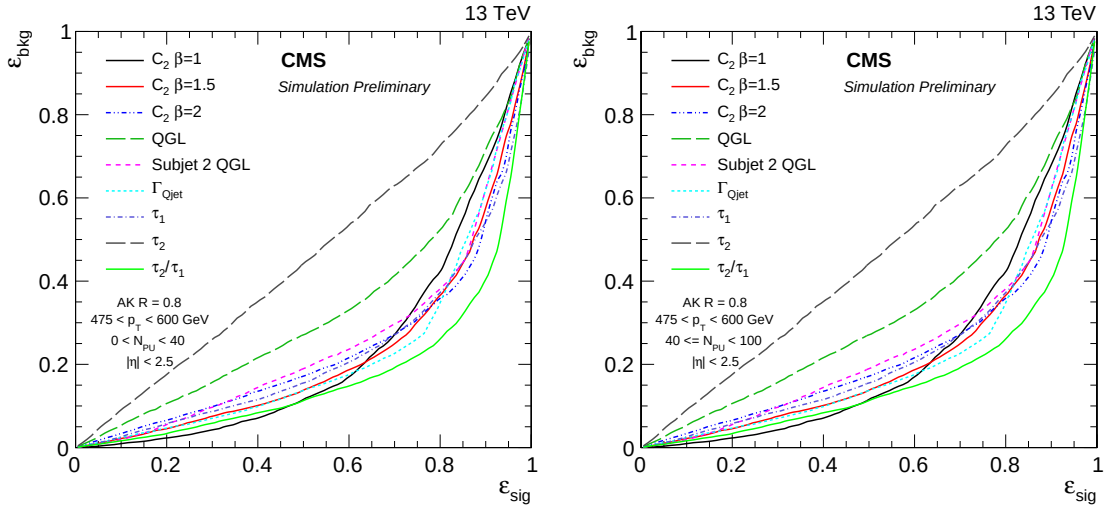

Figure 3.27: Comparison of various substructure observable performance in the 475-600 GeV jet p_T bin, considering W-jet in RS $\mathrm{G} \to \mathrm{WW}$ ($m_\mathrm{G} = 1$ TeV) events as signal. (Left) Comparison in the low pileup region $\mathrm{N_{PU}} < 40$, (right) high pileup one $\mathrm{N_{PU}} > 40$.

$m^{\beta=0}_{\mathrm{softdrop}}$ or $\mathrm{m_{pruned}}$, with τ_2/τ_1, which is consistent with results obtained in 8 TeV data and simulation. The second most powerful combination of variables comes from groomed masses and the newly introduced subjet `QGL`. Figure 3.28 (right) shows `BDT` performance combining triplets of observables, in events with leading jet p_T between 475 and 600 GeV. It's hard to identify the best combination of triplet variables, since a large class of observables lead to roughly the same discrimination power. These include m_{pruned} with τ_2/τ_1 and Γ_{Qjets}, or τ_2/τ_1 and a second mass variable, or subjet `QGL` with $C_2^{\beta=1}$. The expected difference on the Z score, between the best triplet and the combination of all the observables into a single `BDT`, labelled as "all" in Figure 3.28 as well as in the following ones, is around 15-20%.

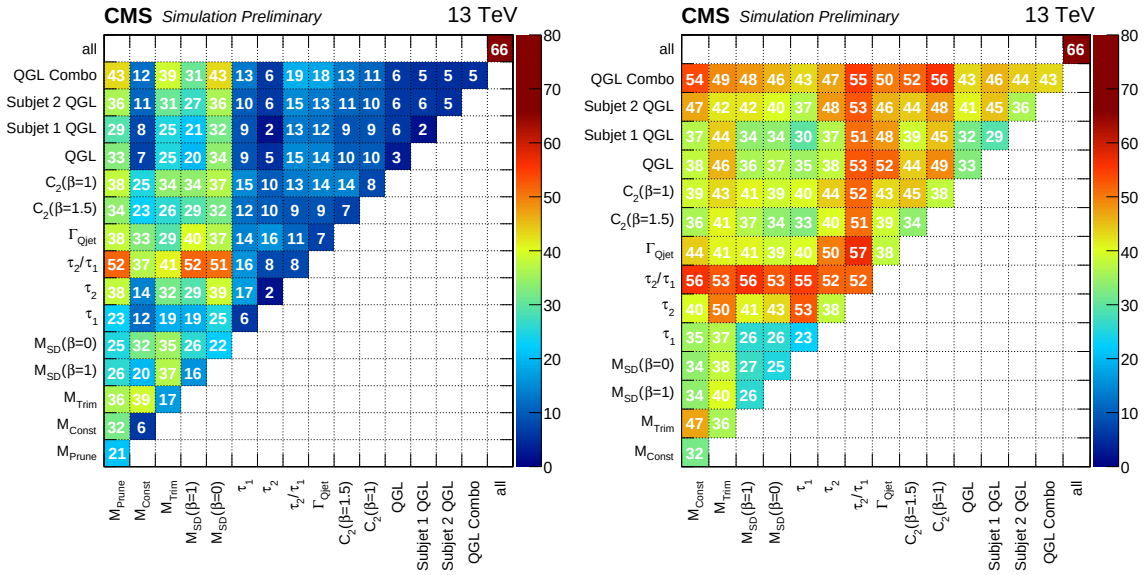

Figure 3.28: Performance score Z for V-tagging using BDT based on different pairs (left) and triplets (right) of substructure observables, in events with leading AK8 jet p_T in the range 475-600 GeV, for the low pileup bin $N_{PU} < 40$. The score Z is defined as $1/\varepsilon_B$, where ε_B is the background efficiency for a signal efficiency of 50%. "All" stands for the performance achieved combining all the considered observables into a single BDT.

When extending to higher pileup, a degradation of the overall discrimination is observed. The performance loss of each variable is assessed by taking the ratio between the Z score, evaluated in high pileup scenario ($N_{PU} > 40$), and the one in the low pileup region ($N_{PU} < 40$). This ratio is shown in Figure 3.29, in the form of a two-dimensional map for BDT pair and triplet trainings. The W-tagging performance degradation with pileup is limited at the level of 20%, when all variables are combined into a single discriminator. This difference is statistically significant since the uncertainty on the Z score, associated to each pairs of triplets, is around 5-7%. A few variables show a rather large degradation in the pileup performance, where the worst cases are represented by Γ_{Qjet} and the energy correlation function $C_2^{\beta=1}$. Both these variables lead to degradation around 30-40%, over a change in the average pileup of 16 interactions. This study was also extended to lower p_T jets, between 300-450 GeV, where a smaller degradation as a function of pileup was found.

Eventually, correlations represent an important means through which the discrimination of the individual observables is better understood, providing an estimation of their ranking when they are combined into multi-dimensional discriminators. To minimize the effect of hidden non linear correlations, each variable is transformed by running a BDT training, which transforms the input observable into a monotonically increasing function in signal yield versus background one. The resulting discriminant is used as an input to the following correlation study. Measuring the degree of linear correlation between single BDT variables and the total combination, observables are ranked in terms of separation power. First, mass variables appear as strongly correlated between each other, then subjet QGL is not correlated with any of the other observables. Finally, energy correlation functions C_2^{β} are also correlated with each other, but show limited correlations with other variables. As already shown in both Fig-

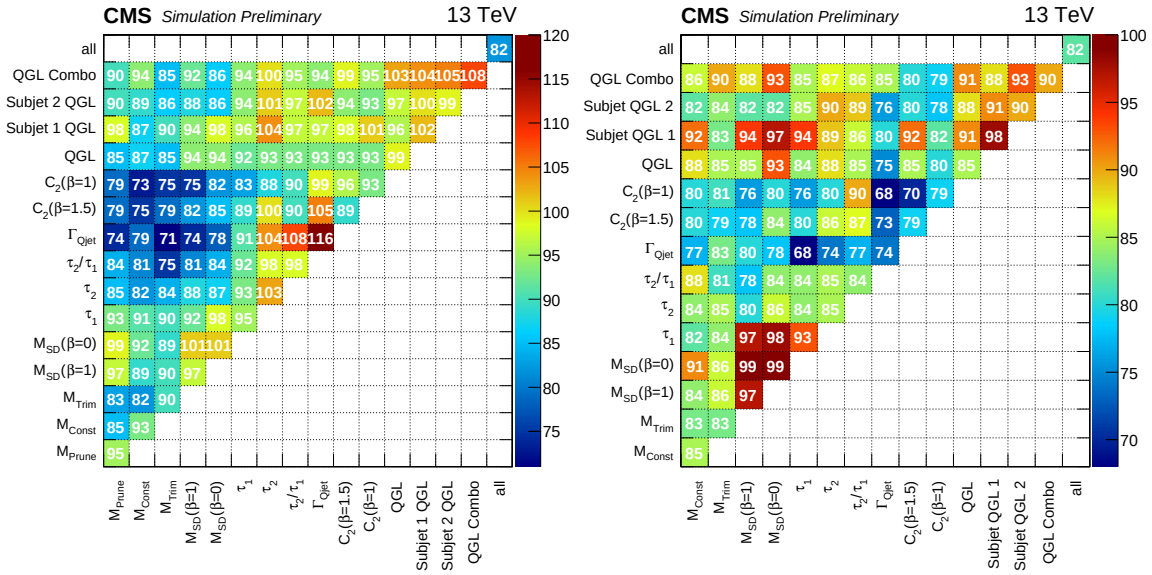

Figure 3.29: Percent ratio in Z score between high and low pileup scenarios: (left) result on BDT trained on pairs of variables, (right) result on BDT trained on triplets. The triplet includes all the pairwise combinations with m_{pruned}.

ure 3.28 and Figure 3.29, this indicates that building larger dimensional discriminators will benefit from additional information, associated with these uncorrelated variables.
To determine the effect of pileup, independent correlation matrices are extracted and compared between the two pileup regions, as reported in Figure 3.30 for both signal and background events. The variables which show the largest correlation reduction are $C_2^{\beta=1}$ and the subjet quark gluon discriminators, implying that their pileup dependence is causing the overall degradation in performance of the global discriminant. Finally, the largest positive correlation between single observables and the total combination is coming from the pruned mass and soft drop ($\beta = 1$).

3.3.2 V-tagging at high transverse momentum

With the increase of center of mass energy from 8 to 13 TeV, searches and measurements of jets with transverse momentum up to 4 TeV will be kinematically allowed. From the 8 TeV measurements, described in Section 3.2.4, the observed loss in the V-tagging performance for high p_T jets is related to breakdowns in the PF reconstruction, rather than substructure algorithms. PF jet reconstruction is based on a topological link between calorimetric clusters and tracker tracks, as illustrated previously in Section 2.4.3. In high p_T jets, particles are collimated with high momentum, thus track resolution becomes poor and the HCAL granularity is not sufficient to measure their direction. Since ECAL granularity is roughly five times better than the HCAL one, this can be used to improve the spatial resolution needed for substructure at high p_T. After reconstructing charged hadrons, if an energy excess is found in ECAL or HCAL, three different strategies have been tested:

- **Merge neutrals**: the excess is assigned to a single merged neutral candidate, which is either a photon or a neutral hadron.

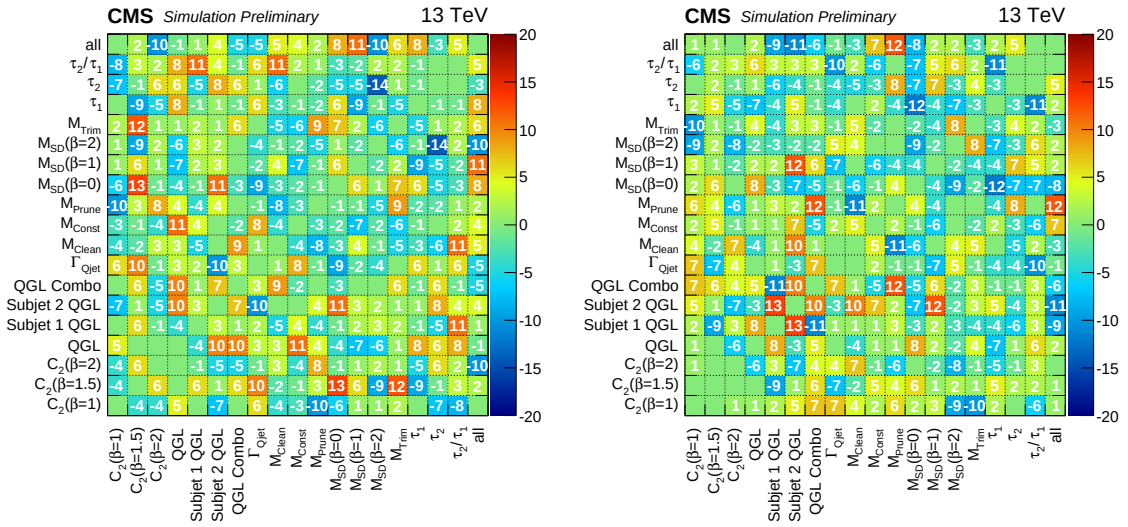

Figure 3.30: Difference between high pileup ($N_{PU} > 40$) and low pileup ($N_{PU} < 40$) in the correlation matrix of the BDTs, trained on single variables, with respect to each other for signal (left) and background (right) jets.

- **Split photons**: in case of an ECAL only excess, a photon candidate is created for each ECAL PF cluster.
- **Split photons and neutrals**: in case of an excess in both HCAL and ECAL, neutral hadrons and photons are created in the direction of each ECAL cluster. The total calorimetric energy is distributed to neutral hadrons according to the energy fraction of each cluster with respect to the total ECAL energy.

The effect on the performance of jet substructure reconstruction is quantified by analyzing the pruned jet mass and the N-subjettiness τ_2/τ_1. Figure 3.31 shows how the three scenarios of neutral particle reconstruction perform for W-jets of p_T around 2 TeV. The pruned mass resolution is significantly improved by splitting both photons and neutral hadrons, while τ_2/τ_1 distribution is pushed to lower values close to generator level predictions, improving the discrimination with respect to QCD jets. This means that ECAL granularity dominates jet substructure reconstruction when tracks are not fitted with a good resolution, reaching similar performance achieved in the standard tracking-dominated reconstruction.
In addition, Figure 3.32 shows the pruned jet mass scale and resolution as a function of the W-jet candidate p_T. The scale and resolution are obtained by fitting the mass response distribution with a Gaussian function, in the range of $\pm$ 1.5 σ around its mean value. The pruned jet mass scale appears to be stable up to the highest jet p_T, for both W-jets and quark/gluon ones. In contrast, the mass resolution increases from 5 to 8 GeV, while increasing the jet p_T from 0.5 to 3.5 TeV. The pruned jet mass resolution for quark/gluon-jets is about 4 GeV higher than the one for W-jets, due to the fact that QCD jets have lower masses fixing the same p_T range.
Finally, in Figure 3.33, the W-tagging efficiency and fake rate, defined requiring the pruned jet mass to be within $60 < m_{\text{pruned}} < 100$ GeV and $\tau_2/\tau_1 < 0.6$, is shown. The pruned jet mass selection efficiency increases up to a p_T of 1 TeV, because in a low boosted regime the jet cone does not fully contain the shower products of the two quarks produced by the W-boson decay,

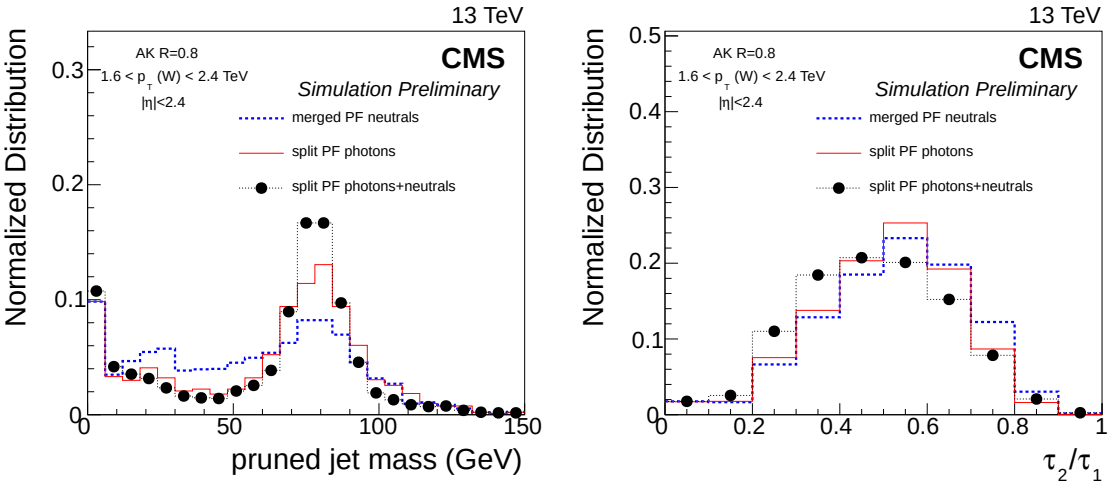

Figure 3.31: Reconstructed pruned jet mass (left) and N-subjettiness τ_2/τ_1 (right) for W-jets of p_T around 2 TeV, with an average of 40 simulated pileup interactions. Three PF neutral particle splitting configurations are compared.

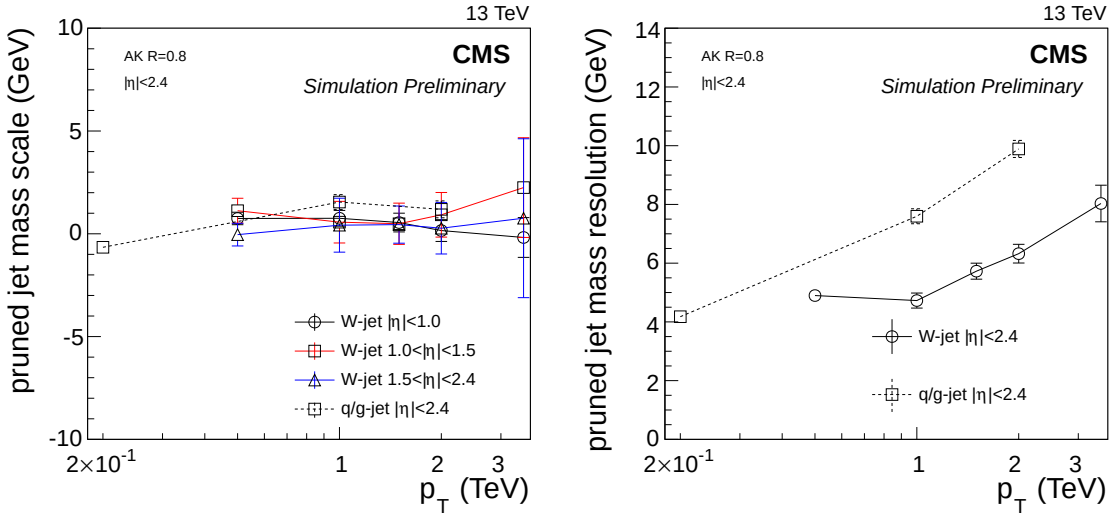

Figure 3.32: Reconstructed pruned jet mass scale (left) and resolution (right) of W-jets and quark/gluon-jets, with and average of 40 simulated pileup interactions.

while the fake rate remains roughly constant up to p_T of 3 TeV. When the pruned mass is combined with τ_2/τ_1, the efficiency drops of about 10-15% moving from 1 to 3.5 TeV, while the fake rate remains still approximately constant. This feature is attributed to jet properties and not to detector effects, since one observes that the generator-level distribution shifts as well as a function of p_T. Thus, the W-tagging performances are expected to be, differently from what observed during 8 TeV data taking, roughly stable up to jets with transverse momentum of 4 TeV, providing substructure observables across the full phase space covered by Run-II data.

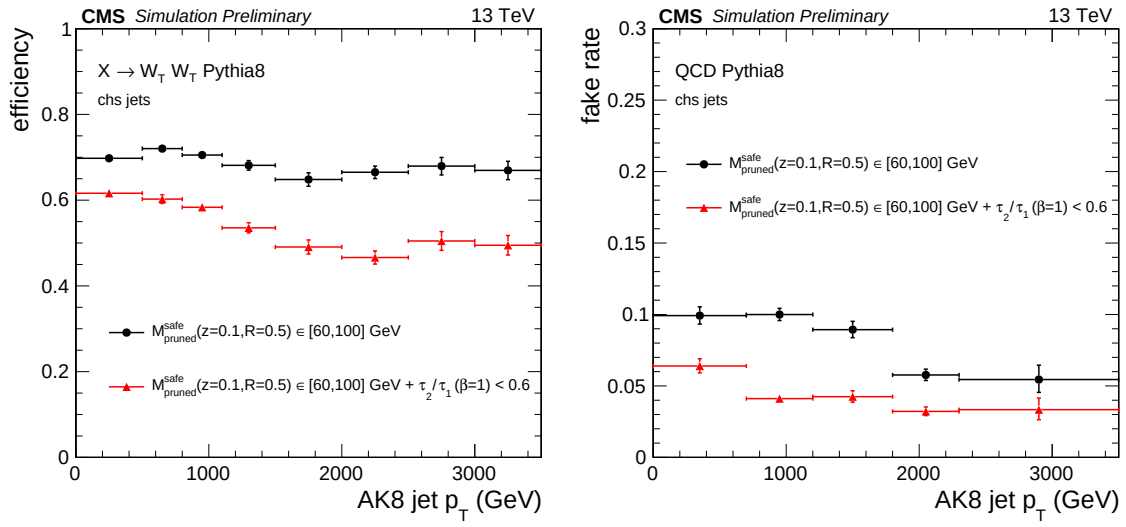

Figure 3.33: (Left) Efficiency of pruned jet mass $60 < m_{\text{pruned}} < 100$ GeV and N-subjettiness $\tau_2/\tau_1 < 0.6$ selection on W-jets as a function of p_T, with an average of 40 simulated pileup interactions. (Right) Fake rate for the pruned jet mass and τ_2/τ_1 selection on q/g-jets as a function of p_T.

3.4 PUPPI commissioning with first Run-II data

Several methods for dealing with pileup have been proposed and successfully applied by both ATLAS and CMS experiments on Run-I data. Many of them, such as four vector area subtraction [85], four vector safe subtraction [134], grooming techniques and pileup jet identification [82], follow a "global approach" trying to subtract pileup from high level objects which are obtained clustering the reconstructed particles into jets or during the clustering stage.
PileUp Per Particle Identification (PUPPI) attempts to combine global information of the event with local ones to identify pileup contamination at particle level, rather than just correct jet quantities, giving a new interpretation of the event filtering away pileup candidates [136]. In fact, PUPPI re-weights each particle four momentum by means of a measurement of its probability to be produced by a pileup interaction, estimated through a pileup sensitive metric (α). This leads to a pileup corrected event, where one can apply jet finding algorithms, evaluate jet shapes, compute lepton isolations and missing transverse energy ideally without any further pileup corrections. The algorithm works as follows:

- A local metric α, evaluated for each particle of the event, is used to distinguish hard scattering products from pileup radiation:

$$\alpha_i = \log \sum_{j}^{N_{evt}} \left(\frac{p_{T,j}}{\Delta R_{ij}} \right)^2 \Theta(R_{min} \leq \Delta R_{ij} \leq R_0) \tag{3.16}$$

  where N_{evt} is the number of particles reconstructed in the event, while ΔR_{ij} is the distance, in the η-φ plane, between the particle i and another PF candidate inside a R_0 cone around i, optimized to be $R_0 = 0.4$. In addition, particles closer to i than $R_{min} = 0.01$ are discarded to avoid divergences due to collinear emissions, while if no particles enter in the sum the infrared divergence is solved fixing α_i to zero, i.e. the particle i

is assumed to come from a pileup interaction. Due to the strong collinear emissions from the parton shower, a particle i belonging to the hard scattering is usually close to other ones, so that α_i tends to be large. In contrast, α_i tends to be smaller for pileup particles, since they originate mainly from soft QCD interactions and are uncorrelated with the hard scattering products.

- In the central region ($|\eta| < 2.5$), tracking information allow to distinguish charged particles from the leading vertex with respect to pileup ones. Thus, the density of primary vertex charged particles around a candidate i is used to increase the performance of the algorithm, changing the metric definition as:

$$\alpha_i = \log \sum_{j}^{N_{\mathrm{PV}}^{ch}} \left(\frac{p_{T,j}}{\Delta R_{ij}} \right)^2 \Theta(R_{min} \leq \Delta R_{ij} \leq R_0) \quad |\eta_i| < 2.5 \tag{3.17}$$

  where $\mathrm{N}_{\mathrm{PV}}^{ch}$ are the charged particles associated to the leading vertex of the event.

- To define a pileup-like weight for a particle i, the following χ^2 is computed:

$$\chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{\mathrm{PU}})^2}{\sigma_{\mathrm{PU}}^2} \Theta(\alpha_i - \bar{\alpha}_{\mathrm{PU}}) \tag{3.18}$$

  where $\bar{\alpha}_{\mathrm{PU}}$ represents the median of the event based α_i distribution evaluated considering only pileup particles, i.e. charged tracks originating from pileup interactions. Since the pileup density varies along η, a proper extrapolation, based on minimum bias interactions, is used to move from ($\bar{\alpha}_{\mathrm{PU}}$, σ_{PU}^2) evaluated in the central region to the forward one. Fluctuations below the median are considered to be pileup-like, assigning a weight equal to zero. Assuming α_i to be Gaussian distributed, the particle weight is defined as:

$$\omega_i = F_{\chi^2,\mathrm{ndf}=1}(\chi_i^2) \tag{3.19}$$

  Leading vertex information are almost ignored by the algorithm to minimize any sample dependance, resulting in a robust algorithm over different physics topologies.

- The probability, defined by Equation 3.19, is used to re-scale particles four momentum as $p_i^\mu \rightarrow \omega_i p_i^\mu$. Charged particles compatible with the primary vertex are not re-weighted ($\omega_{\mathrm{PV}}^{ch} = 1$), while charged pileup candidates are removed ($\omega_{\mathrm{PU}}^{ch} = 0$). Particles with a weight ω_i smaller than 1% are rejected, as well as neutral soft ones. The selection on the scaled p_T of neutral candidates is tuned in three different η regions, following the changes in the CMS granularity, as a function of the number of reconstructed vertices ($\mathrm{N_{PV}}$) to keep both jet p_T and mass responses close to unity against pileup:

$$\begin{cases} p_T > (0.1 + 0.015 \cdot \mathrm{N_{PV}}) \text{ GeV} & \text{for : } |\eta| < 2.5 \\ p_T > (1.5 + 0.070 \cdot \mathrm{N_{PV}}) \text{ GeV} & \text{for: } 2.5 < |\eta| < 3.0 \\ p_T > (2.0 + 0.070 \cdot \mathrm{N_{PV}}) \text{ GeV} & \text{for: } |\eta| > 3.0 \end{cases} \tag{3.20}$$

3.4.1 Performances on large cone jets

PUPPI performances are investigated first on large cone jets in simulated events, clustering the reconstructed PF candidates via the anti-k_t algorithm with $R = 0.8$. For both PF and PF+CHS jets, four momenta are corrected with the safe 4-vector area subtraction [134]. The leading AK8 jets in simulated RS graviton events, decaying into a pair of W-bosons, are considered as W-jet candidates (signal jets), while multi-jet production is taken as source of high p_T QCD jets. Only events with at least one reconstructed AK8 jet with p_T in the range 200-600 GeV are considered.
Figure 3.34 shows the leading jet mass response with respect to the generator level prediction for different pileup mitigation algorithms, where PUPPI provides the best mass response and resolution for both signal and background jets. Moreover, the W-jet mass peak and resolution dependence on pileup is reported in Figure 3.35. PF+CHS shows a stable response against pileup, but it suffers from the largest bias in the reconstructed mass, while PUPPI provides the best mass resolution as well as a good stability as a function of N_{PV} in both scale and resolution, which can be improved by a further tunings of the algorithm parameters.

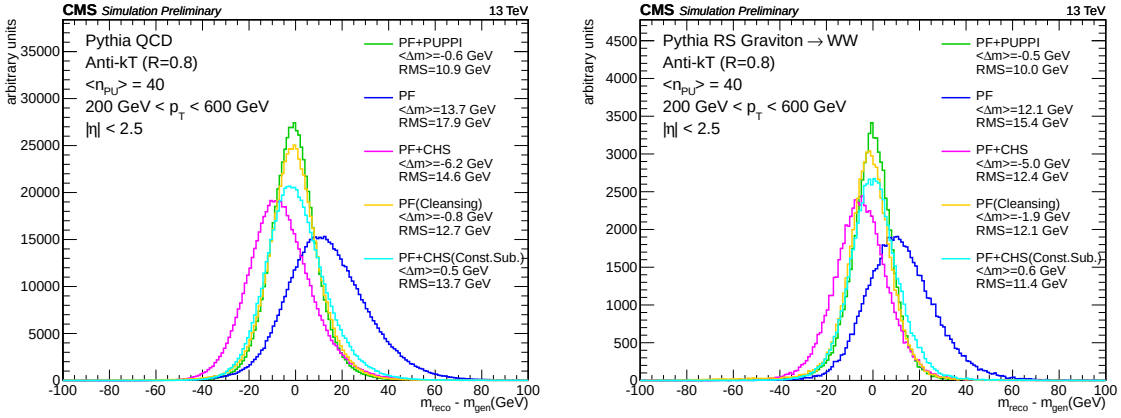

Figure 3.34: Comparison of the leading AK8 jet mass response, with respect to the generator level prediction, for different algorithms in simulated QCD multi-jet events (left) and for W-jets produced by the decay of RS gravitons (right).

After this, PUPPI performance on high p_T jets has been tested looking at the first Run-II collisions at $\sqrt{s} = 13$ TeV, performed with 50 ns buch spacing. Data are collected from a set of multi-jet and H_T triggers, for a total integrated luminosity of about 40 pb^{-1}. Events are selected offline requiring at least one reconstructed AK8 jet with $p_T > 100$ GeV and $|\eta| < 1.3$, which consists in selecting a sample enriched of high p_T QCD jets. Due to the trigger pre-scales, simulated multi-jet events are normalized to the measured yields in data. Dedicated jet energy correction are applied on both data, including residual AK8 corrections, and simulation for PF+CHS jets, while PUPPI jets are not calibrated since no AK8 PUPPI jet corrections were available so far. Figure 3.36 shows the AK8 jet mass distribution (left) as well as the dependence of the mean jet mass value as a function of N_{PV} (right). PUPPI shows both stable performances against pileup and lower jet mass values, close to the Sudakov mass peak produced by the parton shower. A reasonable agreement between data and simulation is observed for the jet mass shape, where discrepancies in case of PUPPI jets will be mitigated once jet energy residual corrections will be applied.

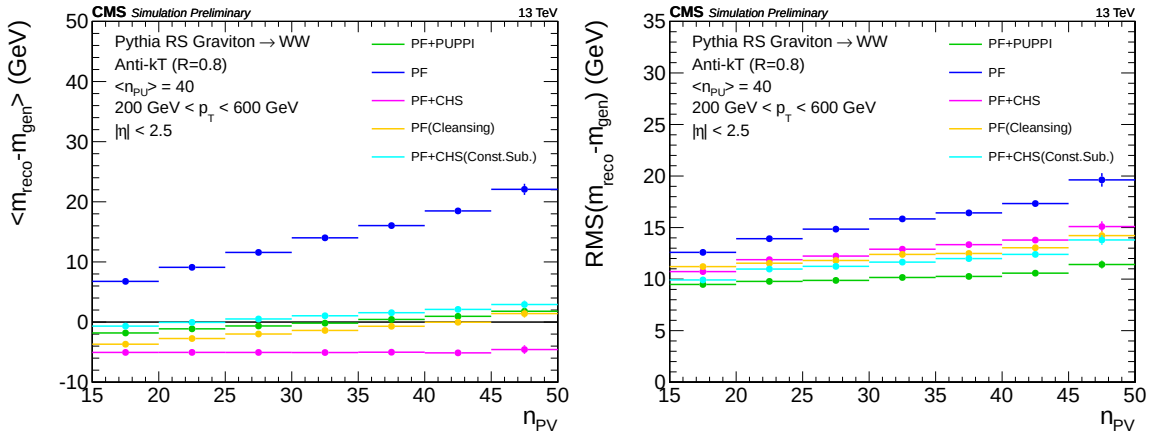

Figure 3.35: Mass response $\langle m_{\text{reco}} - m_{\text{gen}} \rangle$ (left) and resolution (right), defined as the RMS of the $m_{\text{reco}} - m_{\text{gen}}$ distribution, for simulated W-jets as a function of the number of reconstructed vertices.

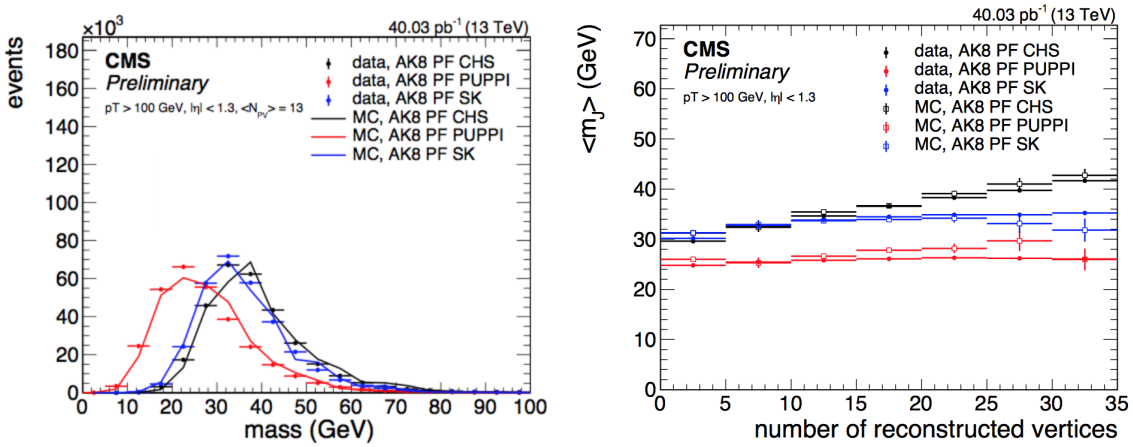

Figure 3.36: Comparison between data and simulation for the jet mass distribution (left) and the jet mass pileup dependance (right) for leading AK8 jets with $p_T > 100$ GeV and $|\eta| < 1.3$, reconstructed by PF+CHS and PUPPI in a QCD enriched sample.

3.4.2 Performances on missing transverse energy

The missing transverse energy measurement plays a crucial role in many searches for new physics at the LHC, representing the primary experimental signature for supersymmetric and dark matter candidates. The $\vec{\not{E}}_T$ evaluation is sensitive to pileup contamination, as well as to various detector and reconstruction effects resulting in mis-measurements of particle transverse momenta. To investigate the performance of PUPPI on the $\vec{\not{E}}_T$ scale and resolution, a pure sample of $Z/\gamma^* \to \mu\mu$ events is used, which is characterized by a well defined reference object $\vec{p}_T(\mu\mu)$ and $\vec{\not{E}}_T$ expectation. The bulk of $Z/\gamma^* \to \mu\mu$ events contain no genuine $\vec{\not{E}}_T$, where the hadronic balance to the well measured vector boson dominates the measured $\vec{\not{E}}_T$ properties. In fact, noise in the calorimeters, particle mis-reconstruction, finite detector and jet energy resolution can generate large $\vec{\not{E}}_T$ in the event, as well as the presence of pileup and underlying event activity.

These $Z/\gamma^* \to \mu\mu$ events are collected from an online trigger which requires the presence of

two identified muons with p_T larger than 17 and 8 GeV, respectively. Only certified runs at $\sqrt{s} = 13$ TeV with 50 ns bunch spacing are considered, for a total integrated luminosity of about 42 pb^{-1}. Two opposite charged muon candidates must be reconstructed both in the tracker and in the muon chambers, with transverse momentum $p_T > 25\,(20)$ GeV and $|\eta| < 2.4$ to ensure to be on the trigger efficiency plateau, and isolated from additional activity in the detector to veto muon candidates from non-prompt sources. Eventually, additional loosely identified leptons (e or μ) are vetoed and the selected di-muon pair is required to have an invariant mass around the Z-boson one, between 60-120 GeV.
Simulated $Z/\gamma^* \to \ell\ell$ events ($\ell = e, \mu, \tau$) are generated via `amc@nlo` at `NLO QCD` [114], using `NNPDF3.0` [137] and interfaced with `Pythia8`. Background contaminations come from fully leptonic $t\bar{t}$ events, which are generated with `POWHEG+PYTHIA8` and WZ/ZZ productions simulated via `POWHEG` and `Pythia8`, respectively. For all processes, the detector response is modelled using a detailed description of the `CMS` experiment based on `GEANT4`. The selected simulated events are re-weighted to represent the pileup distribution as measured in the data.
Jets are clustered from reconstructed `PF` candidates through the anti-k_t algorithm with $R = 0.4$, while `PUPPI` jets are obtained applying the same clustering on the particles produced by the `PUPPI` algorithm. Both `PF` and `PUPPI` jets four momenta are corrected for by dedicated jet energy calibrations as described in Section 2.4.4.
Particle-flow and `PUPPI` $\vec{\not{E}}_T$ are corrected for jet energy calibration (`Type-I`) according to Equation 2.9, where the threshold on the corrected jet p_T has been optimized to be 15 (20) GeV for `PF` (`PUPPI`) jets. Particle-flow candidates with $|\eta| > 3$ are not considered in the $\vec{\not{E}}_T$ evaluation, since `HF` calibration is not commissioned for the analyzed data.
The identified muon pair is removed from the event before running the `PUPPI` algorithm. In fact, since in the selected events muon candidates are required to originate from the leading vertex, pileup particles close to them will be interpreted by `PUPPI` as primary vertex like, causing a slightly degradation of the $\vec{\not{E}}_T$ response and resolution in events with low boosted Z-bosons, where the $\vec{\not{E}}_T$ measurement is dominated by pileup.
Figure 3.37 (left) shows a nice agreement between data and simulation for the invariant mass distribution of the selected $Z/\gamma^* \to \mu\mu$ candidates, while, on the right, the comparison is presented for the measured $\not{E}_T$ obtained from the `PUPPI` algorithm.
Using the di-muon transverse momentum as reference object $\vec{p}_T(\mathrm{Z})$, the hadronic recoil $\vec{u}_T$ is defined as the vector sum of the transverse momentum of all the particles in the event, except the vector boson candidate. The hadronic recoil is then projected with respect to the $\vec{p}_T(\mathrm{Z})$ axis, yielding a parallel ($u_{||}$) and a perpendicular ($u_\perp$) component. These kinematic definitions and related projections are displayed in Figure 3.38.
Figure 3.39 (left) shows a comparison between data and simulation for the $u_\perp$ distribution, computed starting from the `PUPPI` recoil. It is expected to be centered at zero and symmetric since it arises mainly from pileup and random detector effects. At the same time, the $|\vec{p}_T|(\mathrm{Z}) + u_{||}$ distribution is reported in Figure 3.39 (right), which is also expected to be centered around zero but particularly sensitive to the jet energy scale.

3.4.2.1 $\vec{\not{E}}_T$ performance in simulated $Z/\gamma^* \to \mu\mu$ events

Particle-flow $\vec{\not{E}}_T$ performances are compared to `PUPPI` ones on the basis of $Z/\gamma^* \to \ell\ell$ simulated events, which include the small contamination from leptonically decaying τ-leptons. The

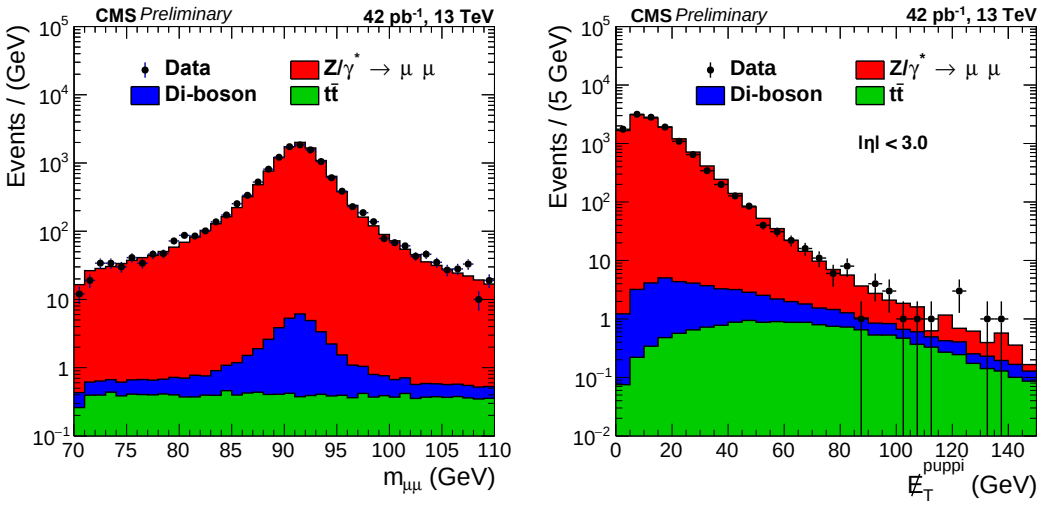

Figure 3.37: (Left) Di-lepton invariant mass distributions for events surviving the $Z/\gamma^* \to \mu\mu$ selections in both data and simulation. (Right) Comparison between data and MC of the measured PUPPI $\vec{\not{E}}_T$ in a $Z/\gamma^* \to \mu\mu$ enriched phase space.

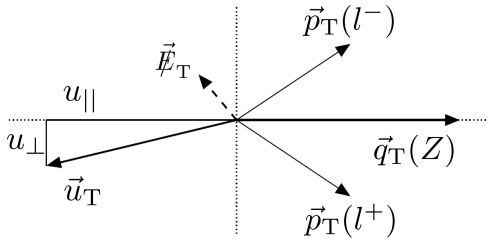

Figure 3.38: Illustration of $Z/\gamma^* \to \ell\ell$ event kinematics in the transverse plane. The vector $\vec{u}_T$ denotes the vectorial sum of the transverse momentum of the all particles reconstructed in the event except for the two leptons produced by the Z-boson decay.

$\vec{\not{E}}_T$ response, defined as -$\langle u_{||} \rangle / p_T(\mathrm{Z})$ as a function of $p_T(\mathrm{Z})$, is shown in Figure 3.40 (left) for both raw and Type-I corrected missing transverse energy. Type-I corrected recoils show a stable plateau around 85% for $p_T(\mathrm{Z})$ larger than 50 GeV, where the bias from unity originates from exploiting only particles reconstructed within $|\eta| < 3$ in the missing energy estimation. In contrast, raw PUPPI response show a much larger bias than PF one. Below 50 GeV, the uncorrected unclustered energy contribution, due to particles not contained within jets, starts to be significant compared to the one from recoiling jets, leading to an underestimation of the response. The bias in the low boosted region is much larger for PUPPI than PF, which is probably too aggressive in removing soft and isolated neutral particles produced by the parton shower. Furthermore, the larger bias is also related to a higher p_T threshold adopted for PUPPI $\vec{\not{E}}_T$ Type-I correction, which is motivated by a not well tuned set of corrections for low p_T PUPPI jets. At the same time, Figure 3.40 (right) reports the $\vec{\not{E}}_T$ response as a function of the number of reconstructed vertices for events with $p_T(\mathrm{Z}) > 50$ GeV, where either PF or PUPPI show a stable performance.

The transverse resolution in bins of $p_T(\mathrm{Z})$ (left) and $\mathrm{N_{PV}}$ (right) is shown in Figure 3.41, which is estimated as the width of Gaussian fits to the $u_\perp$ distribution. The perpendicular resolution ($\sigma_\perp$) slightly increases as a function of $p_T(\mathrm{Z})$, which is much more evident for PUPPI than PF

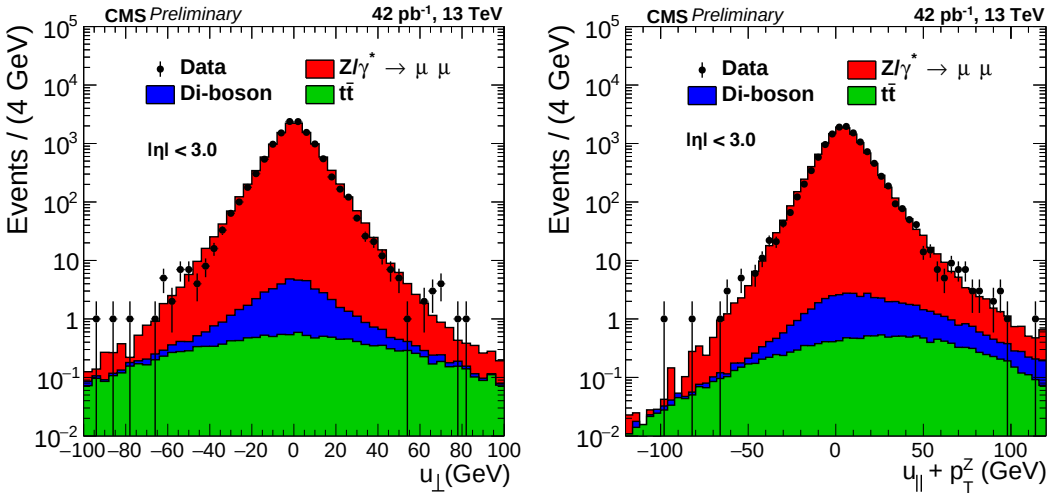

Figure 3.39: Distributions of $u_\perp$ (left) and $|\vec{p}_T|(\mathrm{Z}) + u_{||}$ (right) for the PUPPI recoil in the selected $Z/\gamma^* \to \mu\mu$ events for both data and simulation.

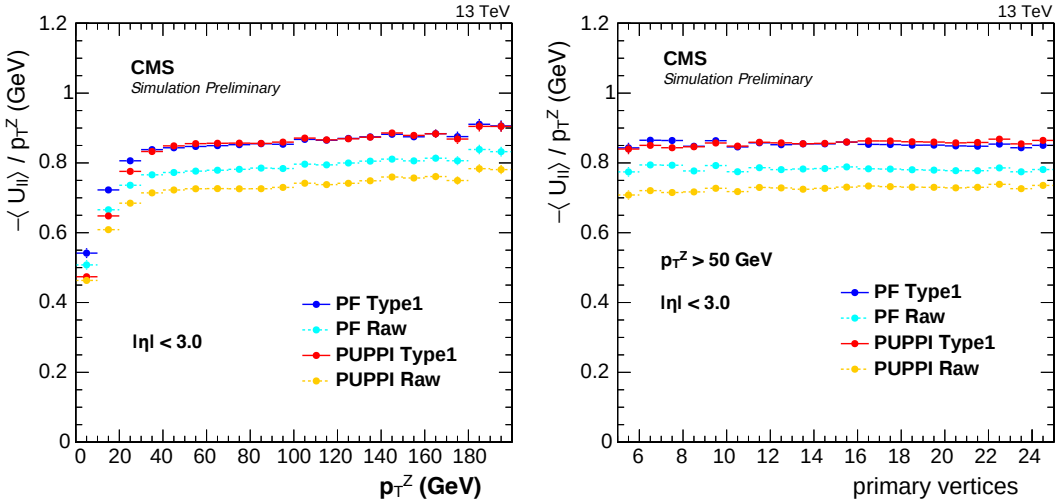

Figure 3.40: Response curves as a function of $p_T(\mathrm{Z})$ (left) and $\mathrm{N_{PV}}$ (right) for both PF and PUPPI $\vec{\not{E}}_T$ in simulated events where a Z-boson decays into a muon pair.

$\vec{\not{E}}_T$. PUPPI provides an important gain in resolution, from 40% to 20% when $p_T(\mathrm{Z})$ ranges between 40-200 GeV. In addition, PUPPI provides an almost stable performance as a function of pileup, while PF perpendicular resolution degrades by 40% moving from 6 to 24 additional pileup interactions, considering only events lying on the response plateau ($p_T(\mathrm{Z}) > 50$ GeV). The degradation in the resolution after applying the Type-I correction is correlated to the increase in the $\vec{\not{E}}_T$ response.

Eventually, Figure 3.42 shows the same trends for the resolution of the longitudinal recoil component ($u_{||}$). Also in this case, PUPPI is better performing than the standard particle-flow measurement especially for events with a low boosted Z-boson candidate. Then, PUPPI performance approches asymptotically the PF one since the resolution becomes dominated by the one of the hard recoiling jets. Finally, PUPPI $u_{||}$ resolution shows a stable performances as a

function of pileup, while PF one degrades by 30% when N_{PV} ranges from 5 to 25.

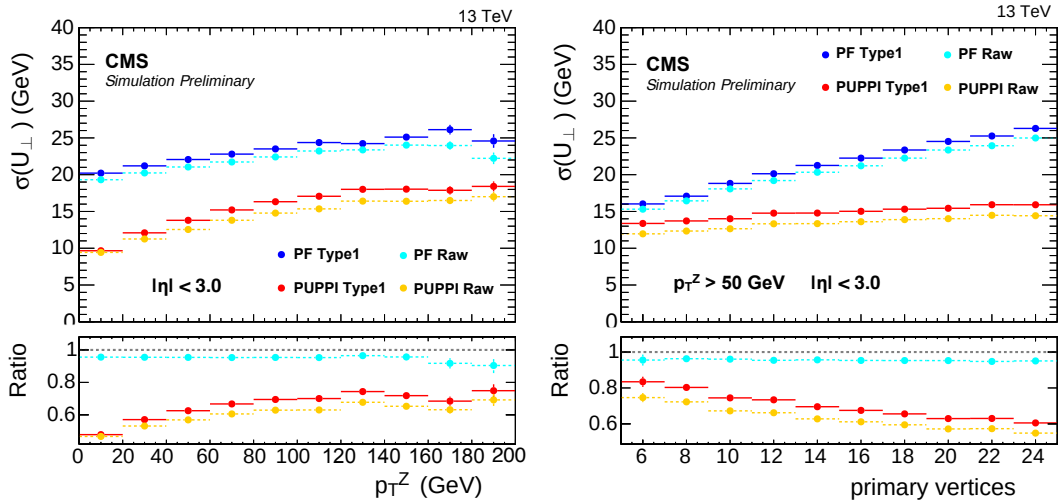

Figure 3.41: Resolution on the perpendicular recoil component ($u_\perp$) as a function of p_T(Z) (left) and the number of primary vertex (right) for simulated events where a Z-boson decays into a muon pair. The lower frame shows the ratio of alternative recoil resolutions with respect to the PF Type-I estimate.

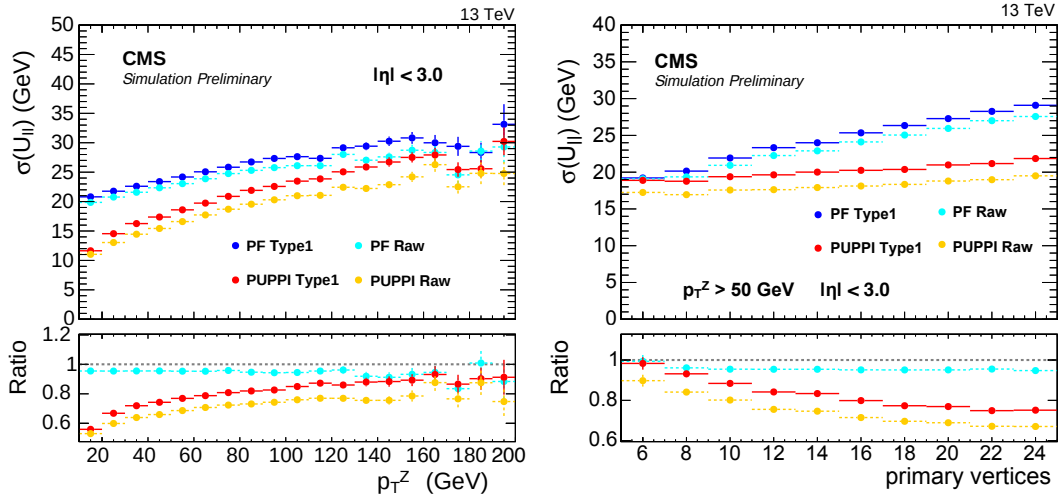

Figure 3.42: Resolution on the longitudinal recoil component ($u_{||}$) as a function of p_T(Z) (left) and the number of primary vertex (right) for simulated events where a Z-boson decays into a muon pair. The lower frame shows the ratio of alternative recoil resolutions with respect to the PF Type-I estimate.

3.4.2.2 PUPPI $\vec{\not{E}}_T$ performance in data

As already shown in Figure 3.37 and Figure 3.39, a good agreement between data and simulation for PUPPI $\vec{\not{E}}_T$, in terms of both shape and absolute normalization, is observed, although the comparison for p_T(Z) > 100 GeV is limited by statistics. Background contamination is expected to be small except for events belonging to the upper tail of the $\vec{\not{E}}_T$ distribution, as reported in Figure 3.38. Figure 3.43 shows a comparison between data and MC for both PF and PUPPI missing transverse energy, which proves how the PUPPI algorithm is able to reduce the measured $\vec{\not{E}}_T$ consistently between data and simulation, even for events in the tail.

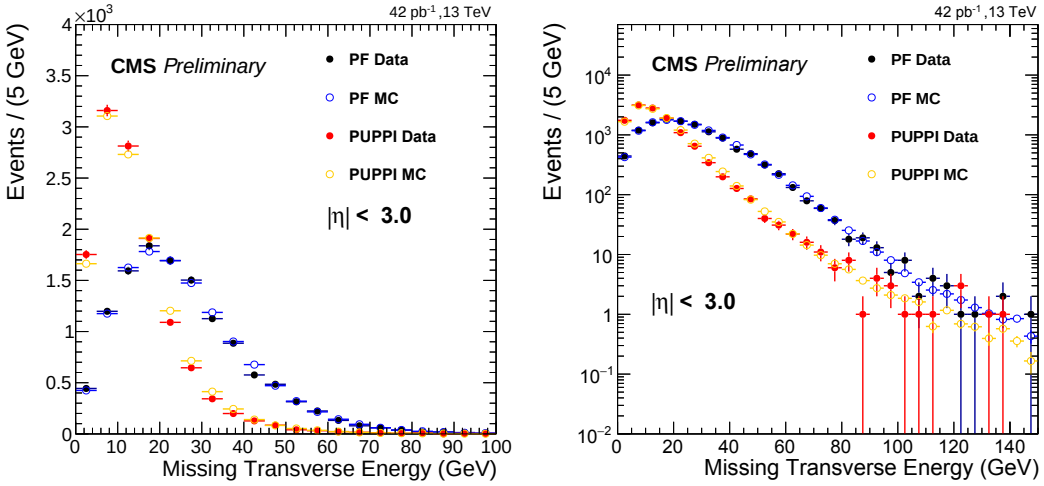

Figure 3.43: $\vec{\not{E}}_T$ distributions as obtained in data and simulation through the PF and the PUPPI algorithms. On the left, distributions are reported in linear scale, while, on the right, in logarithmic one to better compare tail events.

Moreover, Figure 3.44 shows the pileup dependance of PF and PUPPI transverse recoil resolution for events in which the selected Z-boson p_T is larger than 50 GeV. In this case, the resolution is corrected for the non unity $\vec{\not{E}}_T$ response in each N_{PV} bin, to recover the correct recoil energy scale. A nice agreement is visible between data and simulation for both PF and PUPPI quantities and, as already proved by simulation only studies, PUPPI is able to mitigate at best pileup effects yielding to a stable performance as a function of pileup. Eventually, PUPPI $\vec{\not{E}}_T$ will represent an important tool for the next Run-II collisions, where the instantaneous luminosity is going to increase providing an average of 30-40 in-time pileup interactions per bunch crossing.

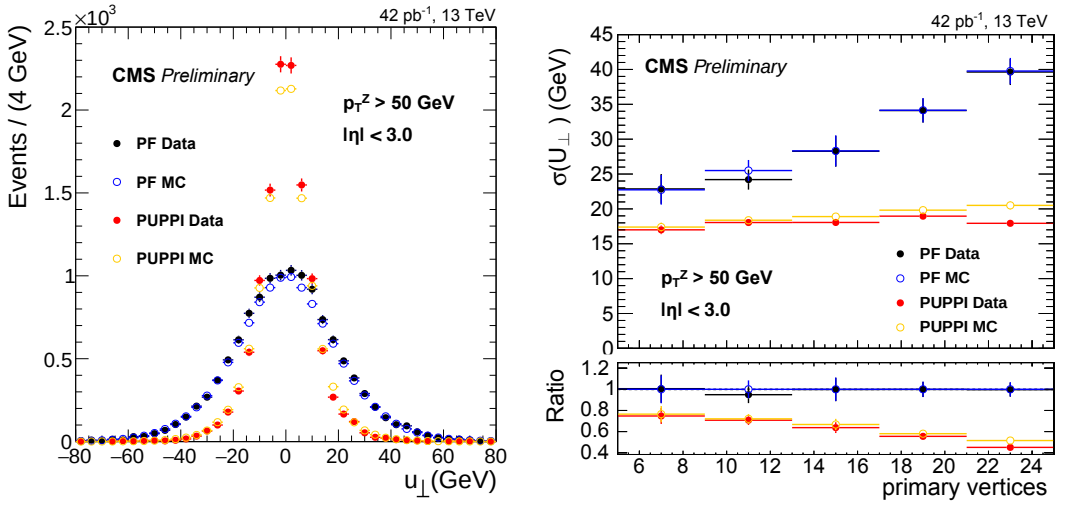

Figure 3.44: Comparison between data and simulation for the measured perpendicular $u_\perp$ recoil and its resolution, obtained from both PF and PUPPI, as a function of the number of primary vertexes in the selected $Z\gamma^* \to \mu\mu$ events. The lower frame shows the performance of the different methods relative to the PF recoil resolution as extracted from simulated events.

Chapter 4

Search for exotic particles decaying into semi-leptonic V-boson pairs

In this Chapter, a search for new resonances decaying into pairs of vector bosons (WW, WZ, ZZ) are presented, in final states where one of the two V-boson decays leptonically, while the other one goes into hadrons. Results are based on the data recorded in pp collisions at $\sqrt{s} = 8$ TeV by the CMS experiment [138], corresponding to an integrated luminosity of 19.7 fb^{-1}. In many theoretical extensions of the SM, the spontaneous breaking of the electroweak symmetry is associated with a new strong dynamic appearing at the TeV scale, which predicts the presence of new heavy resonances coupled with massive vector bosons. As described in Section 1.5, the production of Kaluza-Klein (KK) excitations of spin-2 bosons, in particular Randall-Sundrum and Bulk gravitons, decaying into vector boson pairs can be sizable and detected at hadron colliders.
The final states considered in this study are either $\ell\nu q\bar{q}'$ or $\ell^+\ell^- q\bar{q}$, resulting in events with a charged lepton, a neutrino and a single boosted V-jet ($\ell\nu$+V-jet channel) or two same flavor opposite charged leptons and a single V-jet ($\ell\ell$+V-jet channel). In particular, the search is limited to final states where $\ell = \mu$ or e, implicitly including the cases where W $\to \tau\nu$ or Z $\to \tau\tau$ and the τ-lepton is then decaying leptonically ($\tau \to \ell\nu\nu$).
Even if my work was focused on the single lepton channel, the $\ell\ell$+V-jet analysis is also reported since I participated to the $\ell\ell$+V-jet/$\ell\nu$+V-jet combination, to part of the software development and for a better description of the physics context.
Only the gluon fusion production has been considered, since its predicted cross section is about ten times larger than the VBF one, as shown in Figure 1.9. Figure 4.1 shows two Feynman diagrams related to the gluon fusion production and decay of a generic resonance (X). The analysis is optimized for resonances decaying into WW and ZZ pairs, although the requirement on the pruned jet mass and the V-jet mass resolution brings it to be also sensitive to charged resonances decaying into WZ pairs.
In Section 4.1, an overview of the simulated samples, datasets and triggers used in the analysis is presented. Then, Section 4.2 provides a description of the object reconstruction and the event selection, while Section 4.3 is dedicated to signal modeling. Section 4.4 and Section 4.5 describe, respectively, the background estimation and the determination of the systematic uncertainties considered in the ultimate statistical interpretation.

Eventually, the result of a search for Bulk gravitons and its model independent re-interpretaion are presented in Section 4.6. In the former case, these semi-leptonic analyses are combined with limits derived in a companion CMS search for resonances decaying in the fully hadronic channel [123].

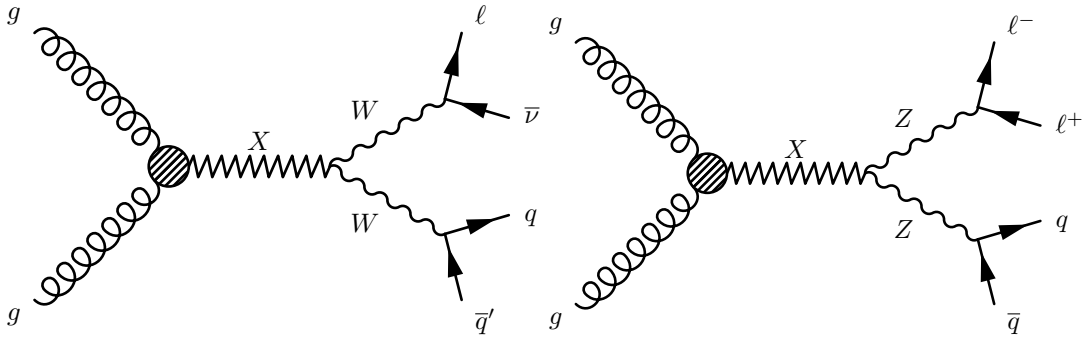

Figure 4.1: Two Feynman diagrams for the production of a generic resonance X decaying into the final states considered in this study: (left) $\ell\nu$+V-jet channel, (right) $\ell\ell$+V-jet one.

4.1 Simulated samples, datasets and triggers

The RS1 and Bulk graviton models are used as benchmark signal processes, with the graviton forced to decay into WW or ZZ pairs. In the former case, vector bosons are produced with a transverse polarization (V_T) in more than 90% of cases, while in the latter mostly a longitudinal one. The considered mass hypotheses lie in the range between 600 and 2500 GeV.
Bulk gravitons are generated with JHUGen 3.1.8 interfaced with Pythia 6, while the RS1 model is fully simulated via Pythia 6 or with MadgGraph interfaced with Pythia 6. Figure 4.2 shows the $pp \to G_{bulk} \to WW(ZZ) \to \ell\nu(\ell\ell) + qq'$ cross section, calculated at LO with CalcHEP 3.4.1 [139], as a function of the graviton mass and for $\tilde{k} = (0.2, 1)$.
A summary of the different signal samples considered in the analysis is listed in Table 4.1.

Model	Generator	$\tilde{k} = k/m_{pl}$	m_G (TeV)	Γ_G/m_G (%)
Bulk	JHUGen	0.2,0.5	0.6-2.5 TeV in steps of 0.1 TeV	$\Gamma_G \ll 1 \ll \sigma_{exp}$
Bulk	JHUGen	0.2,0.5	[1, 1.5, 2.1] TeV	[5, 15, 30]%
RS1	Pythia 6	0.05	[0.75, 1, 1.25, 1.5, 1.75, 2] TeV	$\Gamma_G \ll 1 \ll \sigma_{exp}$
RS1	MadGraph	0.2	[0.6, 1, 1.5] TeV	$\Gamma_G \ll 1 \ll \sigma_{exp}$

Table 4.1: List of signal samples used in the analysis, for both Bulk and RS1 scenarios, as a function of graviton mass, $\tilde{k}$ and resonance width. Events have been generated for both semi-leptonic WW and ZZ final states, taking into account vector boson decays into the three possible lepton flavors (e,μ,τ).

The adopted background samples are already listed in Section 3.2.1. In the $\ell\nu$+V-jet analysis, the largest background contribution is represented by a leptonically decaying W-boson produced in association with jets (W+jets), while for the $\ell\ell$+V-jet channel the largest one corresponds to Drell-Yan (DY+jets) pairs produced in association with jets.
The 19.7 fb^{-1} data used in the analysis are obtained from the event reconstruction performed after the end of Run-I data taking.
In the $\ell\nu$+V-jet channel, signal candidates are selected online with a trigger requiring either

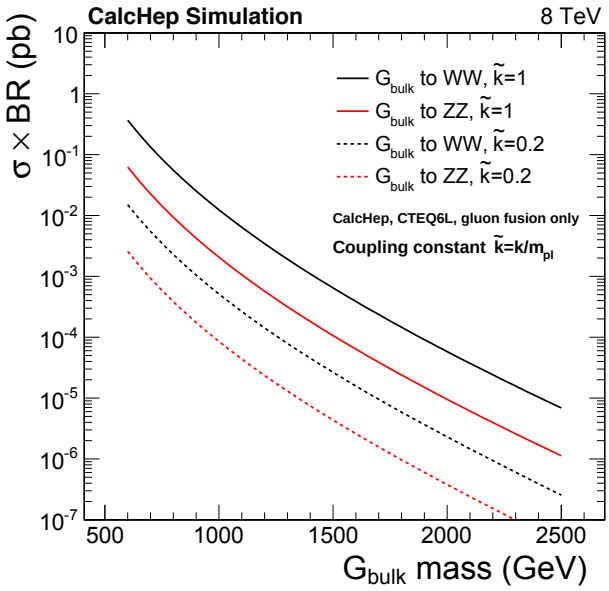

Figure 4.2: The cross-sections for the $pp \to \mathrm{G}_{bulk} \to \mathrm{WW(ZZ)} \to \ell\nu(\ell\ell)+qq'$ process at a centre-of-mass energy of 8 TeV calculated by `CalcHEP` at `LO`.

one muon or one electron, without any isolation requirements and applying a loose identification criterion. The transverse momentum, measured online, must be larger than 40 GeV for muons, while the minimum transverse energy threshold is 80 GeV for electrons. These triggers have been designed to be fully efficient for `BSM` resonance searches with mass above 1 TeV. The trigger selection is not applied on simulated events, thus its efficiency, evaluated with respect to the offline lepton selection criteria, is estimated from data with a tag-and-probe technique on $\mathrm{Z} \to \mu\mu$ and $\mathrm{Z} \to ee$ candidates, as described in [140]. Simulated events are corrected accordingly using the measured trigger efficiency as an event weight. Single-muon trigger efficiency varies from 82% to 94% as a function of η, while the efficiency for the single-electron one is above 98%.
In the $\ell\ell$+V-jet channel, events are selected online with a trigger requiring either two muons or two electromagnetic energy deposits (`ECAL` super-clusters), applying a loose identification criteria in both cases. No lepton isolation requirements are applied at the trigger level. The p_T thresholds are 22 GeV for the leading muon and 8 GeV for the trailing one, while the transverse energy threshold for the `ECAL` clusters is set at 33 GeV. The efficiency of the double-muon trigger varies between 80% and 98%, as a function of η, while the double-electron one is above 99% [141] over the whole (p_T,η) spectrum.

4.2 Event selection and reconstruction

All the events, selected by the chosen online triggers, are required to have at least one primary vertex reconstructed within a 24 cm window, with respect to the estimated interaction point, along the beam axis (d_Z) and with a transverse distance of less than 2 cm (d_0). In addition, each fitted vertex is required to have a number of degrees of freedom larger than four [71]. If more than one reconstructed vertex is surviving these requirements, the primary interaction vertex is selected to be the one with highest $\sum p_T^2$, summing over all its associated tracks.

The selected simulated events are re-weighted to represent the pileup distribution as measured in the data. Different strategies for lepton selection and identification are applied in the WW and ZZ channels:

- **$\ell\nu$ + V-jet channel**: muons are identified by means of the high-p_T selection, while for electrons the `HEEP` selection is considered, as already described in Section 3.2.2. Furthermore, a veto on additional leptons is also required to reduce the contamination from `DY`+jets events.

- **$\ell\ell$ + V-jet channel**: when a resonance with a large mass decays into Z-boson pairs, the two charged leptons produced in the decay are highly collimated because of the Lorentz boost, with a consequent small angular separation (ΔR).
  The same selection criteria adopted for the $\ell\nu$ + V-jet channel are also used to identify lepton pairs in the ZZ one, except for muons. In this case, one of them is required to be reconstructed by the global algorithm, while, on the other leg, a looser selection is applied requiring only a muon track.
  In addition, since lepton reconstruction and identification at high p_T are optimized for the case of well separated leptons, a loss of efficiency is observed when two muons or two electrons are close to each other, due to a mutual overlap inside the same isolation cones. For this reason, the contribution from any other muon or electron candidate inside it has to be excluded, to retain high selection efficiency also in the boosted configuration. Figure 4.3 (left) shows the relative isolation distribution with and without the lepton veto inside the isolation cone.
  Events with at least two opposite charged muons with $|\eta| < 2.4$ and $p_T > 40$ (20) GeV are considered to define the muon channel, while events with two opposite charged electrons with $p_T > 40$ GeV and $|\eta| < 2.5$ are selected in the electron one. Figure 4.3 (right) shows the ultimate di-muon selection efficiency as a function of the Z-boson p_T for different leading muon η bins.

Jets used to identify the hadronically decaying W and Z-bosons are clustered using the Cambridge-Aachen algorithm, with a distance parameter $R = 0.8$, following the outcome of the V-jet tagging studies described in Chapter 3.
To identify b-jets in order to to reduce the $t\bar{t}$ background in the $\ell\nu$ + V-jet channel, anti-k_t jets, clustered with a distance parameter $R = 0.5$, are considered and the combined secondary vertex b-tagging algorithm is applied. The ratio between b-tagging efficiencies measured in data and in simulated events is used as a scale factor to correct the simulation prediction. Dedicated jet energy corrections are applied to both `AK5` and `CA8` jets, where additional quality criteria are used to remove spurious jet-like features originating from isolated noise patterns in the calorimeters or in the tracker system. The efficiency of these requirements for real prompt jets is above 99%. Furthermore, the `CA8` (`AK5`) jets are required to be separated from any identified electron or muon by $\Delta R > 0.8(0.3)$.
The missing transverse energy $\not\!\!E_T$, modified to account for `AK5` jet calibration (`Type-I` correction) and corrected for the observed ϕ-modulation as described in Section 2.4.5, is required to be larger than 40 (80) GeV for the muon (electron) channel in the $\ell\nu$+V-jet analysis, while no selections are applied in the $\ell\ell$+V-jet one.

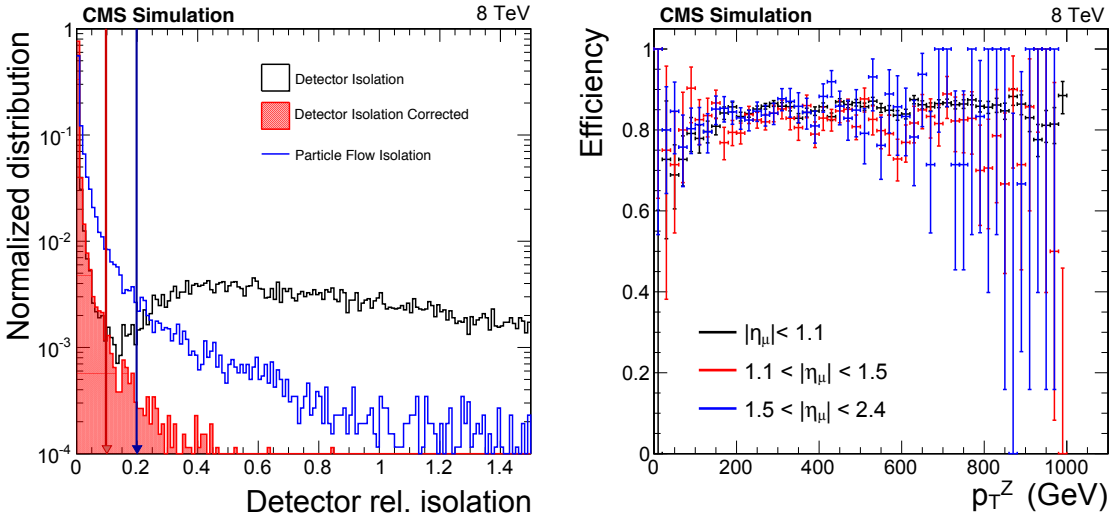

Figure 4.3: (Left) Isolation variables for muons identified in simulated signal events (`Bulk` Graviton with $m_{\mathrm{G}} = 1.5$ TeV): (black) standard track-based isolation, (blue) `PF` isolation with pileup correction, as defined in Equation 3.9, and (red) tracker based isolation, modified removing additional leptons inside the isolation cone. (Right) Selection efficiency obtained requiring one global and one tracker high-p_T muons as function of Z-boson p_T and for different intervals of leading muon η.

4.2.1 W$\to \ell\nu$ and Z$\to \ell\ell$ reconstruction and identification

In the $\ell\nu$+V-jet channel, the identified electron or muon is associated with the lepton produced by the W-boson decay ($\mathrm{W} \to \ell\nu$). The transverse momentum of the undetected neutrino is therefore assumed to be equal to the missing energy $\not\!\!E_T$. The longitudinal component of the neutrino four-momentum (p_z^ν) is obtained by constraining the lepton-neutrino invariant mass to be equal to the known W-boson one [142], which leads to the following second order equation:

$$m_\ell^2 + m_\nu^2 + 2 \cdot (E_\ell E_\nu - p_x^\ell p_x^\nu - p_y^\ell p_y^\nu - p_z^\ell p_z^\nu) = m_{\mathrm{W}}^2 = (80.4)^2 \text{ GeV}^2 \tag{4.1}$$

In case of two real solutions for p_z^ν, the one with the smaller absolute value is chosen, while if there are two complex solutions, only the real part is considered. The same procedure is also applied for $\mathrm{W} \to \tau\nu$ candidates, where the τ-lepton decays to one electron or muon and two neutrinos. In this case, the $\not\!\!E_T$ represents the transverse momentum of the three-neutrino system.

In the $\ell\ell$+V-jet channel, the leptonic Z-boson candidate is reconstructed by combining two opposite charged lepton candidates, and the invariant mass of the di-lepton system is required to be consistent with the Z-boson one, in the range between 70 and 110 GeV. This allows for a significant reduction of Drell–Yan and top-quark backgrounds, at the cost of suppressing the small signal contribution from $\mathrm{Z} \to \tau\tau \to (\ell\nu\nu)(\ell\nu\nu)$.

4.2.2 W$\to \mathrm{q\bar{q}'}$ and Z$\to \mathrm{qq}$ identification using substructure

In $\mathrm{X} \to \mathrm{WW}$ events, the boost factor γ of the W-boson is approximately given by $m_{\mathrm{X}}/2m_{\mathrm{W}}$ [143]. A large boost factor implies that the decay products will be predominantly emitted with a small angular distance $\Delta R \approx 2m_{\mathrm{W}}/p_T^{\mathrm{W}}$, which means that for resonances with mass greater

than 1 TeV the boosted topology is by far the dominant one. The hadronic W or Z-boson candidate is therefore reconstructed as a single massive CA8 jet and substructure techniques, described in detail in Chapter 3, are adopted to identify and select the hadronically decaying boosted V-jets.

4.2.3 Final event categorization

After reconstructing the two vector bosons, in the $\ell\nu$+V-jet ($\ell\ell$+V-jet) channel both leptonic and hadronic V-boson candidates must have a p_T larger than 200 (80) GeV. Higher thresholds are used in the $\ell\nu$+V-jet case because of the larger online single-lepton trigger thresholds, as well as the larger multi-jet background contamination in the low boosted regime.
In contrast, the looser selections adopted in the $\ell\ell$+V-jet channel allow to access to lower masses, even if the acceptance of the boosted category is not optimal for low boosted Z-jets. Thus, the minimum requirement on m_{VV} is equal to 700 (500) GeV for the $\ell\nu$+V-jet ($\ell\ell$+V-jet) channel.
To further reduce the presence of the semi-leptonic $t\bar{t}$ background in the $\ell\nu$+V-jet final state, events are rejected if there is at least one b-tagged AK5 jet with $p_T > 30$ GeV, using a working point tuned to provide a mis-identification rate of light-quark and gluon jets of 1% and a b-jet efficiency of about 70%, preserving about 90% of signal events. To enhance the analysis sensitivity, two V-jet categories are defined:

- **High-purity category** (HP): $65 < m^{\mathrm{W}}_{\mathrm{pruned}} < 105$ GeV ($70 < m^{\mathrm{Z}}_{\mathrm{pruned}} < 110$ GeV) and $\tau_2/\tau_1 < 0.5$;
- **Low-purity category** (LP): $65 < m^{\mathrm{Z}}_{\mathrm{pruned}} < 105$ GeV ($70 < m^{\mathrm{Z}}_{\mathrm{pruned}} < 110$ GeV) and $0.5 < \tau_2/\tau_1 < 0.75$.

Even if the HP category dominates the total sensitivity of the analysis, the LP is retained to recover signal events produced by the decay of large mass resonances, that are not accepted by the standard W-tagging selection due to reconstruction limitation in resolving jet substructure at high p_T, as already investigated in Section 3.2.4.
Eventually, the event categorization is based on four classes of events, depending on the lepton flavor (μ or e) and V-jet purity (HP and LP). The transverse momentum and the N-subjettiness τ_2/τ_1 distributions for the hadronic W (Z) boson candidate in the $\ell\nu$+V-jet ($\ell\ell$+V-jet) channel is shown in Figure 4.4 (Figure 4.5), after requiring $65 < m_{\mathrm{pruned}} < 105$ GeV ($70 < m_{\mathrm{pruned}} < 110$ GeV). The τ_2/τ_1 distribution shows some disagreement between data and simulation, consistently with what reported in Section 3.2.7. Finally, a summary of the final kinematic and jet substructure selection criteria used in the $\ell\nu$+V-jet and $\ell\ell$+V-jet channels are presented in Table 4.2 and Table 4.3, respectively.

4.3 Signal modeling

Bulk graviton events, generated with $\tilde{k} = (0.2, 0.5)$ for several mass hypotheses, are used for modeling the lineshape of the reconstructed di-boson invariant mass distribution (m_{VV}). Figure 4.6 shows the width of the signal lineshape, as predicted in both Bulk and RS1 models, as a function of the curvature parameter ($\tilde{k}$) and different mass hypotheses.
For cases in which $\tilde{k} \leq 0.5$, the relative width of the Bulk graviton resonance ($\Gamma_{\mathrm{G}}/m_{\mathrm{G}}$) is

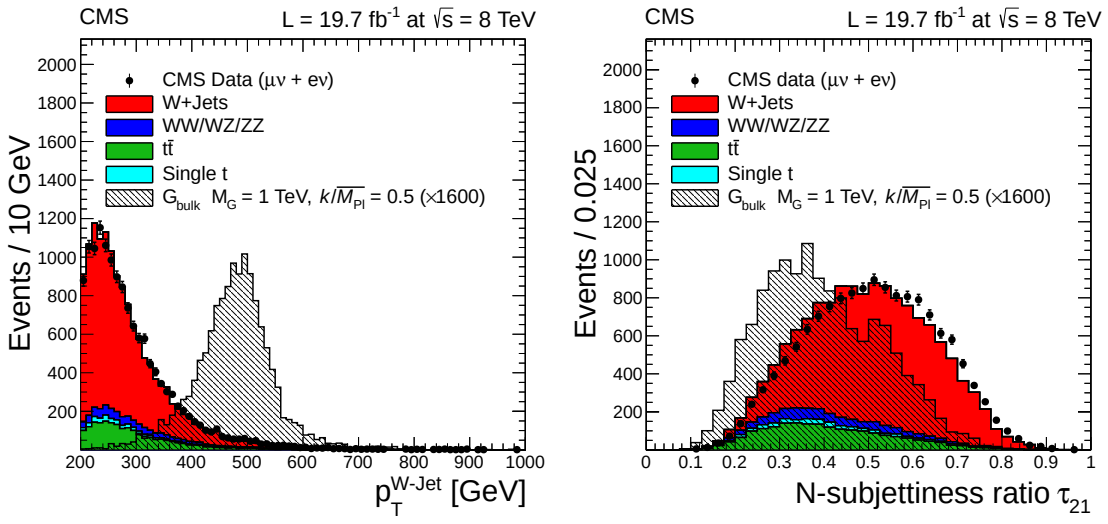

Figure 4.4: Hadronic W-boson p_T and N-subjettiness ratio τ_2/τ_1 distributions obtained requiring $65 < m_{pruned} < 105$ GeV after merging muon and electron channels in the $\ell\nu$+V-jet. The di-boson, $t\bar{t}$, and single-top backgrounds are taken from simulation and are normalized to the integrated luminosity of the data sample. The W+jets background is rescaled so that the total number of background events matches the number of events in data.

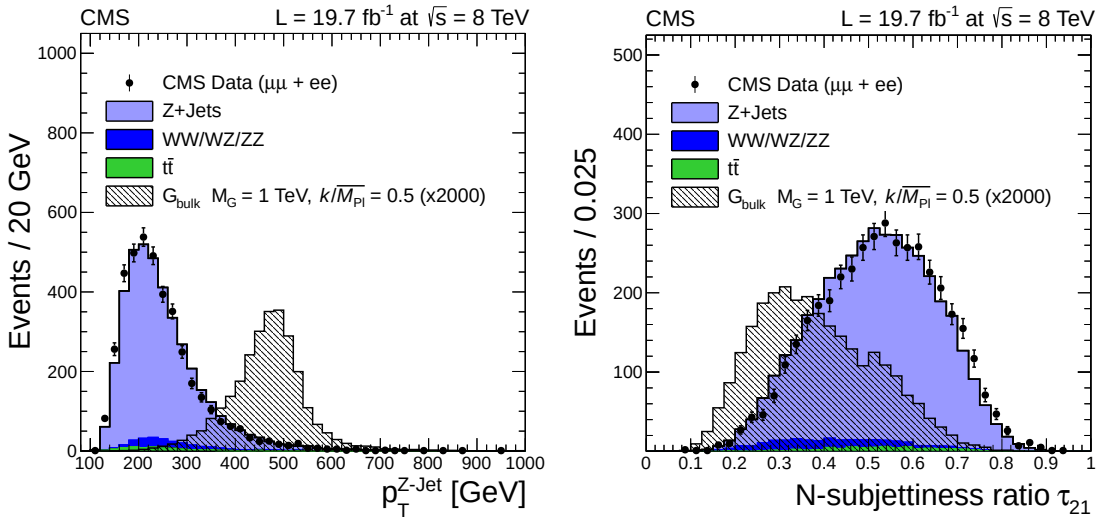

Figure 4.5: Hadronic Z-boson p_T and N-subjettiness ratio τ_2/τ_1 distributions obtained requiring $70 < m_{pruned} < 110$ GeV after merging muon and electron channels in the $\ell\ell$+V-jet. The di-boson and $t\bar{t}$ backgrounds are taken from simulation and are normalized to the integrated luminosity of the data sample. The DY+jets background is rescaled so that the total number of background events matches the number of events in data.

small enough, below 1%, to be neglected when compared to the detector resolution over the whole explored mass range, from 0.6 to 2.5 TeV, making the analysis a search for narrow width resonances along the di-boson invariant mass spectrum. Thus, the modeling of detector effects on the signal lineshape results to be independent from the benchmark $\tilde{k}$ value used for generating the events.

Selection	Value	Comments
`Tight` Lepton selection		
Muon p_T and η	$p_T > 50$ GeV, $\lvert\eta\rvert < 2.1$	
Electron p_T and η	$p_T > 90$ GeV, $\lvert\eta\rvert_{\text{SC}} < 2.5$	avoiding `ECAL` gap
`Loose` Lepton selection		
Electron p_T	$p_T > 35$ GeV	
Muon p_T	$p_T > 20$ GeV	
`CA8` jet selections		
Jet p_T and η	$p_T > 80$ GeV, $\lvert\eta\rvert < 2.4$	hadronic W-boson
`AK5` jet selections		
Jet p_T and η	$p_T > 30$ GeV, $\lvert\eta\rvert < 2.4$	used for b-veto
$\not\!\!E_T$ selections		
$\not\!\!E_T$ (electron ch.)	$\not\!\!E_T > 80$ GeV	
$\not\!\!E_T$ (muon ch.)	$\not\!\!E_T > 40$ GeV	
V-Boson selections		
Leptonic W p_T	$p_T > 200$ GeV	
Hadronic W p_T	$p_T > 200$ GeV	
Di-boson invariant mass	$m_{\text{WW}} > 700$ GeV	
Back-to-back topology	$\Delta R(\ell, \text{W}_{\text{had}}) > \pi/2$, $\Delta\phi(\text{W}_{\text{had}}, \not\!\!E_T) > 2$	
	$\Delta\phi(\text{W}_{\text{had}}, W_{\text{lep}}) > 2$	
Veto		
Number of additional leptons	0	
Number of b-tag jets	0	`CVS` $\varepsilon_{bjet} \approx 70\%$
W-jet selections		
m_{pruned}	$65 < m_{\text{pruned}} < 105$ GeV	
N-subjettiness ratio (high purity)	$\tau_2/\tau_1 < 0.50$	
N-subjettiness ratio (low purity)	$0.50 < \tau_2/\tau_1 < 0.75$	

Table 4.2: Kinematic selections used in $\ell\nu$+V-jet analysis for the final event classification.

The final analysis strategy relies on an unbinned fit of the signal and background models to the reconstructed di-boson invariant mass spectrum, where both the discovery potential and the exclusion power depend on an accurate description of the signal shape.
Since an analytic description of the signal shape provides a much easier access to lineshapes for mass hypotheses not directly covered by the simulation, a double-sided Crystal-Ball (`CB`) function[1] is adopted to parametrize the `CMS` detector resolution. In contrast, simulated signal binned templates are less suitable to achieve this goal, since the small number of available simulated samples necessitates long mass range interpolation between different templates. To take into account differences between muon and electron momentum resolutions, the signal invariant mass distribution is parametrized separately in the two lepton flavor categories. No appreciable differences have been observed in the m_{VV} signal shape between low and high purity categories.
The typical width of the Gaussian core is about 3–5% of the nominal mass in the $\ell\ell$+V-

[1] A double-sided Crystal-Ball (`CB`) function is given by a Gaussian core with power law tails on both sides [144].

Selection	Value	Comments
Lepton selections		
Leading lepton p_T	$p_T > 40$ GeV	for both μ and e.
Subleading lepton p_T	$p_T > 40(20)$ GeV	for μ (e).
Electron η	$\lvert\eta\rvert < 2.5$	avoid the ECAL gap.
Muon η	$\lvert\eta\rvert < 2.4$	
Jet selections		
Jet p_T and η	$p_T > 30$ GeV, $\lvert\eta\rvert < 2.4$	
Veto		
Number of additional leptons	0	
V-Boson selections		
$m_{\ell\ell}$	$70 < m_{\ell\ell} < 110$ GeV	
Leptonic Z-boson p_T	$p_T > 80$ GeV	
Hadronic Z-boson p_T	$p_T > 80$ GeV	
Di-boson mass m_{ZZ}	$m_{ZZ} > 500$ GeV	
Z-jet selections		
m_{pruned}	$70 < m_{\text{pruned}} < 110$ GeV	
N-subjettiness ratio (high purity)	$\tau_2/\tau_1 < 0.50$	
N-subjettiness ratio (low purity)	$0.50 < \tau_2/\tau_1 < 0.75$	

Table 4.3: Kinematic selections used in $\ell\ell$+V-jet analysis for the final event classification.

jet channel, dominated by CA8 jet energy resolution, while it is at the level of 4–6% in the $\ell\nu$+V-jet one, dominated by both jet and $\not\!\!E_T$ contributions. Figure 4.7 shows two examples for the fitted lineshapes through a double-sided Crystal-Ball for Bulk gravitons with mass $m_G = 1.5$ TeV in the HP muon (left) and electron channel (right) in the $\ell\nu$+V-jet analysis. The signal selection efficiency, evaluated for each category, is defined as the number of selected signal events over the number of generated ones, taking into account all the possible lepton flavours (e, μ and τ). The efficiency trend as a function of the mass hypothesis is parametrized with a continuous polynomial function, to obtain an analytic interpolation between mass points in each flavor and purity category. The dependence reflects the W-tagging efficiency shape already shown in Figure 3.10. In fact, considering the selection efficiency in the HP category, it reaches a global maximum close to 20% for m_G around 1.2 TeV in both $\ell\nu$+V-jet ($\ell\ell$+V-jet) and muon (electron) cases. A turn-on is placed at lower masses due to an inefficiency in the merged jet reconstruction for low boosted V-boson, while it drops at high masses due to W-tagging efficiency loss for high p_T V-jets.

4.4 Background estimation

After the full selection, summarized in Table 4.2 for the $\ell\nu$+V-jet channel and in Table 4.3 for the $\ell\ell$+V-jet one, the dominant background comes from SM V+jets events as shown in Figure 4.4 and Figure 4.5. Since both normalization and shape discrepancies are visibile between data and simulation, a data driven method has been developed to estimate this background component. Other minor sources, such as $t\bar{t}$, single-top and di-boson production, are estimated

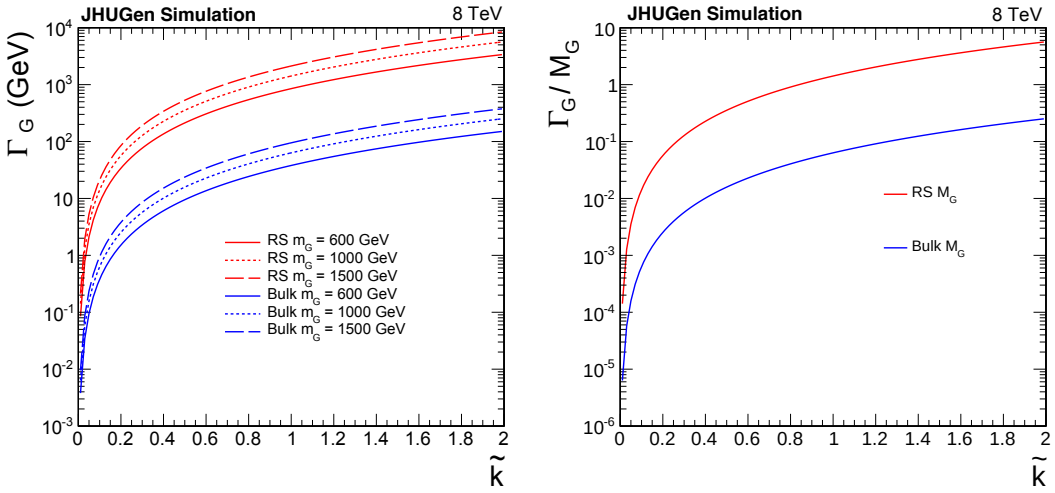

Figure 4.6: (Left) Intrinsic width of gravitons as a function of the coupling constant $\tilde{k}$ and for various mass hypotheses for both `RS1` and Bulk `WED` models. (Right) The same dependence is expressed as relative fraction of the signal width with respect to a reference graviton mass of 1 TeV.

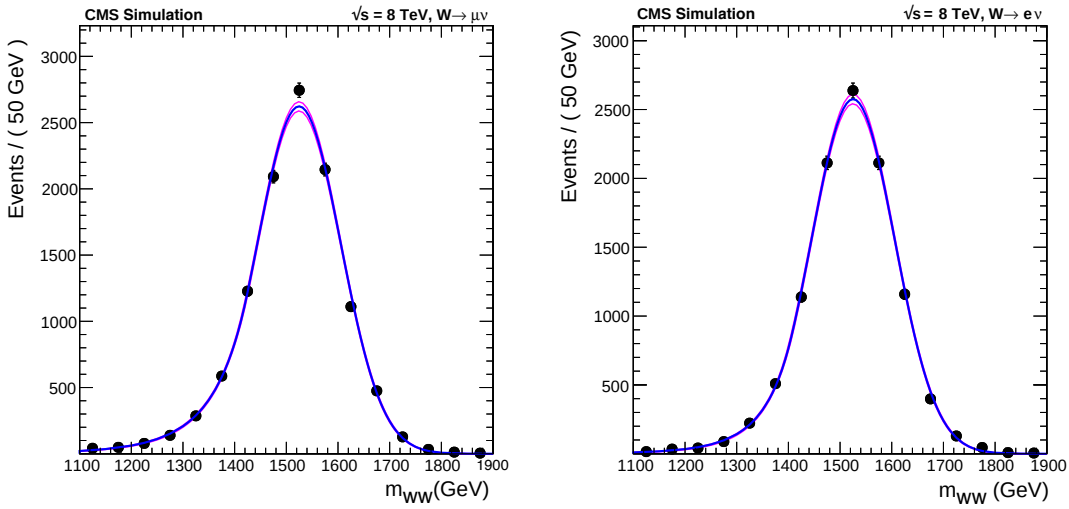

Figure 4.7: Modeling of the reconstructed signal shape with a double-sided Crystal Ball function for Bulk gravitons in the $\ell\nu$+V-jet analysis: `HP` muon category (left), `HP` electron one (right). In both cases, a `Bulk` graviton sample with a generated mass $m_G = 1.5$ TeV is considered.

using simulated samples, after applying correction factors for residual data-to-simulation disagreement measured in control samples selected in data.
The discrepancy observed between data and simulation in both the pruned jet mass and the N-subjettiness τ_2/τ_1 could bias the signal efficiency estimated from the simulated samples. Fot this reason, it is crucial to study this mis-modeling in a V-jet enriched control sample, similar in kinematics to the event characteristics expected for a genuine signal. In this sample, correction factors for the signal efficiency and lineshape are calculated from a direct comparison between data and simulation, as well as their systematic uncertainties. The adopted strategy consists of selecting a sample of high-p_T W-jets in boosted $t\bar{t}$ semi-leptonic events, applying the same requirements described in Section 3.2.7.

Figure 4.8 shows the N-subjettiness τ_2/τ_1 (left) and the pruned jet mass distribution (right) obtained applying $\tau_2/\tau_1 < 0.5$ selection. The only difference with respect to Figure 3.22 stands in removing the ΔR selection between the W-jet and the closest b-jet.

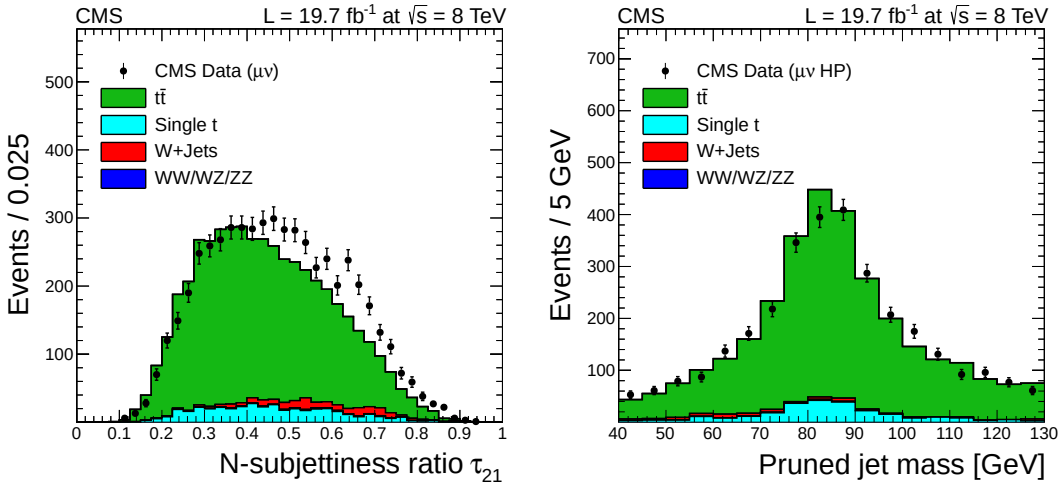

Figure 4.8: Distributions from the semi-leptonic $t\bar{t}$ enriched control sample in the muon channel. (Left) N-subjettiness ratio τ_2/τ_1, (right) pruned jet mass m_{pruned} after requiring $\tau_2/\tau_1 < 0.5$. The distributions show some disagreement between data and simulation, in particular for events that are failing the N-subjettiness requirement. The simulation is corrected for these discrepancies, ensuring that the analysis is robust against differences between data and simulation, independently of their sources.

Since jet substructure observables depend on the parton shower model, in this case `Pythia 6` is considered as a benchmark, as it is adopted to generate signal simulated events. Specific corrections for pure W-jets in this sample are derived from a simultaneous fit to the pruned jet mass distribution, for events passing or failing m_{pruned} and τ_2/τ_1 requirements, as previously detailed in Section 3.2.9. The ratio of the efficiency in data and simulation yields W-tagging scale factors used to correct the total signal efficiency predicted by the simulation, measured to be 0.93 ± 0.06 (1.28 ± 0.30) for the `HP` (`LP`) category, combining muon and electron channels. Eventually, the W-jet mass peak and resolution, which are extracted from the same fit, are used to correct the pruned jet mass shape in $t\bar{t}$, di-boson and signal simulated events in the signal region phase space. The same corrections are adopted also in the case where the V-jet is assumed to come from a Z-boson decay, since the kinematic properties of W and Z-jets are similar and the agreement between data and simulation is expected to be equally good.
At the same time, from the comparison between data and simulation predicted yields, a normalization correction factor for $t\bar{t}$ and single-top backgrounds is evaluated for events in the pruned mass region $65 < m_{\text{pruned}} < 105$ GeV. The measured scale factor is 0.97 ± 0.02 (0.96 ± 0.03) in the muon (electron) channel for the `HP` category, while it is measured to be 1.31 ± 0.05 (1.39 ± 0.08) in the `LP` one. These scale factors, including both the resonant and the combinatorial contributions, are used directly to correct the normalization of the $t\bar{t}$ and single-top simulated backgrounds in the analysis signal region, since data-to-`MC` differences, introduced by the inversion of the b-tag selection when moving from the $t\bar{t}$ control sample to the signal region, are taken into account by a re-weighting simulated events with proper b-tagging scale factors [145].
To estimate the V+jets background from data, signal free m_{pruned} sideband regions are consid-

ered. In the $\ell\nu$+V-jet channel, lower and upper sidebands are defined when the m_{pruned} ranges within [40,65] GeV and [105,130] GeV, respectively. In the $\ell\ell$+V-jet case, the two regions are defined when m_{pruned} lies between [50,70] GeV or [110,130] GeV. The width of the m_{pruned} signal region, 20 GeV around the expected W-jet (Z-jet) pruned mass peak, allows to keep the signal contamination inside the sideband regions below 3% over the whole graviton mass hypothesis range, since it corresponds to roughly 2.5 times the W-jet (Z-jet) mass resolution as measured in Section 3.2.9.

The normalization of the V+jets background inside the signal region is estimated from a fit to the m_{pruned} distribution, using only observed data in the lower and upper sidebands. The analytical parametrization for the V+jets is chosen from a preliminary fit to pure simulated V+jets events, performed over the full m_{pruned} range [40,130] GeV. In contrast, the other backgrounds ($t\bar{t}$, di-boson and single-top) are estimated from simulation, after applying both normalization and the pruned mass shape corrections previously described. As expected the m_{pruned} shapes are independent from the lepton flavor, but strongly dependent on the N-subjettines category, the final state topology ($\ell\nu$+V-jet or $\ell\ell$+V-jet channels) and on the threshold applied on the reconstructed di-boson invariant mass (m_{VV}). A summary of the analytical functions used to parametrize the m_{pruned} distribution are listed in Table 4.4, defined as follows:

$$\text{F}_{\text{ErfExp}}(x) = e^{c_0 x} \cdot \frac{1 + \text{Erf}((x-a)/b)}{2}$$

$$\text{F}_{\text{ExpGaus}}(x) = e^{c_0 x} \cdot \text{Gaus}(x, x_1, \sigma_1)$$

$$\text{F}_{\text{ErfExp2Gaus}}(x) = e^{c_0 x} \cdot \frac{1 + \text{Erf}((x-a)/b)}{2} + c_1 \cdot \text{Gaus}(x, x_1, \sigma_1) + c_2 \cdot \text{Gaus}(x, x_2, \sigma_2)$$

$$\text{F}_{\text{4Gaus}}(x) = c_1 \cdot \text{Gaus}(x, x_1, \sigma_1) + c_2 \cdot \text{Gaus}(x, x_2, \sigma_2) + c_3 \cdot \text{Gaus}(x, x_3, \sigma_3) + c_4 \cdot \text{Gaus}(x, x_4, \sigma_4)$$

Channel	m_{VV} (TeV)	Category	V+jets	$t\bar{t}$	single-top	di-boson
$\ell\nu$+V-jet	[0.7-3.0]	HP	F_{ErfExp}	$\text{F}_{\text{ErfExp2Gaus}}$	$\text{F}_{\text{ExpGaus}}$	F_{4Gaus}
$\ell\nu$+V-jet	[0.7-3.0]	LP	F_{ErfExp}	$\text{F}_{\text{ExpGaus}}$	$\text{F}_{\text{ExpGaus}}$	$\text{F}_{\text{ExpGaus}}$
$\ell\ell$+V-jet	[0.5-2.8]	HP	F_{ErfExp}	$\text{F}_{\text{ExpGaus}}$	$\text{F}_{\text{ExpGaus}}$	F_{4Gaus}
$\ell\ell$+V-jet	[0.5-2.8]	LP	F_{ErfExp}	F_{Exp}	$\text{F}_{\text{ExpGaus}}$	$\text{F}_{\text{ExpGaus}}$

Table 4.4: Summary of the shapes used to fit the m_{pruned} spectra of each background component taking into account: m_{VV} fit range, N-subjettines category and final state topology.

Figure 4.9 and Figure 4.10 show the result of this fit procedure for the $\ell\nu$+V-jet and $\ell\ell$+V-jet analyses, respectively, while Table 4.5 and Table 4.6 report the predicted number of background events inside the pruned mass signal region, comparing them to data.

The m_{VV} shape for the V+jets background inside the signal region is extracted from the low mass m_{pruned} sideband through an extrapolation function $\alpha_{\text{MC}}(m_{\text{VV}})$, which is derived from V+jets simulated events as:

$$\alpha_{\text{MC}}(m_{\text{VV}}) = \frac{F^{\text{V+jets}}_{\text{MC,SR}}(m_{\text{VV}})}{F^{\text{V+jets}}_{\text{MC,SB}}(m_{\text{VV}})} \tag{4.2}$$

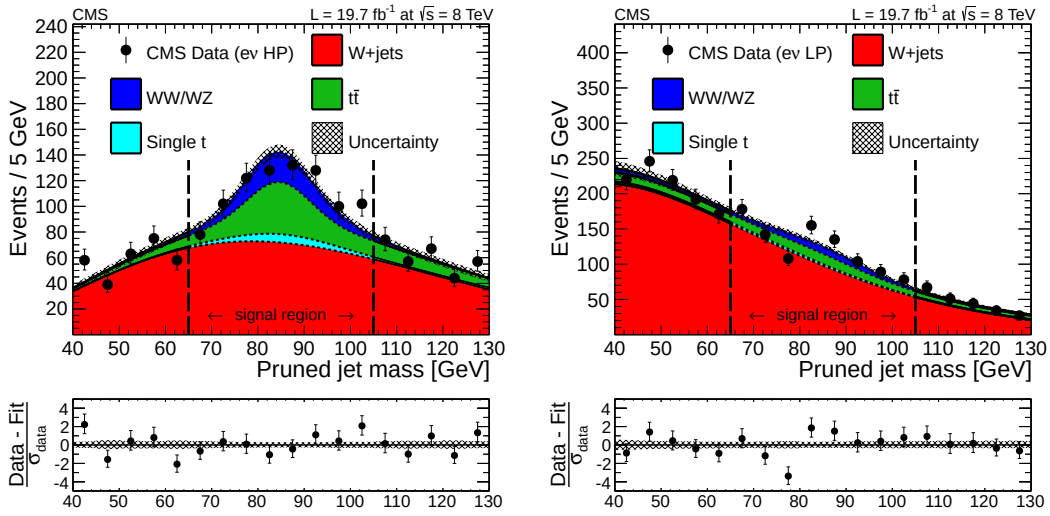

Figure 4.9: Distributions of the pruned jet mass, m_{pruned}, in the $\ell\nu$+V-jet analysis for the electron channel, where the left (right) panel shows the HP (LP) category. The prediction of the non-resonant W+jets background comes from a fit excluding the signal region (limited by vertical dashed lines), while the minor background yields come from the simulation. The MC resonant shapes are corrected using the differences between data and simulation in the W-jet peak position and resolution as measured in a dedicated $t\bar{t}$ control region. At the bottom of each plot, the bin-by-bin fit residuals, (data-fit)/σ_{data}, are reported together with the uncertainty band of the fit normalized by σ_{data}.

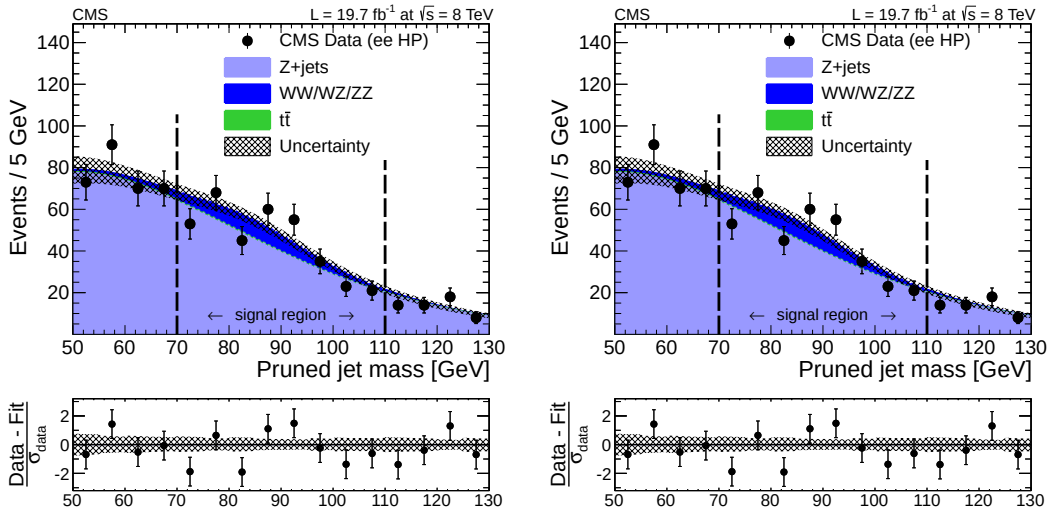

Figure 4.10: Distributions of the pruned jet mass, m_{pruned}, in the $\ell\ell$+V-jet analysis for the di-electron channel, where the left (right) panel shows the HP (LP) category. The prediction of the non-resonant Z+jets background comes from a fit excluding the signal region (marked using vertical dashed lines), while the minor background yields come from the simulation. The MC resonant shapes are corrected using the differences between data and simulation in the Z-peak position and resolution as measured in a dedicated $t\bar{t}$ control region. At the bottom of each plot, the bin-by-bin fit residuals, (data-fit)/σ_{data}, are reported together with the uncertainty band of the fit normalized by σ_{data}.

	$\mu\nu$ + V-jet HP	$\mu\nu$ + V-jet LP	$e\nu$ + V-jet HP	$e\nu$ + V-jet LP
Observed yield	1483	1546	892	988
Expected background	1434 ± 38	1644 ± 41	878 ± 30	978 ± 31
Graviton ($\tilde{k} = 0.5$)	Signal expectation (MC)			
$m_{\mathrm{G}} = 0.8$ TeV	12.8	5.1	10.1	3.9
$m_{\mathrm{G}} = 1.2$ TeV	0.92	0.43	0.79	0.37

Table 4.5: Observed and expected yields for the $\ell\nu$+V-jet analysis, quoted in the range $0.7 < m_{\mathrm{WW}} < 3$ TeV. The expected W+jets background is obtained from the sideband procedure. The uncertainties in the background prediction are statistical in nature, as they mainly depend on the selected number of events in the sidebands. Statistical uncertainties in the signal expectations are negligible.

	$\mu\mu$ + V-jet HP	$\mu\mu$ + V-jet LP	ee + V-jet HP	ee + V-jet LP
Observed yield	575	338	360	233
Expected background	622 ± 29	338 ± 22	370 ± 22	207 ± 17
Graviton ($\tilde{k} = 0.5$)	Signal expectation (MC)			
$m_{\mathrm{G}} = 0.8$ TeV	2.4	0.5	2.0	0.4
$m_{\mathrm{G}} = 1.2$ TeV	0.16	0.04	0.14	0.035

Table 4.6: Observed and expected yields for the $\ell\ell$+V-jet analysis, quoted in the range $0.5 < m_{\mathrm{ZZ}} < 2.8$ TeV. The expected Z+jets background is obtained from the sideband procedure. The uncertainties in the background prediction are statistical in nature, as they mainly depend on the selected number of events in the sidebands. Statistical uncertainties in the signal expectations are negligible.

where $F^{\mathrm{V+jets}}_{\mathrm{MC,SB}}(m_{\mathrm{VV}})$ and $F^{\mathrm{V+jets}}_{\mathrm{MC,SR}}(m_{\mathrm{VV}})$ are the probability density functions which parametrize the simulated di-boson invariant mass spectra in the low mass sideband region and in the signal one, respectively.

The high m_{pruned} sideband is not considered to exclude possible contamination from BSM resonances decaying into a V-boson and a SM Higgs boson, as well as it suffers from a larger $t\bar{t}$ contamination. This contribution from $\mathrm{X} \to h\mathrm{V}$ resonances, in which the Higgs boson is reconstructed as a single jet in the CMS detector ($h \to b\bar{b}$ or $h \to \mathrm{WW}^* \to 4q$) and the V-boson decays leptonically, would populate the high mass sideband region of both the $\ell\nu$+V-jet (m_{pruned} in 105-130 GeV) and the $\ell\ell$+V-jet (m_{pruned} in 110-130 GeV) analyses.

Since the lower sideband region does not represent a perfectly pure sample of V+jets events in data, the presence of minor backgrounds is subtracted from the observed m_{VV} distribution relying on the simulation prediction, to obtain an estimation of the V+jets shape in the sideband control region $F^{\mathrm{V+jets}}_{\mathrm{Data,SB}}(m_{\mathrm{VV}})$. Finally, the shape of the V+jets background distribution in the signal region is obtained by rescaling $F^{\mathrm{V+jets}}_{\mathrm{Data,SB}}(m_{\mathrm{VV}})$ with $\alpha_{\mathrm{MC}}(m_{\mathrm{VV}})$. The ultimate background prediction into the signal region, $N^{bkg}_{\mathrm{SR}}(m_{\mathrm{VV}})$, is given by:

$$N^{bkg}_{\mathrm{SR}}(m_{\mathrm{VV}}) = N^{\mathrm{V+jets}}_{\mathrm{SR}} \times \alpha_{\mathrm{MC}}(m_{\mathrm{VV}}) \times F^{\mathrm{V+jets}}_{\mathrm{Data,SB}}(m_{\mathrm{VV}}) + \sum_k N^k_{\mathrm{SR}} \times F^k_{\mathrm{MC,SR}}(m_{\mathrm{VV}}) \qquad (4.3)$$

where k runs over the non V+jets background contributions, $N^{\mathrm{V+jets}}_{\mathrm{SR}}$ and N^k_{SR} represent the yields of the dominant V+jets and of the other minor sources estimated inside the analysis

signal region, respectively. The ratio $\alpha_{\mathrm{MC}}(m_{\mathrm{VV}})$ accounts for small kinematic differences between the signal region and the sideband one, which are mostly independent from the cross section prediction. The sample produced via `Herwig++` has been used to determine this sideband-to-signal region correlation function, since it is known to better describe the pruned mass than `Pythia6`, as shown in Figure 3.16.

4.4.1 Validation of V+jets background extraction

To maximize the stability and the performance of the analysis results, minimizing possible sources of bias in the signal estimation along the explored m_{VV} spectrum, different tests have been performed:

- The m_{VV} range adopted in the fit determines the region of masses probed by these searches. This range is chosen to ensure a smoothly falling background spectrum, sufficiently far from the kinematic turn-on generated by the acceptance selections, allowing for a good stability and a robust control of the background estimation. For the $\ell\nu$+V-jet analysis, the fits are performed in the range $0.7 < m_{\mathrm{VV}} < 3$ TeV, while in the $\ell\ell$+V-jet channel the ranges for the `HP` and `LP` categories are $0.5 < m_{\mathrm{VV}} < 2.8$ TeV and $0.65 < m_{\mathrm{VV}} < 2.8$ TeV, respectively.

- To parametrize this smoothly falling background distribution, a generic exponential function, called leveled-exponential parametrization, is adopted:

$$f(m_{\mathrm{VV}}) = N \times \exp\left(\frac{-m_{\mathrm{VV}}}{a + b \cdot m_{\mathrm{VV}} + c \cdot m_{\mathrm{VV}}^2}\right) \tag{4.4}$$

  For each analysis category, considering both data in the low sideband region and V+jets simulated events in the signal one, a set of unbinned fits is performed using alternative class of functions with an increasing number of degrees of freedom (dof). To estimate how many parameters are needed to properly describe this background, a `F-test` has been performed [146]. In particular, taking two alternative models with different dof belonging to the same class of functions (exponential, polynomials or power laws), the model with more parameters will be able to fit the data better than the alternative one, but it is not guaranteed that the improvement is significant. Considering, as a null hypothesis (h_0), that the model with less parameters is enough to correctly describe the data, the following test statistics is used for testing against the null hypothesis:

$$F(h_1, h_0) = \frac{(\sum_i(\ data_i - fit(h_0)_i\ )^2 - \sum_i(\ data_i - fit(h_1)_i\ )^2)/(n(h_1) - n(h_0))}{\sum_i(\ data_i - fit(h_0)_i\ )^2/(N - n(h_1))} \tag{4.5}$$

  Assuming that $F(h_1, h_0)$ follows asymptotically a Fisher distribution with ($n(h_1)$-$n(h_0)$,N-$n(h_1)$) degrees of freedom, an confidence level is computed as:

$$\mathrm{CL}(h_0, h_1) = 1 - \int_{-\infty}^{F(h_1,h_0)} \mathrm{Fisher}(n_1 - n_0, N - n_1) \tag{4.6}$$

  where the null hypothesis is rejected when the observed confidence level is smaller than 0.05. Results show that a leveled-exponential function with two parameters represents the best compromise for all the categories. This test is also applied to study the shapes of

minor backgrounds, relying only on simulated events. Eventually, V+jets and di-boson events are parametrized through a leveled exponential with two parameters, while single-top and $t\bar{t}$ with a simple exponential shape.

- To test the validity and the robustness of the data driven method for the V+jets, closure tests are performed extracting this background from a sideband inside a signal free control region, where the compatibility in both shape and normalization with data can be measured. The first test is performed dividing the low mass sideband into two parts: $40 < m_{\text{pruned}} < 55$ GeV is used as sideband (A), while $55 < m_{\text{pruned}} < 65$ GeV is considered as a signal region (B), so that the W+jets rate is predicted fitting data in (A) and extrapolating the fit result inside (B) through a proper $\alpha_{\text{MC}}(m_{\text{VV}})$ function. In contrast, the same test is performed extrapolating W+jets events from the low mass sideband into a sub-region of the upper one, defined by $105 < m_{\text{pruned}} < 115$ GeV (C).
  Figure 4.11 (left) shows the pruned jet mass fit to data using only $\mu\nu$+W-jets HP events inside the region (A) and the upper sideband (US), indicated by solid colors, to extract the W+jets prediction inside the region (B). At the same time, Figure 4.11 (right) shows the fit obtained when only events in the low mass sideband (LS) and in the region $115 < m_{\text{pruned}} < 130$ GeV (D) are considered.

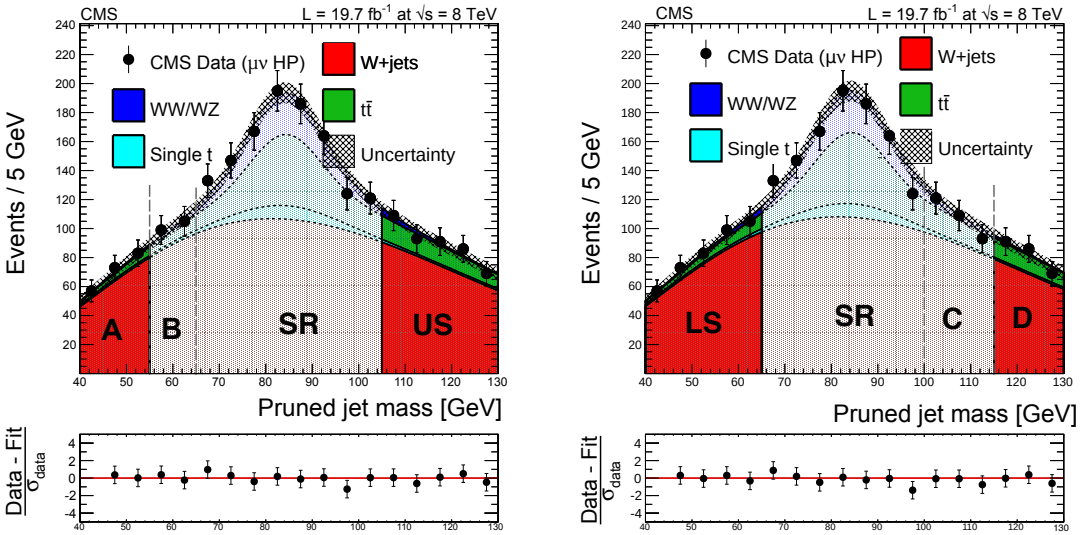

Figure 4.11: Fit to the pruned mass distribution inside the $\mu\nu$+V-jet channel considering only events in data with pruned mass between 40-55 GeV (A) or 105-130 GeV (US) on the left, while pruned mass within 40-65 GeV (LS) or 115-130 GeV (D) on the right.

Figure 4.12 (left) shows the $\alpha_{\text{MC}}(m_{\text{WW}})$ function obtained from a simultaneous fit of W+jets simulated events in the sideband (A) and in the region (B), using the leveled-exponential parametrization defined in Equation 4.4. Figure 4.12 (right) shows the m_{WW} distribution for events in data with $40 < m_{\text{pruned}} < 55$ GeV (region A) in the $e\nu$+V-jet HP category, where the W+jets shape is parametrized again according to a leveled-exponential function.
Finally, Figure 4.13 shows a comparison between the predicted total background, obtained through Equation 4.3, with the data inside the signal free region B (left) and C (right). Looking at the bin-by-bin fit residuals, (data-fit)/σ_{data}, a good agreement is found over the whole m_{VV} range for both the extrapolations. The test is repeated in

all the different analysis categories showing always consistency between the predicted background and the one observed in data. Thus, the test closes validating the proposed strategy for the V+jets background estimation.

- The effect of using alternative functional forms for the background extraction has been investigated. No significant biases are measured on the predicted number of background events when different parametrizations, leveled-exponential or power laws, are adopted.

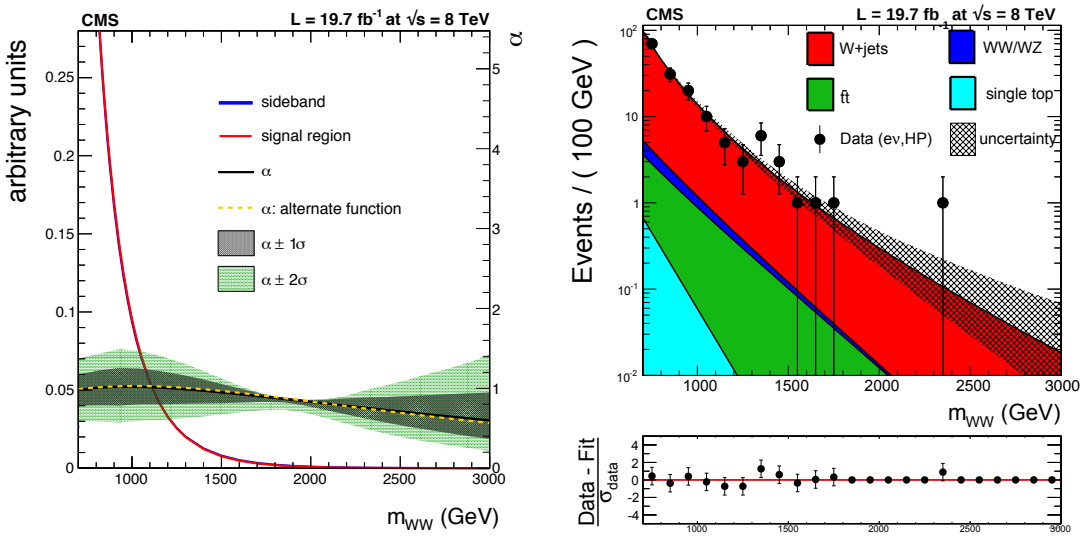

Figure 4.12: (Left) The extrapolation functions $\alpha_{\mathrm{MC}}(m_{\mathrm{WW}})$ obtained by simultaneously fitting the simulated m_{VV} distributions inside the sideband (A) and signal region (B). (Right) Di-boson invariant mass distribution m_{WW} for events with $40 < m_{\mathrm{pruned}} < 55$ GeV (A). The W+jets shape is fitted, after subtracting contaminations from minor backgrounds, by means of a leveled-exponential function.

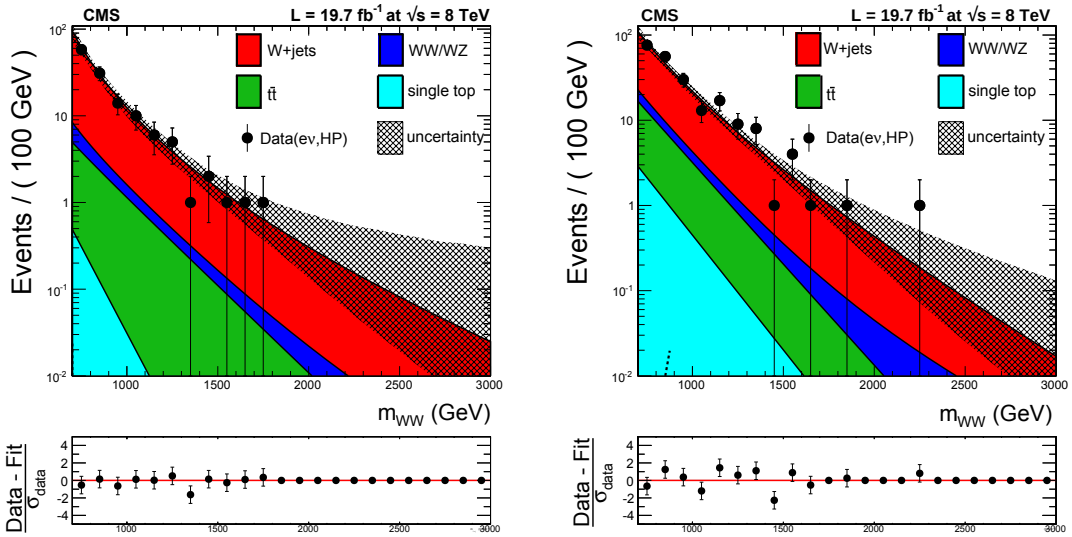

Figure 4.13: Distributions in m_{WW} for data and the expected backgrounds in the $e\nu$+V-jet (HP) category inside the pruned mass region defined by $55 < m_{\mathrm{pruned}} < 65$ GeV (left) and $105 < m_{\mathrm{pruned}} < 115$ GeV (right). The W+jets background shape has been extracted using events within $40 < m_{\mathrm{pruned}} < 55$ GeV in the former, while adopting the entire low mas sideband in the latter.

4.4.2 V+jets background extraction

Figure 4.14 shows the $\alpha_{\mathrm{MC}}(m_{\mathrm{VV}})$ obtained in the HP and LP category in the $\mu\nu$+V-jet channel. The blue and the red dashed dotted lines are the fits to the m_{VV} distribution in the signal and in the low-mass sideband region, performed using the two parameters leveled-exponential function. These two samples are simultaneously fitted and, to account for any correlation between them, the parameters used to model the signal region shape are correlated with the sideband ones. The obtained extrapolation funcrion $\alpha_{\mathrm{MC}}(m_{\mathrm{VV}})$ is shown as a solide black line, while the dark (light) shaded region corresponds to the $1\,\sigma$ ($2\,\sigma$) uncertainty bands of the fit. The structure of the uncertainty band is related to the function used to model the m_{VV} spectrum and it represents only the shape uncertainty of the function, since additional normalization uncertainties are not displayed.
In Figure 4.15, the fit result to the data in the m_{pruned} low-mass sideband is shown for the HP (left) and LP (right) $\mu\nu$+V-jet category. Results of the final background extraction in the signal region will be presented in Section 4.6.

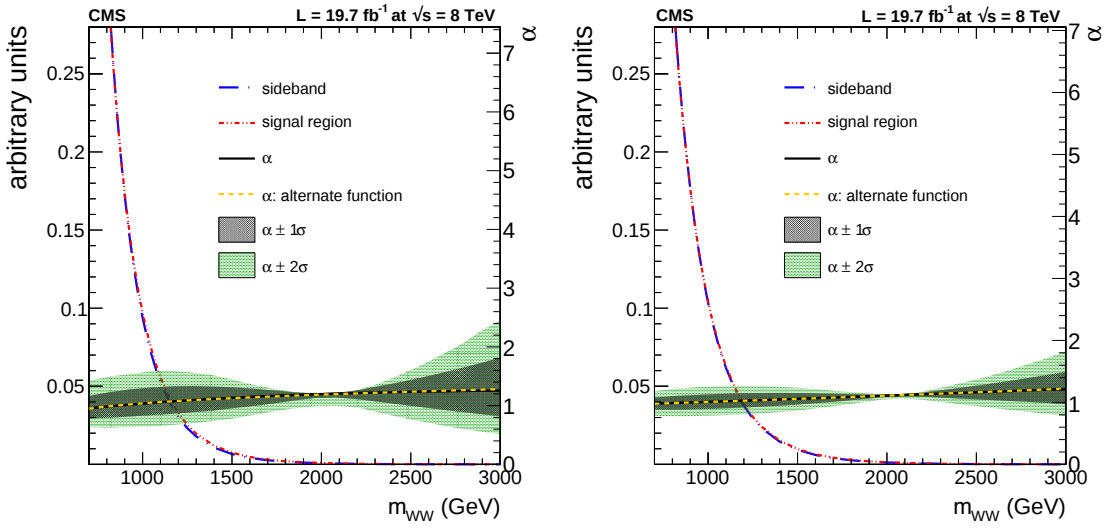

Figure 4.14: The extrapolation function $\alpha_{\mathrm{MC}}(m_{\mathrm{VV}})$ from the low mass sideband to the signal region is shown for high (left) and low (right) purity $\mu\nu$+V-jet category. The grey area displays the statistical uncertainty band of the fit parameters, while systematic effects related to parton shower and fit function models are indicated by a dashed purple and a dashed yellow line, respectively. The shape uncertainty becomes close to zero around 2 TeV as a consequence that $\alpha_{\mathrm{MC}}(m_{\mathrm{VV}})$ is a probability density function. In fact, the extremes of the envelope have to cross each other to conserve the total probability.

4.5 Systematic uncertainties

Systematic uncertainties are described here separately for signal and background predictions.

4.5.1 Systematic uncertainties on the background estimation

Uncertainties in the estimation of the background affect both the normalization and shape of the m_{VV} distribution. The uncertainty on the V+jets background normalization is mainly statistical in nature, related to the uncertainties on the parameters extracted from the pruned jet mass sideband fit. This strongly depends on the amount of data in the pruned mass

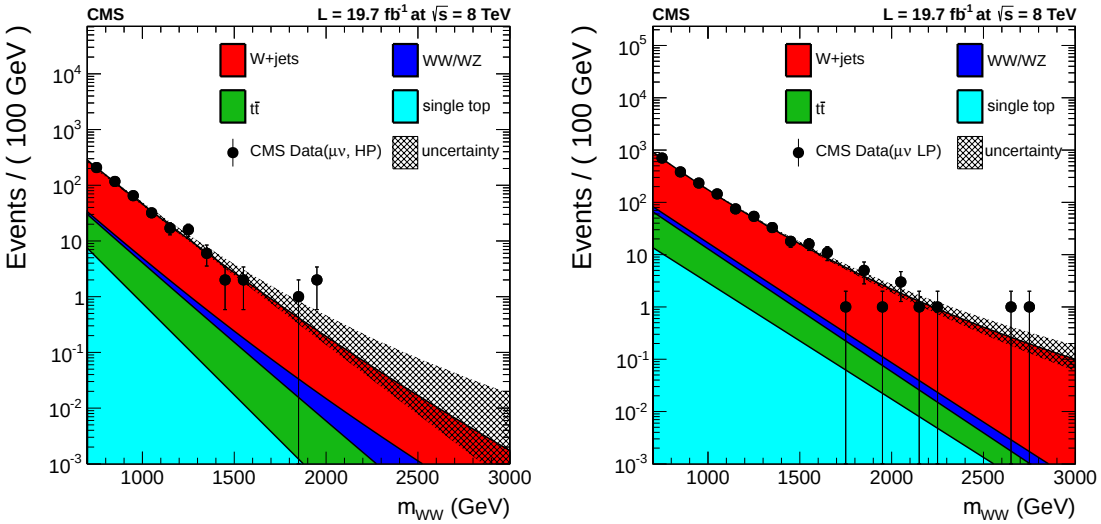

Figure 4.15: The fits to estimate $F_{\mathrm{data,SB}}(m_{\mathrm{VV}})$ for both high (left) and low (rigth) purity categories in the $\mu\nu$+V-jet channel. Minor backgrounds ($t\bar{t}$, di-boson and single-top) are estimated from simulation, while the W+jets shape is the fit result from the data.

sideband regions. Additional smaller effects are also considered and added in quadrature to the pure statistical term: differences in the V+jets estimation arising from a different pruned mass shape prediction between `Herwig++` and `Pythia 6`, as well as differences arising when alternative parametrizations are adopted to model the V+jets m_{pruned} distribution. The total uncertainty on the V+jets yields remains below 10% in all the explored final states.
Furthermore, systematic uncertainties in the V+jets shape are estimated from the covariance matrix of the fit to the data performed in the low mass sideband ($40 < m_{\mathrm{pruned}} < 65$ GeV) and from the uncertainties in the modeling of $\alpha_{\mathrm{MC}}(m_{\mathrm{VV}})$. They are both mainly statistical in nature, as they are driven by the available amount of data in the sideband and by the number of simulated V+jets events fulfilling the analysis selection, as shown in Figure 4.15 and Figure 4.14, respectively.
In addition, differences in the extrapolated V+jets shape due to alternative shower models, used for the $\alpha_{\mathrm{MC}}(m_{\mathrm{VV}})$ estimation, or alternative parametrizations of the m_{VV} spectrum are accounted for by enlarging the fitted parameter errors to cover their effects. To perform this, the covariance matrix of the fitted parmeters is diagonalized by means of a suitable eigenvector base, then to each transformed shape parameter a Gaussian contraint is associated where its width is properly enlarged to cover the estimated systematic effects.
The systematic uncertainty in the $t\bar{t}$ normalization comes from the uncertainties in the data-to-simulation scale factors evaluated in the top-quark enriched control sample, which is measured to be around 6%.
The uncertainty in the WW inclusive cross section in this boosted regime is assigned to be 20%, estimated from the uncertainty on the published `CMS` cross section measurement at $\sqrt{s} = 8$ TeV [5]. In addition, the uncertainty on the V-tagging scale factor, described in Section 3.2.10, is considered also for the WW background. The same uncertainties derived for WW production are also used for WZ and ZZ processes.
The lepton trigger, reconstruction and identification efficiencies and the related uncertainties are estimated with a dedicated tag-and-probe analyses, looking at $\mathrm{Z} \to \mu\mu$ (ee) events. The

uncertainties in trigger and identification efficiencies for high-p_T muons are estimated to be 3% and 4%, respectively, while the total uncertainty in the electron trigger, reconstruction and identification efficiency is 3%.
An additional systematic, affecting the normalization of non data driven background, is represented by the uncertainty in the data-to-simulation scale factors for b-jet identification, derived following [125]. The uncertainty related to the total integrated luminosity is 2.6% [108].
The accuracy on energy and momentum measurements for leptons and jets represents an important source of systematic uncertainties for simulated backgrounds samples. In particular, muon momentum scale and resolution, electron energy scale and resolution, jet energy scale and resolution, and the unclustered energy in the event are considered.
Leptons four momenta are varied within one standard deviation according to the muon momentum scale [147] or the electron energy scale [79] uncertainties, or applying a Gaussian smearing in case of resolution uncertainties. For each variation, the event selection is reapplied. The largest relative change in the background efficiency compared to the default value is considered as the systematic uncertainty for that specific source. The same procedure is also applied for the jet four-momenta using the corresponding energy scale and resolution uncertainties [84]. During this procedure, variations in the lepton and jet four-momenta are propagated coherently to the $\vec{\not{E}}_T$ through:

$$\vec{\not{E}}_T^{*} = \vec{\not{E}}_T + \sum_{i}^{objects} (\vec{p}_T - \vec{p}_T^{*}) \tag{4.7}$$

where $\vec{p}_T^{*}$ stands for the transverse momentum obtained after the objects, lepton or jets, variation (shift or smear).
Eventually, Table 4.7 reports a complete summary of all the normalization uncertainties considered for each single background contribution in both $\ell\nu$ + V-jet ($\ell\ell$ + V-jet) channels and HP (LP) categories.

4.5.2 Systematic uncertainties on the signal estimation

As described for the background processes the previous Section, muon momentum scale and resolution, electron energy scale and resolution, jet energy scale and resolution uncertainties are sources of systematic uncertainty also on the signal efficiency. They are evaluated independently for each considered signal sample as a function of the graviton mass hypothesis. Furthermore, uncertainties on lepton trigger and identification efficiencies, V-tagging scale factors and luminosity are also considered and their effect on the signal event prediction is estimated. The systematic contribution from the pileup model is evaluated by re-weighting the simulated signal samples so that the distribution of the number of interactions per bunch crossing is shifted up and down by the uncertainty on minimum bias cross section, using what has been measured in data [132].
The impact of the proton PDF uncertainties on the signal acceptance is evaluated following the PDF4LHC [68] prescription, using MSTW and NNPDF PDF sets.
Finally, the effect of scale and resolution uncertainties of physics objects on the reconstructed m_{VV} signal shape has been evaluated, showing a small effect on both peak position and width of the Gaussian core. The jet energy scale and resolution introduce a relative uncertainty of about 3% (2%) in the signal width for the $\ell\nu$ + V-jet ($\ell\ell$ + V-jet) channel. In the $\ell\nu$ + V-jet

Source	V+jets		$t\bar{t}$		Single-top		Di-boson	
	$\ell\nu + \mathrm{V}$	$\ell\ell + \mathrm{V}$	$\ell\nu + \mathrm{V}$	$\ell\ell + \mathrm{V}$	$\ell\nu + \mathrm{V}$	$\ell\ell + \mathrm{V}$	$\ell\nu + \mathrm{V}$	$\ell\ell + \mathrm{V}$
Experimental sources								
Luminosity	-	-	2.6%	2.6%	2.6%	×	2.6%	2.6%
Muons trig+ID	-	-	2%	5%	2%	×	2%	5%
Ele trig+ID	-	-	3%	3%	3%	×	3%	3%
Muon scale	-	-	0.1%	0.2%	0.12%	×	0.15%	0.2%
Muon resolution	-	-	<0.1%	0.3%	<0.1%	×	<0.1%	0.5%
Ele scale	-	-	<0.1%	0.2%	<0.1%	×	<0.1%	0.15%
Ele resolution	-	-	<0.1%	0.15%	<0.1%	×	<0.1%	0.21%
Jet scale	-	-	1.5%	2%	2.5%	×	2.1%	1.8%
Jet resolution	-	-	0.45%	0.5%	0.5%	×	0.35%	0.4%
V-tagging	-	-	-	-	-	×	9 (24)%	9 (24)%
b-tagging	-	-	2.5%	2%	1.5%	×	-	-
Background estimate								
V+jet norm.	6 (7.2)%	7.5 (8.5)%	-	-	-	×	-	-
$t\bar{t}$ norm.	-	-	5 (6)%	5 (6)%	-	×	-	-
Single-top norm.	-	-	-	-	20%	×	-	-
Di-boson norm.	-	-	-	-	-	×	20%	20%

Table 4.7: Summary of the systematic uncertainties affecting background yields, relative to the expected number of background events. All systematic uncertainties in the list are treated as uncorrelated, with log-Normal priors in the ultimate statistical interpretation. Numbers inside square parenthesis refer to specific LP category values. Systematic effects that are varying the yield of a background source by less than 0.1% are neglected.

channel, the unclustered energy scale introduces a 1–3% uncertainty in the signal width, larger at low resonance masses. In the $\ell\ell + \mathrm{V}$-jet channel, the muon resolution causes an additional relative uncertainty of 2% in the signal width.
The uncertainty in the peak position of the signal is estimated to be less than 1%, except for the jet energy scale contribution which reaches at most 1.5%. Only the significant sources that are affecting the signal shapes are considered and treated as nuisance parameters in the final statistical interpretation. In Table 4.8, the primary uncertainties in both the signal shape and normalization are presented.

4.6 Statistical interpretation

Figure 4.16 shows the final observed m_{WW} spectrum for the selected events in the four categories of the $\ell\nu$+V-jet analysis. The observed data and the predicted background agrees with each other. The highest-mass event in the $\ell\nu$+V-jet channel is selected in the electron LP category, with a reconstructed m_{WW} around 3.2 TeV. This event is not included in the final statistical interpretation, which is performed up to a reconstructed mass of 3 TeV. The impact of this excluded event on the reported result is negligible since it belongs to a low sensitive category and limits are first interpreted in a narrow bulk graviton scenario, with a resonance mass up to 2.5 TeV. The observed event is compatible, at the $1\,\sigma$ level, with the background prediction for m_{WW} above 2.5 TeV.
The di-boson invariant mass m_{ZZ} distribution of the selected events in the $\ell\ell$+V-jet analysis is presented in Figure 4.17. Also in this case, an overall good description, in both normalization and shape, of the data by the background only estimation is observed. The highest mass event

Source	Normalization		Shape	
	$\ell\nu$ + V-jet	$\ell\ell$ + V-jet	peak position	width
Experimental sources				
Luminosity	2.6%	2.6%	-	-
Muons trig+ID	2%	5%	-	-
Muon scale	1%	2%	<0.1%	<0.7%
Muon resolution	<0.1%	0.5%	<0.1%	<0.1%
Electrons trig+ID	3%	3%	-	-
Electron scale	0.5%	0.5%	<0.1%	<0.2%
Electron resolution	0.1%	0.1%	<0.1%	<0.1%
Jet scale	1-3%	1%	<1.5%	<3%
Jet resolution	<0.5%	<0.1%	<0.1%	<2.5%
Unclustered energy scale	<0.5%	-	<0.1%	<3%
Pileup	0.5%	0.5%	-	-
V-tagging	9 (24)%	9 (24)%	-	-
Theoretical sources				
PDF	<0.5%	<0.5%	-	-

Table 4.8: Summary of systematic uncertainties affecting signal yield, relative to the expected number of observed signal events, and lineshape estimated after the full analysis selections. All systematic uncertainties in the list are treated as uncorrelated in the ultimate statistical interpretation. Their values for each specific mass point are not indicated, only the maximum range of variation as a function of the graviton mass is reported.

in the $\ell\ell$+V-jet channel belongs to the electron HP category, whichhas $m_{ZZ} = 2.6$ TeV.
To test the presence of a resonance decaying into vector bosons, a comparison between the m_{VV} distribution observed in data and the background prediction is used. Upper limits on the production cross section of a heavy narrow resonances decaying into the WW or ZZ pairs have been set, by combining the four event categories of the $\ell\nu$+V-jet analysis or the $\ell\ell$+V-jet one, respectively. The modified frequentist method (CL_S method), described in [149, 150], is adopted where signal and background yields are extracted by an unbinned shape analysis of the m_{VV} spectrum. Systematic uncertainties are treated as nuisance parameters and profiled in the statistical interpretation using log-normal priors, while Gaussian priors are used for shape parameters only.

4.6.1 Upper limit on narrow Bulk gravitons

Exclusion limits are set in the context of the Bulk graviton model, under the assumption of an intrinsic width negligible with respect to the experimental resolution. Figure 4.18 shows the expected and the observed 95% confidence level (CL) exclusion limits as a function of the graviton mass hypothesis, for the $\ell\nu$+V-jet (left) and $\ell\ell$+V-jet (right) channels. These limits are compared to the cross section multiplied by the branching fraction to WW and ZZ predictions for a Bulk graviton with $\tilde{k} = 0.2$ and $\tilde{k} = 0.5$.
These results were cross-checked with an alternative background estimation, directly extracted from data fitting to the m_{VV} distribution for the same selected events in the signal region with a smoothly falling function. This approach, already adopted in previously published CMS searches [127, 123, 151], provides results very close to the baseline method previously described, further strengthening our confidence in the robustness of the adopted background estimation method.

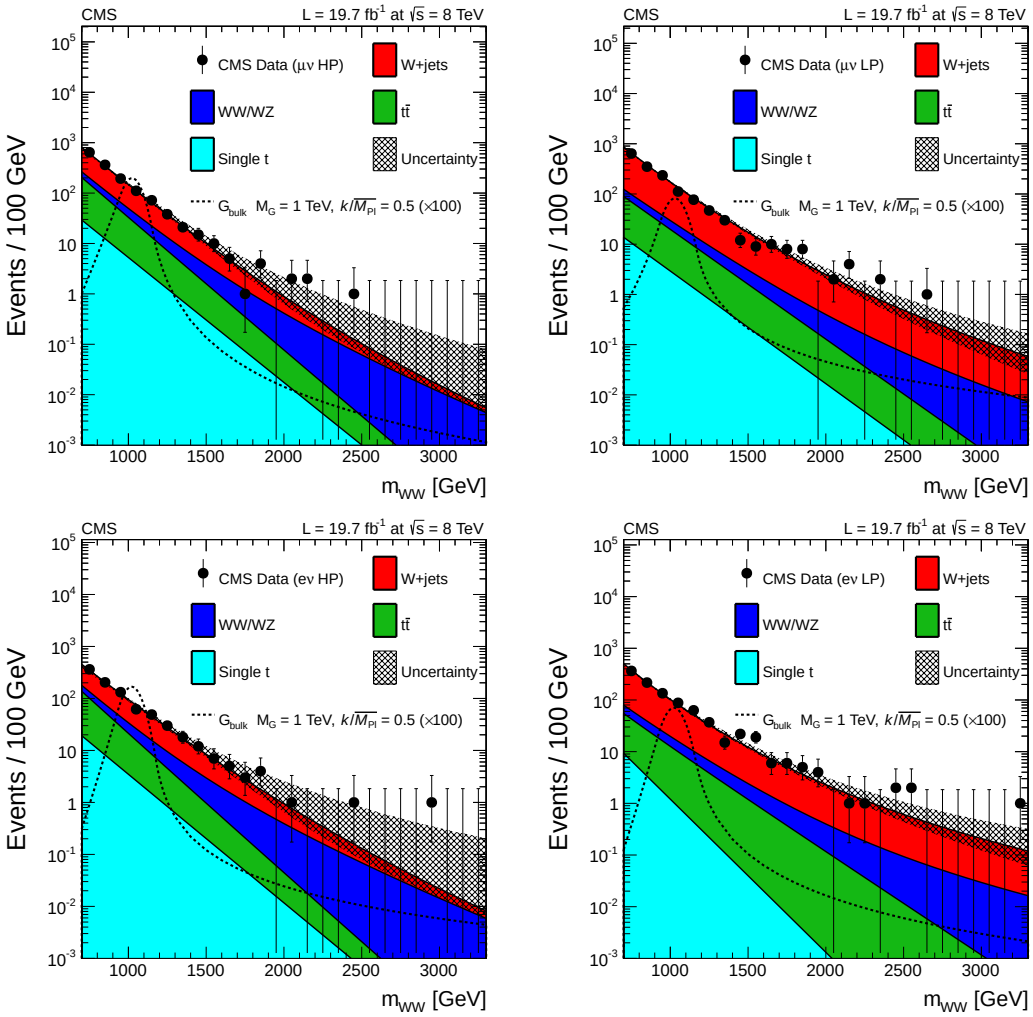

Figure 4.16: Final distributions in m_{WW} for data and expected backgrounds for both the muon (top) and the electron (bottom) channels, HP (left) and LP (right) categories for the $\ell\nu$+V-jet final state. On data points, the 68% error bars for Poisson event counts are obtained from the Neyman construction as described in [148]. An hypothetical Bulk graviton signal, with a mass of 1 TeV and $\tilde{k} = k/m_{pl} = 0.5$ is shown with a dashed black line. The normalization of the signal distribution is scaled up by a factor of 100 for a better visualization.

Eventually, $\ell\nu$+V-jet and $\ell\ell$+V-jet analyses are combined with a complementary CMS search in the VV $\to (q\bar{q}')(q\bar{q}')$ final state [123] (di-V-jet channel), to enhance the sensitivity of the search for this specific model. The fully hadronic analysis uses the same techniques to identify V-jets and the systematic uncertainties in jet energy scale and resolution, V-tagging scale factors and luminosity are considered correlated among the three channels in the final combination. The systematic uncertainties in electron and muon identification and energy scale/resolution are considered fully correlated between the $\ell\nu$+V-jet and $\ell\ell$+V-jet channels. The resulting 95% CL upper limits on the signal cross section are shown in Figure 4.19 (left). The $\ell\ell$+V-jet channel is the only one contributing to the limit for resonance masses below 800 GeV. Then, in the range 800–2500 GeV, the $\ell\nu$+V-jet channel dominates the sensitivity, as shown in Figure 4.19 (right), although the $\ell\ell$+V-jet and di-V-jet channels give significant contributions

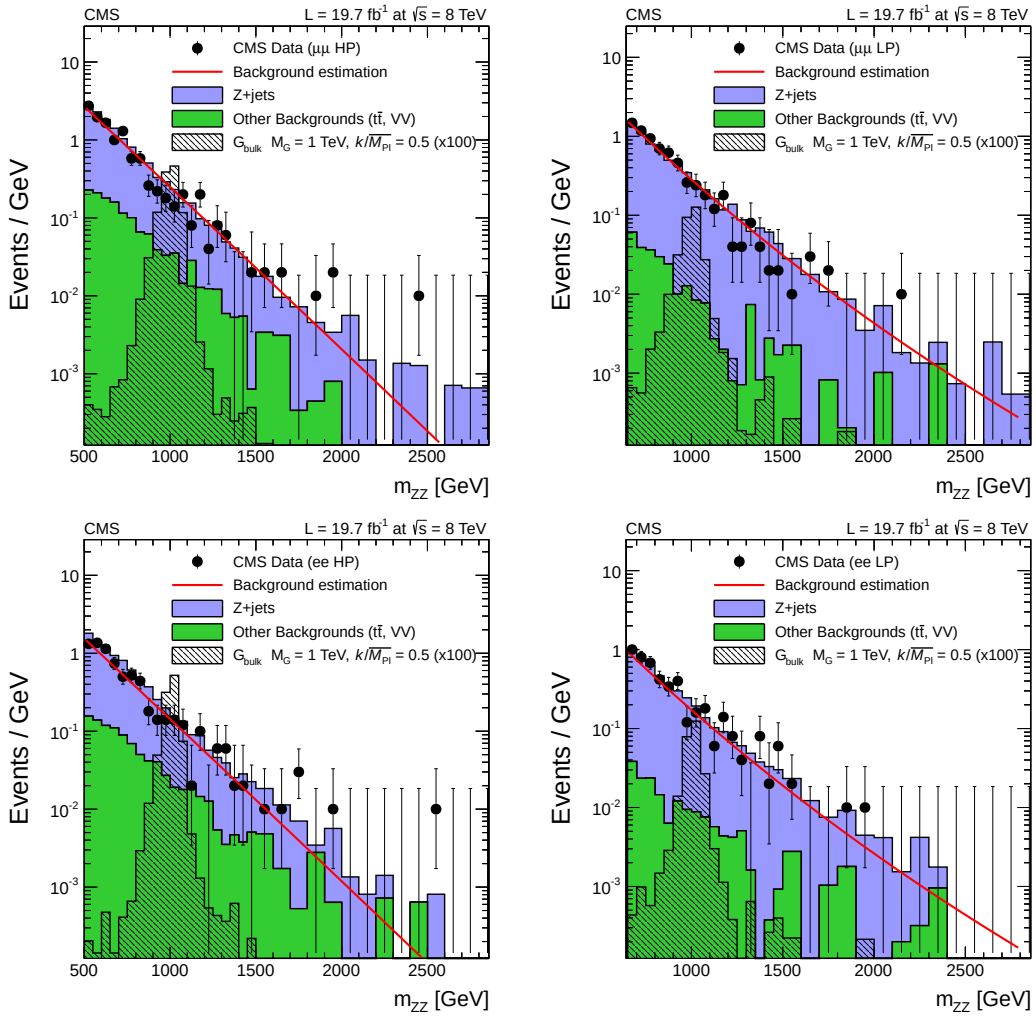

Figure 4.17: Final distributions in m_{ZZ} for data and expected backgrounds for both the muon (top) and the electron (bottom) channels, HP (left) and LP (right) categories for the $\ell\ell$+V-jet final state. On data points, the 68% error bars for Poisson event counts are obtained from the Neyman construction as described in [148]. An hypothetical Bulk graviton signal, with a mass of 1 TeV and $\tilde{k} = k/m_{pl} = 0.5$ is shown with a dashed black line. The normalization of the signal distribution is scaled up by a factor of 100 for a better visualization.

to the combined limit in the region below and above 1.3 TeV, respectively. Because of the combination, the expected upper limits on the cross section becomes more stringent by about 15–20%, compared to the individual $\ell\nu$+V-jet one, depending on the resonance mass. The integrated luminosity of the sample is not enough to reach the exclusion sensitivity of the Bulk graviton models with $k/m_{pl} = 0.2$ or 0.5. No excesses, with local significances larger than two standard deviations, are observed over the entire mass spectrum.

4.6.2 Model independent limits

As presented in previous Section 4.6.1, the analysis is specific to the case of a narrow Bulk graviton production, which is not the only possible extension of the SM predicting massive resonances decaying into vector boson pairs. Although gravitons are spin-2 particles, this

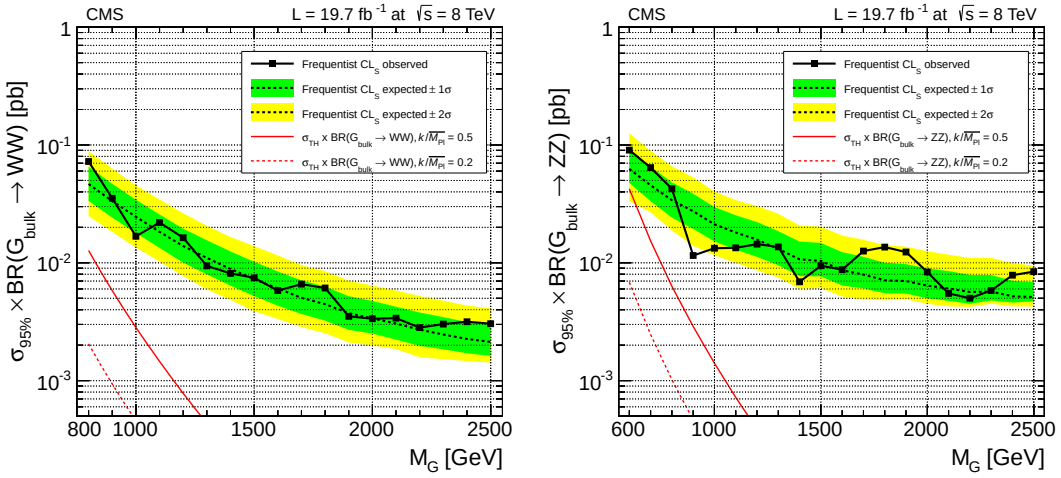

Figure 4.18: Observed (solid) and expected (dashed) 95% CL upper limits on the narrow Bulk graviton production cross section times the branching fraction of $G_{bulk} \to WW$ (left) and $G_{bulk} \to ZZ$ (right). The predicted cross section times branching fraction for the Bulk model is shown as a red solid (dashed) curve for $\tilde{k} = 0.5\,(0.2)$, respectively.

result can be interpreted in a generic spin-independent scenario since the selection strategy does not rely on the spin related properties, like angular correlations between decay products. Thus, to produce a model independent result without making any assumption on the resonance decay widths, exclusion limits on the visible number of signal events are evaluated. The final result becomes reproducible only if reconstruction and identification efficiencies for the

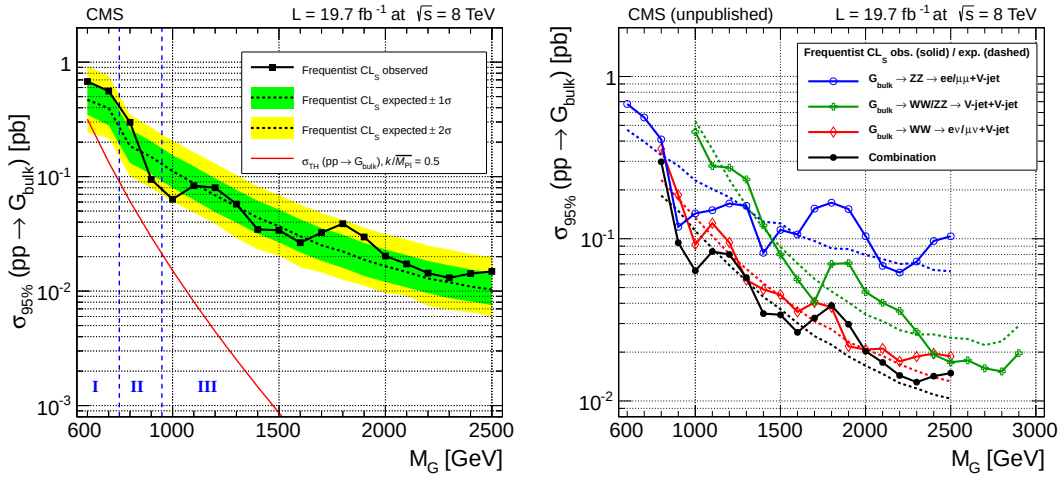

Figure 4.19: (Left) observed (solid) and expected (dashed) 95% CL upper limit on the Bulk graviton production cross section obtained combining $\ell\nu$+V-jet and $\ell\ell$+V-jet channels with the all-hadronic analysis [123]. The cross section for the production of a Bulk graviton with $\tilde{k} = 0.5$ is shown as a red solid curve. In region I, only the $\ell\ell$+V-jet channel contributes thanks to lower online thresholds. In the region II, both $\ell\ell$+V-jet and $\ell\ell$+V-jet channels contribute. Finally, in region III, both the semi-leptonic and all-hadronic analysis are combined. (Right) Observed (solid lines) and expected (dashed lines) 95% CL upper limit for each exclusive di-boson search ($\ell\nu$+V-jet, $\ell\ell$+V-jet and di-V-jet), as well as their statistical combination. The combination is shown only in the mass range where at least two analyses contribute.

final state objects, measured inside the kinematic acceptance of the analysis, are provided. Furthermore, to avoid any dependence on the assumptions used to define the analysis categorization, the event classification is reduced to one single category. Muon and electron channels are merged and the LP category is dropped, since the signal population in the N-subjettiness categories depends on intrinsic properties of the resonance model, like V-jets p_T spectrum and polarization.
Since a generic model cannot restrict to narrow lineshapes, limits in a bi-dimensional plane, as a function of both mass (m_X) and natural width (Γ_X) of the new resonance, are provided. The generated signal lineshape is parametrized according to a Breit–Wigner function (BW) with an intrinsic width Γ_X. The BW core is then convoluted with a double-sided CB, which is used to model the detector resolution. While different values of Γ_X are scanned for a given mass point, the parameters of the double-CB function are kept fixed to those determined in the narrow-width approximation. The scan is performed at regular steps of the relative width (Γ_X/m_X), starting from the narrow width limit up to $\Gamma_X/m_X = 0.40$, in steps of 0.05. It was checked, on specific simulated samples listed in Table 4.1, that the parametrization of the detector resolution factorizes from the resonance width, being stable when Γ_X increases.
Selection efficiencies are evaluated as a function of the vector boson kinematics (p_T,η), as they depend significantly on the production and decay properties of the new resonance. These efficiencies are extracted from Bulk graviton samples, generated for the baseline analysis.
At first, simulated signal events are pre-selected according to the acceptance requirements of the analysis, reported in Table 4.9, which are applied on generator level objects.

Object	**Selections**	
	$\ell\nu$+V-jet	$\ell\ell$+V-jet
Muons	$\lvert\eta\rvert < 2.1$	$\lvert\eta\rvert < 2.4$
	$p_T > 50$ GeV	$p_T(\ell_1) > 40$ GeV
		$p_T(\ell_2) > 20$ GeV
Electrons	$\lvert\eta\rvert < 2.5$	$\lvert\eta\rvert < 2.5$
	$p_T > 90$ GeV	$p_T > 40$ GeV
$\sum \vec{p}_T(\nu)$ (Muon ch.)	$\not{E}_T > 40$ GeV	×
$\sum \vec{p}_T(\nu)$ (Electron ch.)	$\not{E}_T > 80$ GeV	×
W$\to \ell\nu$ or W$\to \tau\nu \to \ell\nu\nu\nu$	$p_T^W > 200$ GeV	×
Z$\to \ell\ell$	×	$p_T^Z > 80$ GeV
	×	$70 < m_{\ell\ell} < 110$ GeV
V$\to qq'$	$\lvert\eta_W\rvert < 2.4$	$\lvert\eta_Z\rvert < 2.4$
	$p_T^W > 200$ GeV	$p_T^Z > 80$ GeV
	$65 < m_{qq'} < 105$ GeV	$70 < m_{qq'} < 110$ GeV
VV system	$0.7 < m_{WW} < 3$ TeV	$0.5 < m_{ZZ} < 2.8$ TeV
	$\Delta R(W_{qq'}, \ell) > \pi/2$	×
	$\Delta\phi(W_{qq'}, \sum \vec{p}_T(\nu)) > 2$	×
	$\Delta\phi(W_{qq'}, W_{\ell\nu}) > 2$	×

Table 4.9: Generator level requirements used to define the acceptance region for the WW and ZZ analysis. The vector sum of the transverse neutrino momenta $\sum \vec{p}_T(\nu)$ is taken over all the neutrinos in the final state, coming either from W $\to \ell\nu$ or W $\to \tau\nu \to \ell\nu\nu\nu$ decays, where $\ell = \mu, e$.

Then, starting from pre-selected events, the reconstructed V-boson candidates are independently checked to fulfill the analysis selection, evaluating the efficiency in bins of p_T and η of the V-boson prior to any simulation of detector effects. The binning is optimized in order to minimize the statistical uncertainty and bins with less than 25 events passing the selections are excluded from the final fit. The impact of these regions is very limited, since they are located in extreme corners of phase space. The V-boson efficiency parametrization includes also all the residual tunings of the simulation to the data, like lepton identification, trigger efficiencies, b-tagging and V-tagging scale factors. The efficiencies of the additional lepton and b-jet vetoes in the $\ell\nu$+V-jet analysis are found to be independent to the di-boson system kinematic, thus a constant efficiency of 91.5% for the b-jet veto and 98.3% for the second-lepton veto are used.
Eventually, the total signal efficiency and the pre-selection acceptance are found to be weakly dependent on the resonance width Γ_X. Figure 4.20 and Figure 4.21 show the resulting efficiencies for both leptonically and hadronically decaying W and Z-bosons, considering only longitudinal polarization state. No contribution from $Z \to \tau\tau \to 2\ell 4\nu$ is taken into account, since hugely suppressed by the $m_{\ell\ell}$ requirement.
As already detailed in Section 3.2.6, V-tagging performances are strongly affected by the decaying V-boson polarization, thus a special care must be given to transversely polarized vector bosons. This effect is studied comparing the efficiencies reported above, with the ones obtained from a set of `MadGraph` `RS1` graviton events, where V-bosons produced from their decays are transversely polarized in more than 90% of the cases. For leptonically decaying W or Z-bosons, the generator level selections embed polarization effects inside the acceptance definition. On the V-jet side, the loss in efficiency has been measured to be largely independent of the V-boson kinematic properties, so that the effect of the transverse polarization is adequately modeled by a constant scale factor of 85%, independent from the kinematic properties (p_T and η) of the $V \to qq'$ decay system.
To validate the procedure, the resulting parametrized efficiencies are used to predict the total signal efficiency for reconstructing Bulk or `RS1` gravitons, comparing this estimation with the exact number obtained from the baseline analysis performed on the fully simulated samples. In all cases, the predicted signal yields from the two methods agrees within 15%, and this deviation is considered as conservative systematics uncertainty on the signal rate when evaluating the model independent limits.
Exclusion limits at 95% `CL` on the number of signal events are reported in Figure 4.22 for both the model independent $\ell\nu$+V-jet analysis and the $\ell\ell$+V-jet one.
Results are not combined to avoid assumptions on the branching fractions of the hypothetic resonance decaying to both WW and ZZ channels. The limits are calculated using an asymptotic approximation of the CL_S method [152]. Under the narrow-width approximation, it is explicitly checked that the central values for the expected and observed limits obtained with the full frequentist method and the asymptotic approximation match over all the mass range. All the systematic uncertainties, considered in the baseline analysis, are included in the calculation of these limits. With increasing width Γ_X, statistical fluctuations in the limit tend to be smoothed out while the overall performance degrades. In fact, for relative widths (Γ_X/m_X) larger than 0.25, the deterioration of the limit is very mild because the sensitivity, coming from the knowledge of the signal shape, is diluted by the large width.
Finally, although optimized for WW and ZZ final states, these analyses are also sensitive to

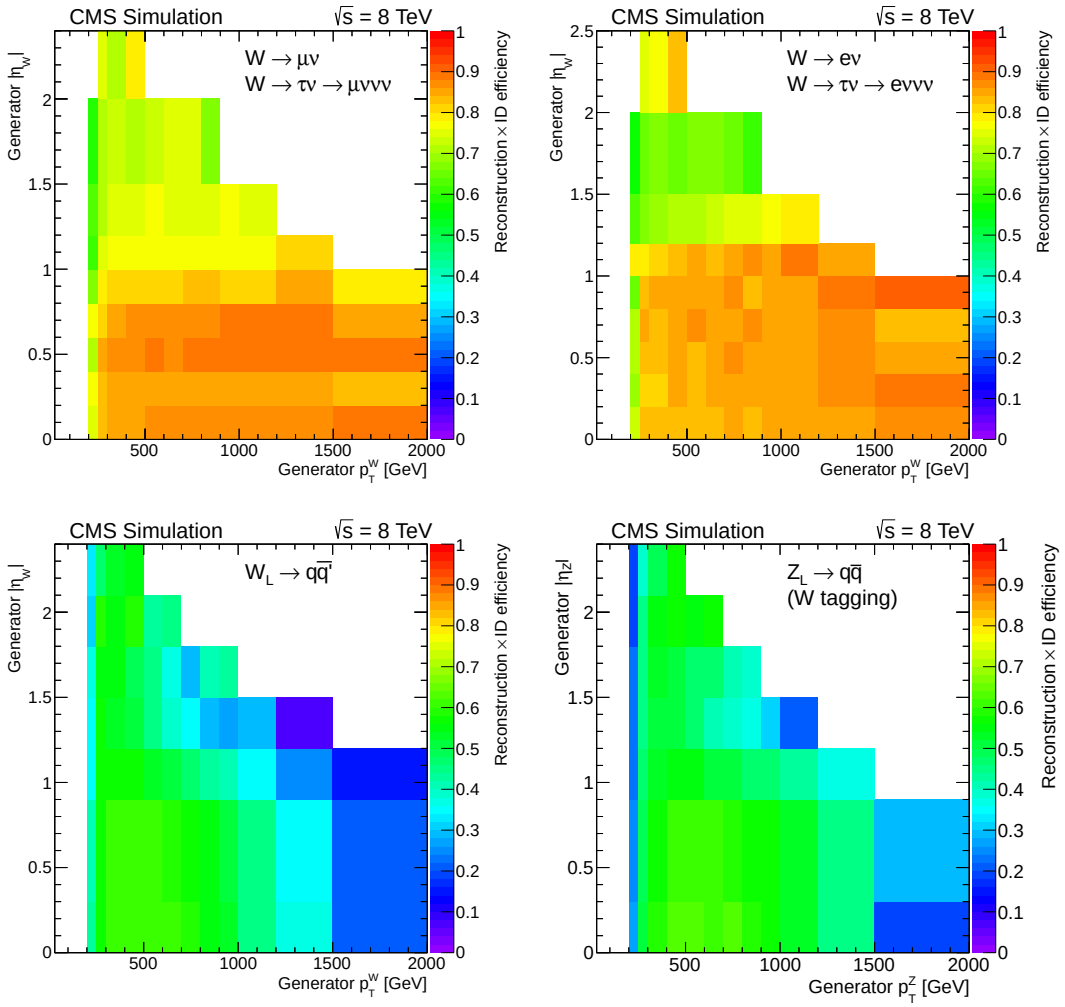

Figure 4.20: Reconstruction and identification efficiencies for $W \to \mu\nu$ and $W \to \tau\nu \to \mu\nu\nu\nu$ (top left), $W \to e\nu$ and $W \to \tau\nu \to e\nu\nu\nu$ (top right), $W_L \to q\bar{q}'$ (bottom left) and $Z_L \to q\bar{q}$ (bottom right) decays as function of generated p_T^V and $|\eta_V|$ using the W-tagging requirements for the hadronic V-boson.

charged resonances decaying to WZ semi-leptonic final states ($\ell\ell$+W-jet or $\ell\nu$+Z-jet), because of the poor V-jet pruned mass resolution which prevent to separate at best W and Z-boson jets. The efficiencies to identify a longitudinally polarized W (Z)-boson using V-tagging requirements are computed from the Bulk graviton samples, as already shown in Figure 4.20 and Figure 4.21. The effect of the transverse polarization of the vector bosons is modeled by multiplying the aforementioned efficiencies by the constant scale factor, measured to be 0.85. In addition, in the $\ell\nu$+V-jet channel, the combined efficiency of the second-lepton veto and b-jet veto is equal to 81%, because of the presence of $Z \to b\bar{b}$ decays, which are rejected by the requirement of the analysis that no jet is tagged as coming from a b-quark.

4.6.3 Generic model independent re-interpretation

The results presented in the previous Section can be used to evaluate the signal yield expected to be observed in the CMS detector coming from a generic model, that predicts a resonance

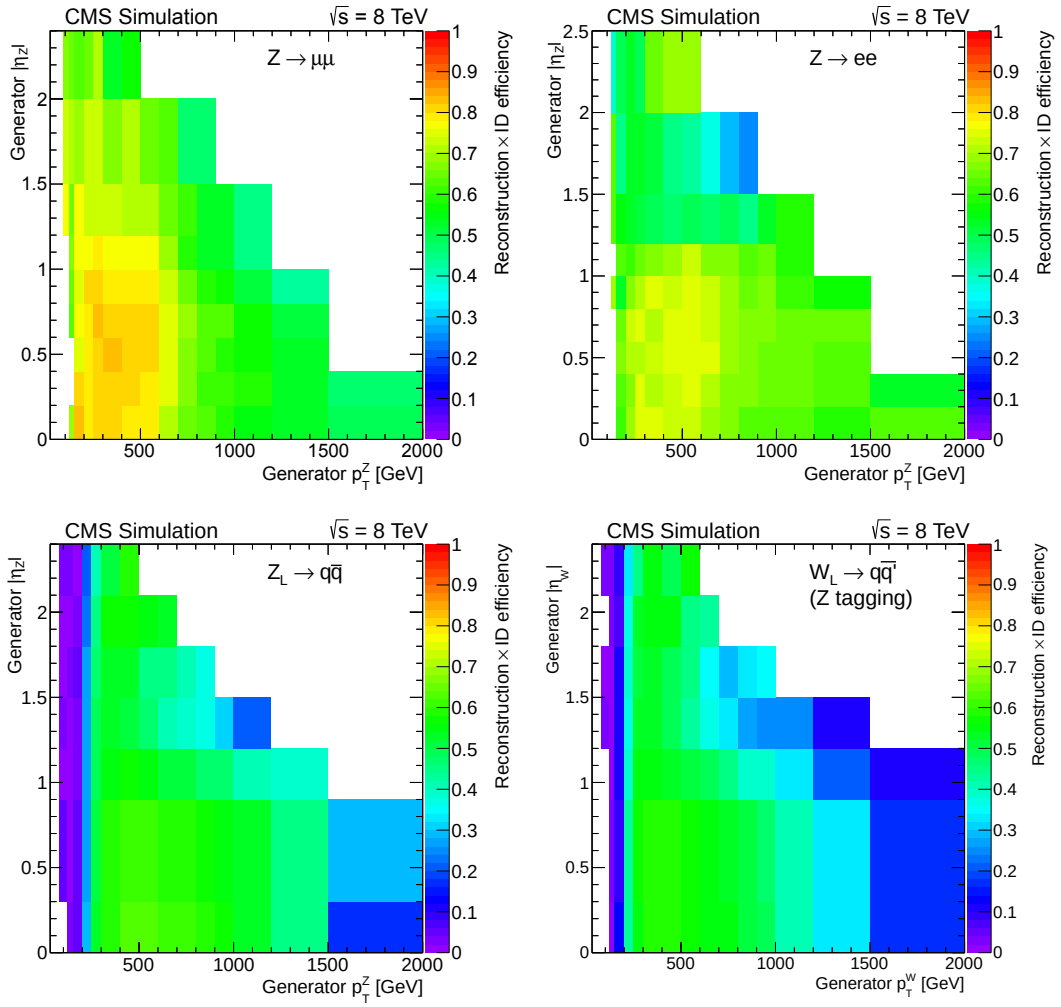

Figure 4.21: Reconstruction and identification efficiencies for the $Z \to \mu\mu$ (top left), $Z \to ee$ (top right), $Z_L \to q\bar{q}$ (bottom left) and $W_L \to q\bar{q}'$ (bottom right) decays as function of generated p_T^V and $|\eta_V|$ using the Z-tagging requirements for the hadronic V-boson.

X decaying into $\ell\nu$+V-jet or $\ell\ell$+V-jet final states. The computed efficiencies are applied as follows:

- According to the chosen model, a set of events are simulated for a given mass (m_X) and width (Γ_X) decaying into a final state with leptons and quarks.
- Generated events are filtered applying the selections listed in Table 4.9, which define the acceptance region of the analysis.
- For each surviving event, the efficiency of reconstructing the leptonically decaying vector boson (ε_ℓ) is evaluated, as well as the one for the hadronically decaying leg (ε_q), taking into account the V-boson polarization state.
- Each event is then weighted by the product of the two identification efficiencies ($\varepsilon_\ell \times \varepsilon_\ell$). In case of the semi-leptonic WV final state is considered ($X \to WW \to \ell\nu$+V-jet), the efficiencies should be combined also with the b-jet and second lepton veto ones.

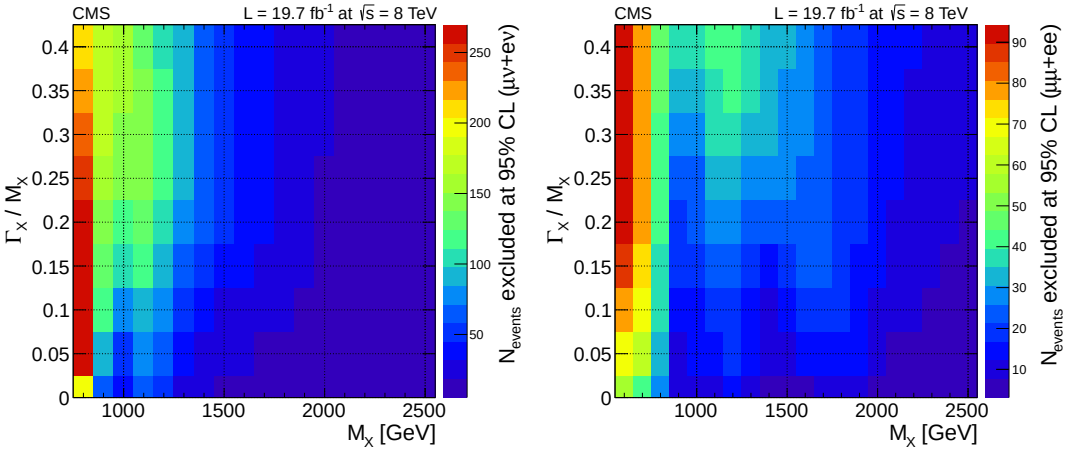

Figure 4.22: Observed exclusion limits at 95% CL on the number of signal events coming from a resonance decaying into WV $\to \ell\nu$+V-jet (left) or ZV $\to \ell\ell$+V-et jet (right), as a function of its mass (m_X) and normalized width (Γ_X/m_X).

- The sum of weights divided by the total number of events in the acceptance region provides an estimation of the total efficiency for the given model.
- The number of signal events, extracted from this procedure, can be compared with the observed limit reported in Figure 4.22 to estimate the exclusion sensitivity to the considered model.

4.7 Conclusion

A search for new heavy resonance decaying into WW, ZZ or WZ pairs has been presented, in which one of the vector bosons decays leptonically and the other one hadronically, i.e. considering $\ell\nu$+V-jet or $\ell\ell$+V-jet final states ($\ell = \mu, e$). Jet substructure are used to reduce backgrounds from V+jets, $t\bar{t}$, di-boson and multi-jet processes.
No significant excess over the background prediction is found, thus upper limits on the production cross section are evaluated as a function of the resonance mass hypothesis, in context of narrow Bulk gravitons. For this specific case, the semi-leptonic channels are combined with complementary fully hadronic searches, resulting in upper limits on the narrow ($\tilde{k} < 0.5$) Bulk graviton production cross section in the range from 700 to 10 fb for masses between 600 and 2500 GeV, respectively.
Finally, the semi-leptonic analyses are re-interpreted in a simplified scenario, providing model-independent limits on the observed number of excluded events at 95% CL. The re-interpretation of these exclusion limits in a generic phenomenological model, including charged resonance decaying into WZ final states ($W' \to WZ$), is allowed by the V-boson reconstruction and identification efficiencies defined in the kinematic acceptance of the analysis.

Chapter 5

Search for a SM-like Higgs decaying into WW $\rightarrow \ell\nu q\bar{q}'$ in exclusive jet bins

The observation of a Higgs boson, with a mass of about 125 GeV and properties (production and decay rates, spin-parity quantum numbers) compatible with those expected from the SM, is consistent with the unitarity constraints on di-boson scattering at high energies [49, 153, 154, 155]. Nevertheless, there is still a possibility that the newly discovered particle is part of a larger Higgs sector, only partially responsible for the EWSB. Several BSM scenarios predict the existence of additional massive resonances, with couplings similar to the SM Higgs boson ones, as described in Section 1.4.

Previous searches at the LHC for heavy SM-like Higgs bosons have been reported by both the ATLAS and the CMS experiments. ATLAS measurements excludes, at 95% CL, the existence of a SM-like heavy Higgs boson within the mass range $131 < m_{\mathrm{H}} < 559$ GeV [156]. The CMS collaboration reported a search in the WW and ZZ decay channels, combining 5.1 fb^{-1} at $\sqrt{s} = 7$ TeV with 5.3 fb^{-1} at $\sqrt{s} = 8$ TeV, exploring the mass range $145 < m_{\mathrm{H}} < 1000$ GeV and excluding SM-like Higgs boson masses up to 710 GeV [157].

This Chapter reports the search for a SM-like Higgs boson decaying into W-boson pairs in the mass range $600 < m_{\mathrm{H}} < 1000$ GeV, looking at the semi-leptonic decay mode and considering both gluon fusion and VBF productions [158]. The semi-leptonic topology is triggered and separated from the QCD multi-jet production thanks to the presence of one high p_T lepton ($\ell = \mu, e$), coming from the leptonic decay of one W-boson. In addition, this decay mode shows the largest branching fraction for high mass heavy Higgs searches, due to the hadronic decay of the other W-boson. Because of the large invariant mass searched for, the W-bosons have a large transverse momentum, which causes the hadronic decay products to be collimated to the point of being reconstructed as a single jet in the detector. Jet substructure techniques are therefore employed for identifying single jets originating from a highly boosted hadronically decaying W-boson, as adopted in the searches described in Chapter 4.

This analysis also features the first selection of the vector boson scattering topology at high masses, paving the way for future investigations of the WW scattering process in the boosted semi-leptonic final state.

The analysis has been performed on the LHC pp collision data recorded by the CMS detector at the center-of-mass energy of 8 TeV, corresponding to an integrated luminosity of 19.3 fb^{-1}.

The obtained results are interpreted as a search for a SM-like heavy Higgs boson, as well as in the context of an electroweak singlet extension of the SM, based on an effective theory where a second scalar boson, besides the resonance at 125 GeV, completes the unitarization of the WW scattering, as already described in Section 1.4.
In Section 5.1 and Section 5.2, an overview on the background and the signal simulated samples, datasets and triggers is given, with particular emphasis on how the interference between the Higgs signal and the continuum SM di-boson production has been accounted for, for both gluon fusion and VBF production modes. Section 5.3.3 is dedicated to the description of the analysis strategy, the event selection and categorization.
In Section 5.4, the adopted signal modelling is described, while Section 5.5 is focused on the background extraction. Systematic uncertainties and the statistical interpretation are reported in Section 5.6 and Section 5.7, respectively.
Eventually, in Section 5.8, the ultimate CMS combination, performed in the mass range $145 < m_{\mathrm{H}} < 1000$ GeV, is presented [159] exploiting both H → WW and H → ZZ decay channels and considering integrated luminosities of up to 5.1 fb^{-1} at $\sqrt{s} = 7$ TeV and up to 19.7 fb^{-1} at $\sqrt{s} = 8$ TeV. In the case of a Higgs candidate decaying into a pair of W-bosons, fully leptonic (H → WW → $\ell\nu\ell\nu$) and semi-leptonic (H → WW → $\ell\nu qq'$) final states are considered, while for a Higgs boson decaying into two Z-bosons, final states with four charged leptons (H → ZZ → $2\ell 2\ell'$), two charged leptons and two quarks (H → ZZ → $2\ell 2q$) and two charged leptons and two neutrinos (H → ZZ → $2\ell 2\nu$) are used, where $\ell = (e, \mu)$ and $\ell' = (e, \mu, \tau)$.

5.1 Datasets, triggers and background simulation

The datasets used in the analysis corresponds to an integrated luminosity of 19.3 fb^{-1}. Data have been collected by the CMS experiment via single lepton triggers, with a typical on-line p_T threshold of 24 GeV for muons and 27 GeV for electrons. The pseudorapidity range for muons (electrons) extends up to $|\eta| < 2.4$ (2.5). The trigger efficiency is around 94% (90%) for muons (electrons) and it is used to correct simulated events taking into account its small dependence, at the level of few percent, measured as a function of p_T and η.
The adopted simulated background samples are reported in Section 3.2.1, since also in this analysis the boosted $\ell\nu$+V-jets topology is explored. As observed in the search for heavy gravitons in the $\ell\nu$+V-jet final state, described in Chapter 4, the largest background contribution is represented by leptonically decaying W-bosons produced in association with jets (W+jets), followed by semi-leptonic decays of top quark pairs. In addition to the background samples listed in Table 3.1, exclusive W-boson production in association with two, three or four jets at matrix element are considered, to increase the available MC statistics in the VBF analysis, as well as $t\bar{t}$ and WW electroweak productions ($\mathrm{O}(\alpha_{ew}^6)$) generated via Phantom [160].

5.2 Signal simulation

The SM-like Higgs boson signal samples have been produced with POWHEG, at NLO QCD+EWK, interfaced with Pythia 6 for parton showering and underlying event simulation, considering both the gluon fusion (ggH) and the vector boson fusion (qqH) mechanisms.
The Higgs lineshape is re-weighted according to the complex pole scheme (CPS) calculation for the Higgs boson propagator [161, 162, 163] for all the considered mass hypotheses, ranging

from 600 to 1000 GeV. The lineshape of an unstable particle is usually well described by a Breit-Wigner distribution, within an uncertainty of the order of Γ/m, which means that this model breaks down in the high mass region, in particular when $m_{\mathrm{H}} > 600$ GeV. For this reason, the Breit-Wigner approximation must be overtaken through the CPS re-weighting, which is computed and applied on an event-by-event basis during the generation. The gluon fusion and VBF inclusive cross sections are normalized to the NNLO QCD and next-to-next-to-leading-log (NNLL) predictions, as reported in Section 1.3.

5.2.1 Higgs interference with the di-boson production

The background continuum ($gg, qq \to \mathrm{WW}$) and its interference with heavy Higgs production are only known at LO [35], which is not consistent with the NLO signal amplitude accuracy provided by POWHEG. To embed higher order corrections in an effective way, providing the correct signal shape and minimizing the uncertainty due perturbative orders in the calculation, three alternative options are considered to correct the Higgs lineshape for interference effects:

- Additive correction:

$$\frac{d\sigma^{\mathrm{NLO}}}{dm_{\mathrm{WW}}} = \frac{d\sigma_{\mathrm{S}}^{\mathrm{NLO}}}{dm_{\mathrm{WW}}} + \frac{d\sigma_{\mathrm{I}}^{\mathrm{LO}}}{dm_{\mathrm{WW}}} + \frac{d\sigma_{\mathrm{B}}^{\mathrm{LO}}}{dm_{\mathrm{WW}}} \tag{5.1}$$

  where $d\sigma_{\mathrm{S}}^{\mathrm{NLO}}/dm_{\mathrm{WW}}$ stands for the NLO heavy Higgs lineshape, $d\sigma_{\mathrm{I}}^{\mathrm{LO}}/dm_{\mathrm{WW}}$ is the LO interference, while $d\sigma_{\mathrm{B}}^{\mathrm{LO}}/dm_{\mathrm{WW}}$ is the LO background only generation ($gg, qq \to \mathrm{WW}$).

- Multiplicative correction:

$$\frac{d\sigma^{\mathrm{NLO}}}{dm_{\mathrm{WW}}} = k_{\mathrm{NLO}} \cdot \left(\frac{d\sigma_{\mathrm{S}}^{\mathrm{LO}}}{dm_{\mathrm{WW}}} + \frac{d\sigma_{\mathrm{I}}^{\mathrm{LO}}}{dm_{\mathrm{WW}}}\right) + \frac{d\sigma_{\mathrm{B}}^{\mathrm{LO}}}{dm_{\mathrm{WW}}} \quad k_{\mathrm{NLO}} = \frac{d\sigma_{\mathrm{S}}^{\mathrm{NLO}}}{dm_{\mathrm{WW}}} / \frac{d\sigma_{\mathrm{S}}^{\mathrm{LO}}}{dm_{\mathrm{WW}}} \tag{5.2}$$

- Intermediate correction:

$$\frac{d\sigma^{\mathrm{NLO}}}{dm_{\mathrm{WW}}} = k_{\mathrm{NLO}} \cdot \left(\frac{d\sigma_{\mathrm{S}}^{\mathrm{LO}}}{dm_{\mathrm{WW}}}\right) + \sqrt{k_{\mathrm{NLO}}} \cdot \frac{d\sigma_{\mathrm{I}}^{\mathrm{LO}}}{dm_{\mathrm{WW}}} + \frac{d\sigma_{\mathrm{B}}^{\mathrm{LO}}}{dm_{\mathrm{WW}}} \tag{5.3}$$

where the intermediate option is used as central value for the lineshape, while the difference between the additive and the multiplicative ones gives the uncertainty on the interference estimate.
For the gluon fusion production, signal normalization and lineshape are corrected for the interference with the SM $gg \to \mathrm{WW} \to l\nu q\bar{q}'$ production, whose effect increases as a function of the Higgs mass hypothesis. The interference contribution is evaluated with MCFM v6.4 [164] at LO, where the resulting cross section is written as:

$$d\hat{\sigma}_{\mathrm{I}} = \frac{\hat{s} - m_{\mathrm{H}}^2}{(\hat{s} - m_{\mathrm{H}}^2)^2 + m_{\mathrm{H}}^2\Gamma_{\mathrm{H}}^2} \mathrm{Re}\left(2\tilde{A}_{\mathrm{H}} A_{\mathrm{B}}^*\right) + \frac{m_{\mathrm{H}}\Gamma_{\mathrm{H}}}{(\hat{s} - m_{\mathrm{H}}^2)^2 + m_{\mathrm{H}}^2\Gamma_{\mathrm{H}}^2} \mathrm{Im}\left(2\tilde{A}_{\mathrm{H}} A_{\mathrm{B}}^*\right) \tag{5.4}$$

where the second term is always positive and proportional to the Higgs width, while the first one is constructive only for $\hat{s} < m_{\mathrm{H}}^2$, destructive for larger $\hat{s}$ values.
Figure 5.1 (left) [35] shows a comparison between different simulations: MCFM ggH at LO (red), ggVV background only at LO (pink) and the total generation with interference (S+B+I at LO), considering a SM-like Higgs with $m_{\mathrm{H}} = 900$ GeV. The interference term is separated by

subtracting signal and background distributions from the S+B+I generation. Then applying a set of NLO k-factors (k_{NLO}) as a function of m_{VV}, the interference corrected Higgs lineshape is reported in Figure 5.1 (right), where the three different corrections previously described are compared.

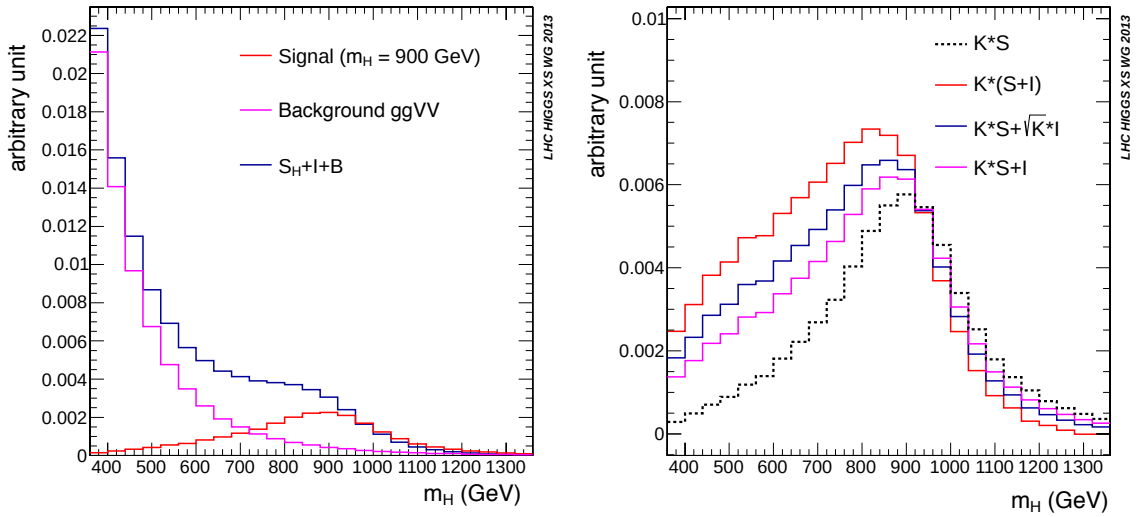

Figure 5.1: (Left) Invariant mass of the di-boson system in $gg(\mathrm{H}) \rightarrow \mathrm{VV}$ events for high mass SM-like Higgs production (red), with $m_{\mathrm{H}} = 900$ GeV, di-boson continuum background (pink) and signal+background+interference (blue). (Right) Comparison between the Higgs boson lineshape at NLO before any interference correction (black dashed) and interference corrected ones, applying additive or multiplicative or intermediate corrections.

Thus, the interference has a huge impact at LO, changing both the normalization and the signal shape as expected from Equation 5.4.
Also in the VBF Higgs production, the interference effects with the continuum SM $qq \rightarrow \mathrm{WW}qq$ production have to be included to obtain a correct description of the signal lineshape, in particular when the intrinsic width of the Higgs candidate becomes large. Also in this case, the adopted strategy consists in re-weighting the POWHEG NLO generation as follows:

- Generate background only events (B, $qq \rightarrow qq\mathrm{WW}$) at LO in $\mathrm{O}(\alpha_{ew^6})$ and $\mathrm{O}(\alpha_{ew}^4\alpha_s^2)$ with Phantom.
- Generate Signal+Background+Interference ($\mathrm{S_H + B + I_H}$ via Phantom, always at LO in both $\mathrm{O}(\alpha_{ew^6})$ and $\mathrm{O}(\alpha_{ew}^4\alpha_s^2)$.
- Generate signal only events ($\mathrm{S_H}$) at LO through MadGraph, since a VBF Higgs signal-only generation was not available in the considered Phantom version.
- The interference between the high mass scalar (H) and the low mass SM Higgs boson (h) is neglected, since the off-shell Higgs boson contribution is rather small compared to the $qq \rightarrow qq\mathrm{WW}$ production, thus this interference term is expected to be small.

By subtracting the background only contribution, the generated m_{WW} distribution for $\mathrm{S_H + I_H}$ and $\mathrm{S_H}$ are fitted with suitable parametrizations, as shown in Figure 5.2 for $m_{\mathrm{H}} = 800$ GeV.

Finally, the interference weight factor is calculated as a continuous function of m_{WW}:

$$\omega(m_{\mathrm{WW}}) = \frac{(\mathrm{S_H} + \mathrm{I_H})(m_{\mathrm{WW}})}{\mathrm{S_H}(m_{\mathrm{WW}})} \tag{5.5}$$

The uncertainty on the interference weights is evaluated by varying the renormalization and factorization scales in the range $m_{\mathrm{H}}/2 < (\mu_r, \mu_f) < 2 \cdot m_{\mathrm{H}}$, showing an effect at the order of 10% on the normalization with a much smaller shape variation.

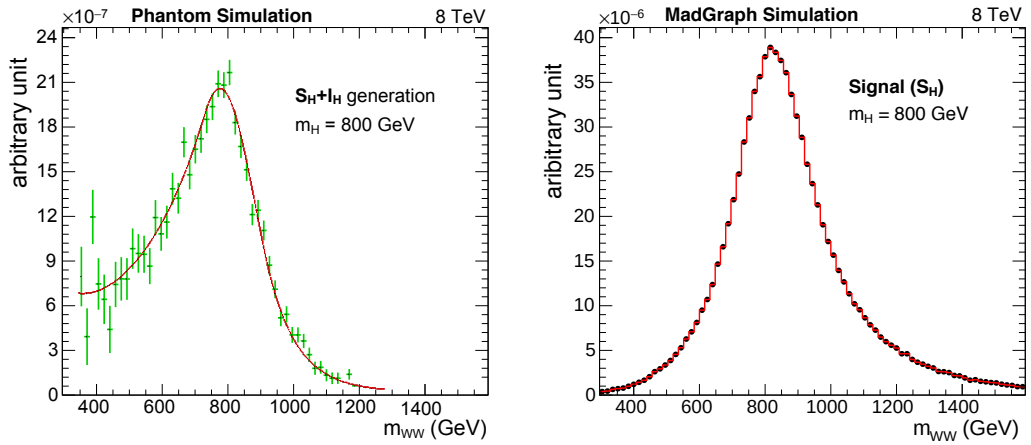

Figure 5.2: Di-boson invariant mass distribution (m_{WW}) for $\mathrm{S_H} + \mathrm{I_H}$ (left) and $\mathrm{S_H}$ (right) as generated at LO by Phantom and MadGraph, respectively, considering a SM-like Higgs boson with $m_{\mathrm{H}} = 800$ GeV produced via VBF.

5.2.2 Electroweak singlet lineshape

In the BSM electroweak singlet model, a heavy scalar boson mixes with the low mass Higgs state and, phenomenologically, the couplings of the two gauge eigenstates become related by unitarity. Thus, the effective coupling strength of the light Higgs boson is reduced with respect to the pure SM case. Defining C_{V} (C'_{V}) as the scale factor couplings of the low (high) mass Higgs to vector bosons, the unitarity condition is expressed as $C_{\mathrm{V}}^2 + C_{\mathrm{V}}'^2 = 1$, which has to be preserved to avoid unitarity violation in the $\mathrm{V_L V_L} \to \mathrm{V_L V_L}$ scattering. In addition, to preserve the unitarization of the SM also when couplings of the two Higgs states with fermions are considered, C_f and C'_f respectively, the following constraint is also required: $C_{\mathrm{V}} C_f + C'_{\mathrm{V}} C'_f = 1$.
When a common scale factor for both fermions and V-bosons is considered, $C = C_{\mathrm{V}} = C_f$, the heavy scalar signal strength (μ') and width (Γ') are defined according to Equation 1.35, where $k = \mu'/\mu_{\mathrm{SM}}$ and $\cos C^2 = \alpha^2$:

$$\Gamma' = C'^2 \cdot \Gamma^{\mathrm{SM}} + \Gamma^{\mathrm{new}} = \frac{C'^2}{1 - \mathrm{BR_{new}}} \Gamma^{\mathrm{SM}} \quad \mu' = \frac{\sigma' \cdot \mathrm{BR}'}{\sigma_{\mathrm{SM}} \cdot \mathrm{BR_{SM}}} = C'^2 (1 - \mathrm{BR_{new}}) \mu \tag{5.6}$$

Indirectly, an upper limit at 95% CL is set on C'^2 to be less than 0.28, from the signal strength fits to the SM Higgs boson at low mass, as exposed in Section 1.4.
Equation 5.6 states that the considered BSM scenario can be obtained by the SM-like one by a re-weighting of the Higgs lineshape to an arbitrary width. The BSM re-weighting is applied on top of the CPS correction, before any inclusion of interference effects, since these strictly depend

on the signal width. The re-weighting is performed analytically, fitting the CPS lineshape with a proper model represented by a Breit-Wigner with a running width defined as:

$$F_{\mathrm{BW}}(s = m^2_{\mathrm{WW}}) = \frac{s \cdot \Gamma/m_{\mathrm{H}}}{(s - m^2_{\mathrm{H}})^2 + (s \cdot \Gamma/m_{\mathrm{H}})^2} \tag{5.7}$$

The fitted width Γ is scaled according to Equation 5.6 for each chosen combination of C'^2 and $\mathrm{BR}_{\mathrm{new}}$. In Figure 5.3, for gluon fusion Higgs production with $m_{\mathrm{H}} = 700$ GeV and 900 GeV, black points represent the m_{WW} distribution as generated by POWHEG, while the red and blue dots are the complex pole scheme and CPS plus interference shapes, respectively. Moreover, the dashed red line represents the analytic fit to the CPS lineshape, while the green lines are obtained by scaling the width according to the chosen C'^2 and $\mathrm{BR}_{\mathrm{new}}$ values.

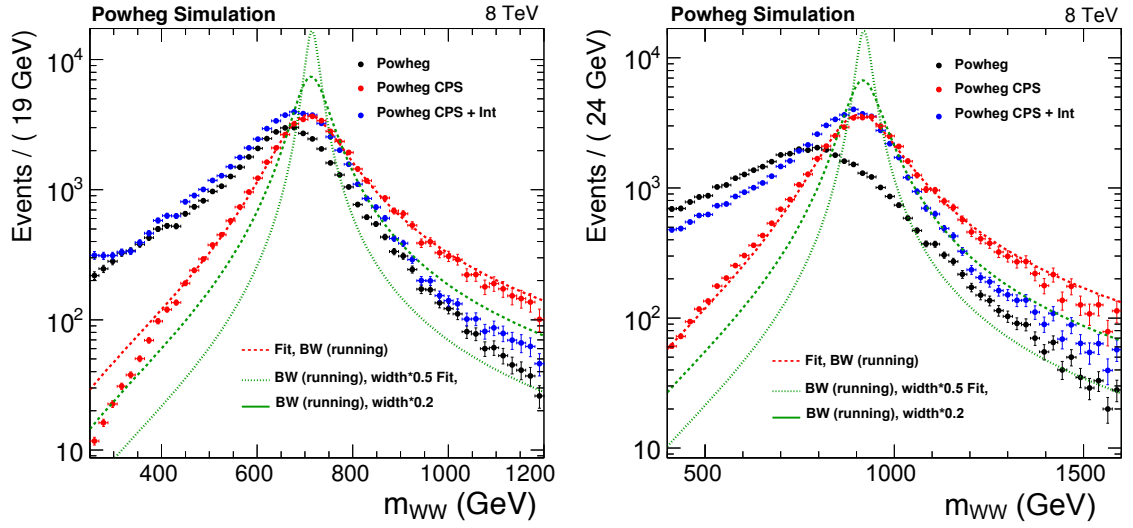

Figure 5.3: SM-like Higgs lineshapes for $m_{\mathrm{H}} = 700$ GeV (left) and 900 GeV (right): black points represents the m_{WW} distribution as generated by POWHEG, while red and blue points are the result obtained after CPS and CPS plus interference re-weight. Dashed red line is the Breit-Wigner running width fit to the CPS Higgs lineshape, dashed green lines are the BSM lineshapes obtained scaling the fitted width according to: $\Gamma' = \Gamma^{\mathrm{SM}} \cdot C'^2/(1 - \mathrm{BR}_{\mathrm{new}})$.

The interference correction must be recomputed for both ggH and qqH production modes after modifying couplings and signal width. For gluon fusion, MCFM is used again to evaluate this contribution for each set of $(C', \mathrm{BR}_{\mathrm{new}})$ values considered in the analysis, then interpolated on the whole $(C', \mathrm{BR}_{\mathrm{new}})$ plane for each mass value. As expected, the effect of the interference is much reduced on both shape and normalization for low C' and $\mathrm{BR}_{\mathrm{new}}$ values since the intrinsic signal width is lower.

Eventually, in the VBF channel, the electroweak singlet model implemented in Phantom is used to correctly generate the interference term for each $(m_{\mathrm{H}}, C'^2, \mathrm{BR}_{\mathrm{new}})$ point considered in the analysis, then interpolated to the complete grid. To correctly account for the unitarity constraint, the VBF SM Higgs off-shell contribution (S_h) and its interference with $qq \to qq\mathrm{WW}$ SM background (I_h) are included, where $\mathrm{S}_h + \mathrm{B} + \mathrm{I}_h$ is computed with Phantom while the off-shell S_h in MadGraph. In this BSM scenario, the interference weight function becomes:

$$\omega(m_{\mathrm{WW}}) = \frac{[(\mathrm{S_H} + \mathrm{I_H} + \mathrm{B})(m_{\mathrm{WW}}) - \mathrm{B}(m_{\mathrm{WW}})] + [(\mathrm{S}_h + \mathrm{I}_h + \mathrm{B})(m_{\mathrm{WW}}) - (\mathrm{S}_h + \mathrm{B})(m_{\mathrm{WW}})]}{\mathrm{S_H}(m_{\mathrm{WW}})}$$

where the SM Higgs off-shell contribution is small and subtracted from the numerator through the MadGraph prediction, while its interference with the SM $qq \to qq$WW background is absorbed in the signal re-weighting, neglecting any mutual interference between the two Higgs states.

5.3 Analysis strategy and event selection

The final state signature of the H $\to$ WW $\to \ell\nu q\bar{q}'$ decay is characterized by an isolated charged lepton, either a muon or an electron, missing transverse energy ($\not\!\!E_T$) and a jet containing the entire hadronic decay of a high p_T W-boson.
Events are categorized depending on the lepton flavour and the additional jet activity outside of the W-jet candidate cone, to enhance the analysis sensitivity over the explored mass range. The analysis is performed studying the distribution of the reconstructed invariant mass of the WW system, m_{WW}, to distinguish between a peaking shape for the expected Higgs signal, increasing in width as a function of the mass hypothesis, and a smooth non-resonant falling background.
A heavy SM-like Higgs boson may be produced through gluon-gluon fusion or VBF mechanisms, where the VBF production is expected to play a big role in the high mass region, reaching almost 50% of the total production cross section for $m_{\mathrm{H}} = 1$ TeV, as shown in Section 1.3. For this reason, the analysis is divided into bins of additional jet activity, to classify if the event is consistent with the ggH or qqH topology. This improves the sensitivity over an inclusive CMS analysis [165] and the VBF category shows a first benchmark of future di-boson scattering measurements in the considered final state.

5.3.1 W$\to \ell\nu$ reconstruction and identification

Muons are measured by means of both the tracker and the muon system, within $|\eta| < 2.4$, while electrons are detected as tracks pointing to energy clusters in ECAL, within $|\eta| < 2.5$, excluding the transition region between the barrel and the endcaps, $1.44 < |\eta| < 1.57$.
Muons (electrons) are required to have a momentum transverse to the beam direction greater than 30 (35) GeV. Lepton candidates are required to be isolated from other detector activities and compatible with the primary vertex of the event, which is chosen as the one with the highest $\sum p_T^2$ of its associated tracks. According to the simulation, this requirement provides the correct assignment for the primary vertex in more than 99% of cases in signal events.
Muons are identified through orthogonal selections with an average efficiency of 80% (tight muon identification) [76], while electrons according to a dedicated MVA selection [79], in contrast to the identification strategies described in Section 3.2.2 and adopted in the analysis previously exposed.
In fact, the tight muons identification differs from the high p_T one by requiring the normalized χ^2 of the global track to be less than ten and to fit at least five hits in the tracker layers. The electron MVA identification is based on a BDT taking as inputs both super-cluster and track information. The adopted working point guarantees an almost flat 80% efficiency on real prompt electrons over a large range of transverse momentum.
The lepton isolation is ensured by applying requirements on the sum of the transverse energies of all reconstructed PF candidates (charged particles, photons and neutral hadrons) within a cone of $\Delta R < 0.4$ (0.3) for μ (e) around the lepton direction, after subtracting the average

pileup energy estimated using a jet area technique on an event-by-event basis [85].
To reduce the background contamination from the DY lepton pairs and SM di-boson production, additional loosely identified leptons are vetoed in the event, with $p_T > 20$ GeV for electrons and $p_T > 10$ GeV for muons.
The leptonic W-boson candidate is obtained combining the lepton momentum with the missing transverse energy measured in the event, obtained from the PF reconstruction with Type-I correction, as described in Section 2.4.5. Finally, the unmeasurable longitudinal component of the neutrino momentum is reconstructed by requiring the lepton-neutrino pair to have the invariant mass of a W-boson, using a kinematic constraint as already done in Section 4.2.1. The ambiguity in the involved second-order equation is resolved by taking the solution that yields the smallest $|p_z|$ value which, when evaluated in simulation, is closest to the true $|p_z|$ approximately 75% of the times.

5.3.2 W→ $q\bar{q}'$ identification using jet substructure

Two different clustering algorithms, the anti-k_t and the Cambridge-Aachen (CA), are used to reconstruct jets in the event, considering the whole detector pseudorapidity coverage up to $|\eta| < 5$. Cambridge-Aachen jets are clustered with a distance parameter of $R = 0.8$ (CA8) and are used for reconstructing the hadronically decaying W-boson (W-jets), while anti-k_t jets, clustered with $R = 0.5$ (AK5), are used for categorizing events as a function of the additional jet activity, as well as for b-quark jet tagging.
Any reconstructed jet that overlaps with an identified lepton, within $\Delta R = 0.5$ (0.8) for AK5 (CA8) jets, is removed to avoid double counting of leptons as a jets. The hadronic W-boson candidate is therefore reconstructed as a single massive CA8 jet and substructure techniques are adopted to tag hadronically decaying boosted V-jets, adopting the same W-jet identification strategy optimized and described in Chapter 3.

5.3.3 Event selection

The large invariant mass of the heavy Higgs resonance translates into a large transverse momentum of the W-bosons produced in the decay. Therefore, the transverse momentum of both the reconstructed leptonically decaying W-boson and the hardest CA8 jet in the event are required to be larger than 200 GeV. At the same time, the missing transverse energy is required to be above 50 (70) GeV for the muon (electron) channel, to further suppress possible contribution to the analysis phase space from QCD multi-jet events.
In addition, other specific topological requirements are added to the event selection, requiring the two W-boson candidates to be back-to-back. In fact, the distance between the lepton and the W-jet should be $\Delta R(\ell, \mathrm{W_{had}}) > \pi/2$, the azimuthal distance between the missing energy and the W-jet should be $\Delta\varphi(\not{E}, \mathrm{W_{had}}) > 2.0$ and, finally, the azimuthal distance between the leptonically decaying W-boson and the W-jet should be $\Delta\varphi(\mathrm{W_\ell}, \mathrm{W_{had}}) > 2.0$.
Considering the reconstructed AK5 jets with $p_T > 30$ GeV and $\Delta R > 0.8$ from the leading CA8 jet, events are rejected if there is at least one b-tagged jet to reduce the amount of $t\bar{t}$ background, adopting a b-tagging working point that provides 70% efficiency on prompt b-jets. This b-veto is only possible inside the tracking-covered region of the detector, up to $|\eta| < 2.5$.
Each event is then categorized according to the additional number of AK5 jets with $p_T >$

30 GeV. If there are zero or one AK5 jets, the event is classified in the 0+1-jet channel, otherwise if there are two or more AK5 jets, it falls in the 2-jet category. This strategy based on first selecting the W-jet candidate as the leading CA8 jet, then look for additional no b-tagged jets to identify the VBF tag-jets, has been tested against alternative assignments, showing the best signal efficiency for the VBF production mode minimizing the gluon fusion contamination inside the 2-jet category.
To profit at best of the available statistics, the 0+1-jet sample is split between lepton flavors ($e\nu$+W-jet and $\mu\nu$+W-jet), while in the 2-jet bin case the two lepton flavors are lumped together.
Figure 5.4 shows a comparison between data and simulation for the number of AK5 jets in the event, used to split gluon fusion and VBF categories after the kinematic selections previously described. In the left plot, the boosted inclusive W+jets sample is adopted, generated via MadGraph interfaced with Pythia 6, which shows a poor agreement with data. In contrast, in Figure 5.4 (middle), a good agreement is visible in the $t\bar{t}$ enriched sample, obtained inverting the b-veto selection. Finally, when using the exclusive W+jets generation (W+2, W+3 and W+4 jets at matrix element), adopted for the W+jets simulation in the 2-jet bin category, data-to-MC agreement improves as shown in Figure 5.4 (right) in the high jet multiplicity bins.

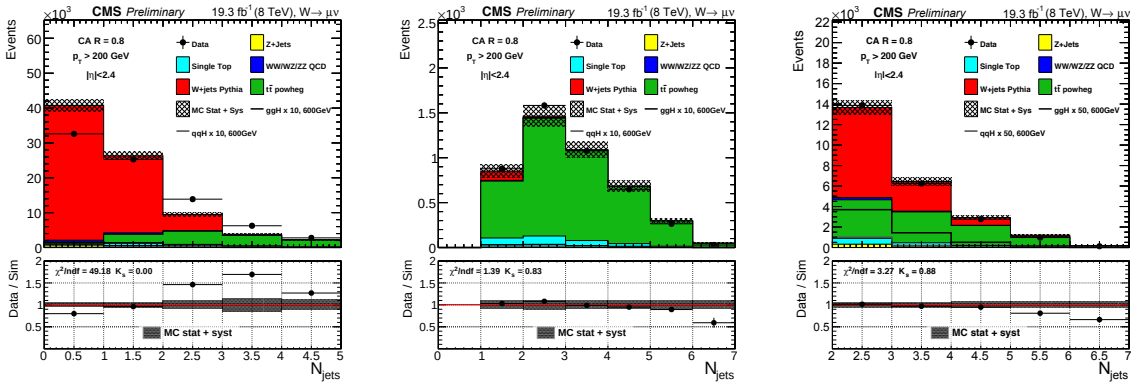

Figure 5.4: Comparison between data and simulation for the number of AK5 jets in the event, with a $p_T > 30$ GeV and reconstructed outside of the leading CA8 jet cone. (Left) The inclusive MadGraph + Pythia 6 W+jets sample is considered in the b-vetoed region, (middle) top enriched control region and (right) the 2-jet category when MadGraph + Pythia 6 W+jets exclusive generation is adopted.

5.3.4 Comparison of data and simulation in the 0+1-jet category

After applying the kinematic selections previously described, without jet substructure requirements except a loose selection on the W-candidate pruned jet mass to be in the range 40-130 GeV, data are compared with the expectation from simulation inside the 0+1-jet category. In Figure 5.5, the lepton p_T (bottom left), the transverse momentum of the leptonic W-boson $p_T^{W_\ell}$ (top left), the W-jet candidate p_T (top right) and its N-subjettiness τ_2/τ_1 (bottom right) are shown in the 0+1-jet bin for the $\mu\nu$+W-jet final state.
Since a sizable disagreement between data and Pythia 6 q/g jets simulation is observed for the pruned mass distribution, as discussed in Section 3.2.7, the W+jets background is normalized to yields in data. This disagreement in the 0+1-jet category is anyway irrelevant because ultimately the W+jets contribution is determined with a data-driven method.

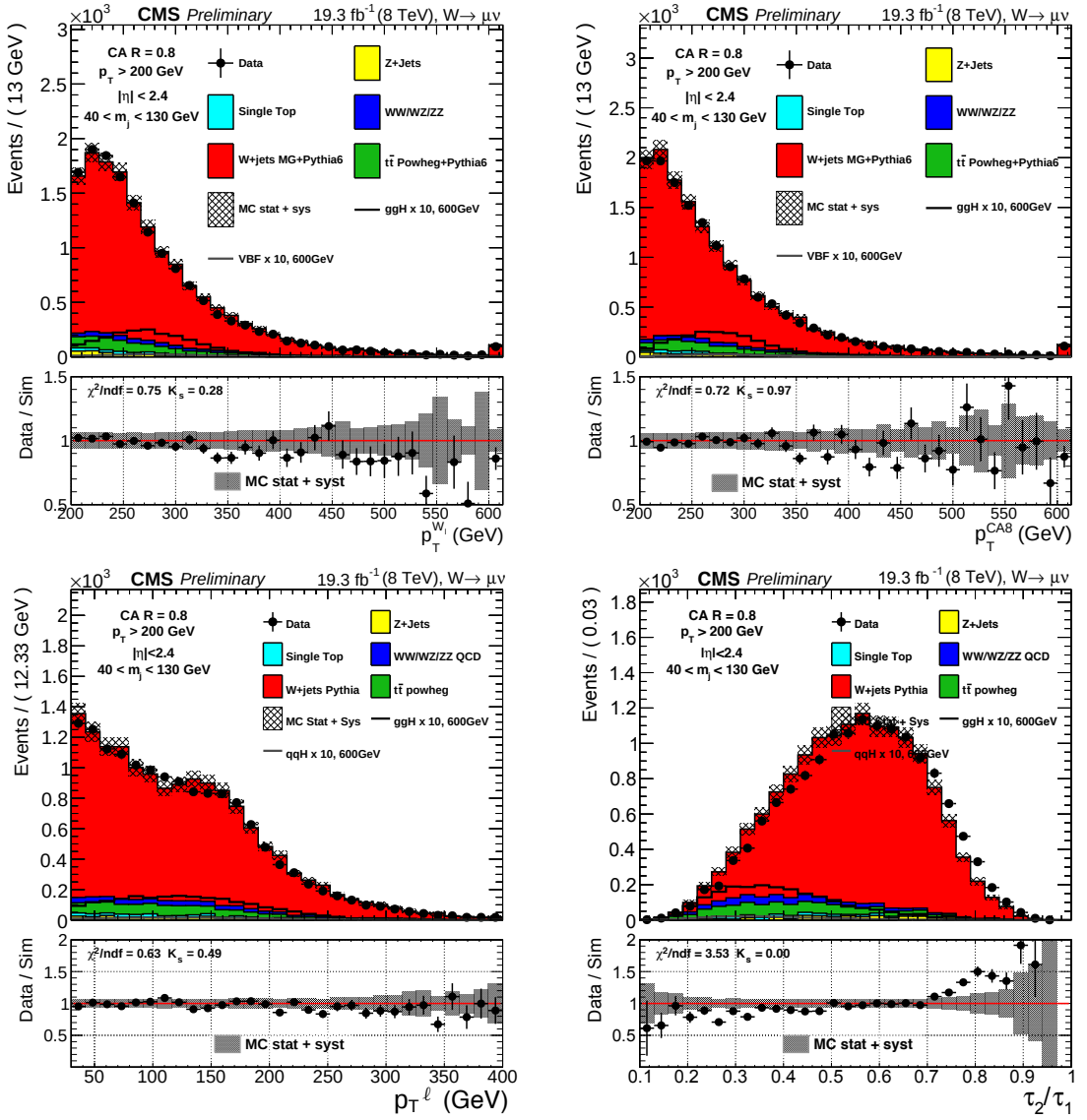

Figure 5.5: Comparison between data and simulation for the muon channel in the 0+1-jet category: leptonic W-boson p_T (top left), leading CA8 jet p_T (top right), muon p_T (bottom left) and N-subjettiness of the W-jet candidate (bottom right). Comparisons are done after a loose mass selection, requiring the leading CA8 pruned mass to range between 40-130 GeV.

5.3.5 Optimizing a VBF selection

In the 2-jet bin, after applying the basic kinematic selections described in the previous Section, a loose pruned mass requirement ($40 < m_{\text{pruned}} < 130$ GeV) and the N-subjettiness selection $\tau_2/\tau_1 < 0.5$, a good agreement between data and simulation is observed, where the largest background is represented by semi-leptonic $t\bar{t}$ production.
To reduce the $t\bar{t}$ component, a top quark mass veto is adopted in addition to the b-veto requirement. In fact, combining the leptonic and the hadronic W-candidate four momenta with the corresponding nearest AK5 jet, $t\bar{t}$ and single-top events show an invariant mass for the W-boson+jet system peaking around the top-quark one. These distributions are reported in

Figure 5.6, where the uncertainty band on the simulation prediction is obtained by combining the statistical uncertainty, due to the limited MC statistics, with a systematic one for the predicted cross section of each background component, estimated to be around 30% for W+jets and di-boson, 15% for single-top and 10% for $t\bar{t}$. Requiring both $m(\mathrm{W_{had}}, j)$ and $m(\mathrm{W}_\ell, j)$ to be larger than 210 GeV allows to reduce the top background contamination by a factor four, loosely affecting the VBF signal acceptance as reported in Table 5.1 for a single generated mass point.

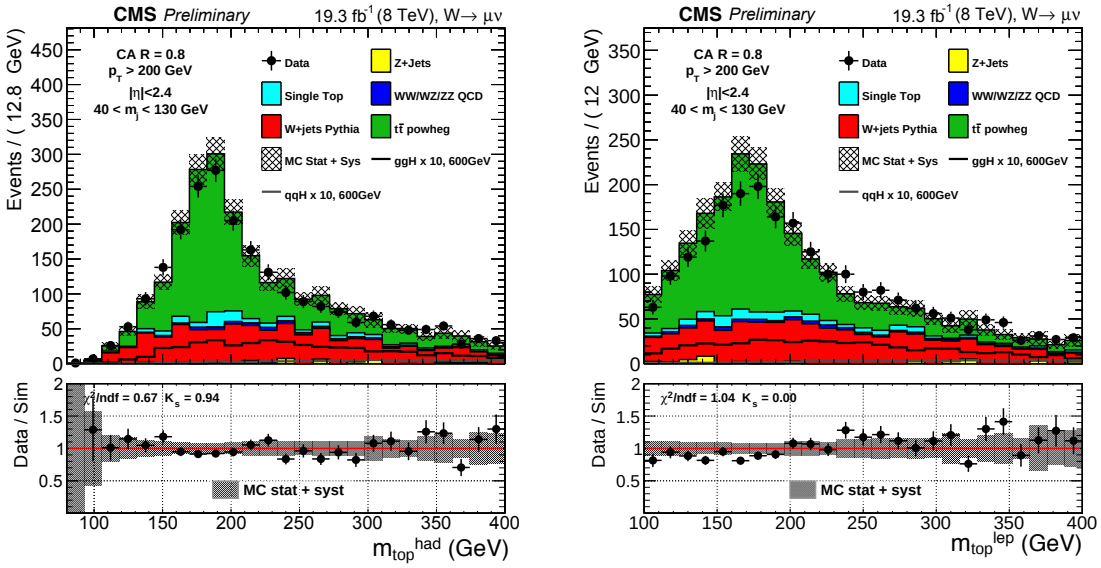

Figure 5.6: Invariant mass of the W-boson candidate and the nearest AK5 jet for the hadronic leg of the decay (left) and the leptonic one (right) in the 2-jet bin category, considering only events with one isolated muon in the final state.

Selection	$\varepsilon_{\mathrm{ggH}}$	$\varepsilon_{\mathrm{qqH}}$	$\varepsilon_{\mathrm{sing\text{-}t}}$	$\varepsilon_{\mathrm{VV}}$	$\varepsilon_{\mathrm{W+jet}}$	$\varepsilon_{t\bar{t}}$	$\mathrm{S_{ggH+qqH}/B}$
b-veto	0.91	0.94	0.44	0.89	0.82	0.40	1.8×10^{-2}
$m(\mathrm{W_{had}}, j)$	0.82	0.96	0.60	0.78	0.69	0.44	2.8×10^{-2}
$m(\mathrm{W}_\ell, j)$	0.72	0.96	0.45	0.68	0.64	0.37	4.0×10^{-2}

Table 5.1: Top killing selection efficiencies for a SM-like Higgs signal with $m_{\mathrm{H}} = 800$ GeV and each background source. Purity, defined as S/B, is reported after each selection, which are applied sequentially. The b-veto efficiency refers to all the selected events in the 2-jet bin category, after applying acceptance requirements and W-jet identification selections, while the top mass selections are applied In the purity evaluation, the m_{WW} of the final decay system is required to be between 550-1500 GeV.

Finally, additional selections are applied to the exclusive 2-jet bin category to further isolate the VBF signal topology, which is characterized by two hard tag-jets with large invariant mass (m_{jj}) and η separation ($\Delta\eta_{jj} = |\eta_{j_1} - \eta_{j_2}|$).
Alternative VBF jet selections are tested by taking the two jets with largest invariant mass or η separation, which show worse performances, in terms of both matching efficiency with generator level partons and signal over background ratio, with respect to the largest p_T criterion. After rejecting a good fraction $t\bar{t}$ events through the top mass veto requirement, $\mu\nu$+W-jet and

$e\nu$+W-jet events are merged together and a simultaneous rectangular selection on m_{jj}-$\Delta\eta_{jj}$ is optimized. A working point that allows 80% qqH signal efficiency is chosen, which corresponds to require $m_{jj} > 250$ GeV and $\Delta\eta_{jj} > 3$.
In Figure 5.7, the $\Delta\eta_{jj}$ (left) and m_{jj} (right) distributions are shown in the 2-jet bin category before any VBF-like requirement, considering only events with an isolated muon in final state. In contrast with the 0+1-jet case, where the background is predominantly due to the W+jets production, in the 2-jet channel there is a more even fraction of W+jets and $t\bar{t}$ backgrounds, although the $t\bar{t}$ is further reduced when both VBF and top mass requirements are applied.
The analysis signal region is defined, in both 0+1-jet and 2-jet categories, requiring the pruned jet mass to range between 65-105 GeV, while the two regions $40 < m_{\text{pruned}} < 65$ GeV and $105 < m_{\text{pruned}} < 130$ GeV are used as sidebands, as already done in other high mass searches described in the Chapter 4. A summary of the final kinematic selection used in the 0+1-jet and 2-jet categories are listed in Table 5.2.

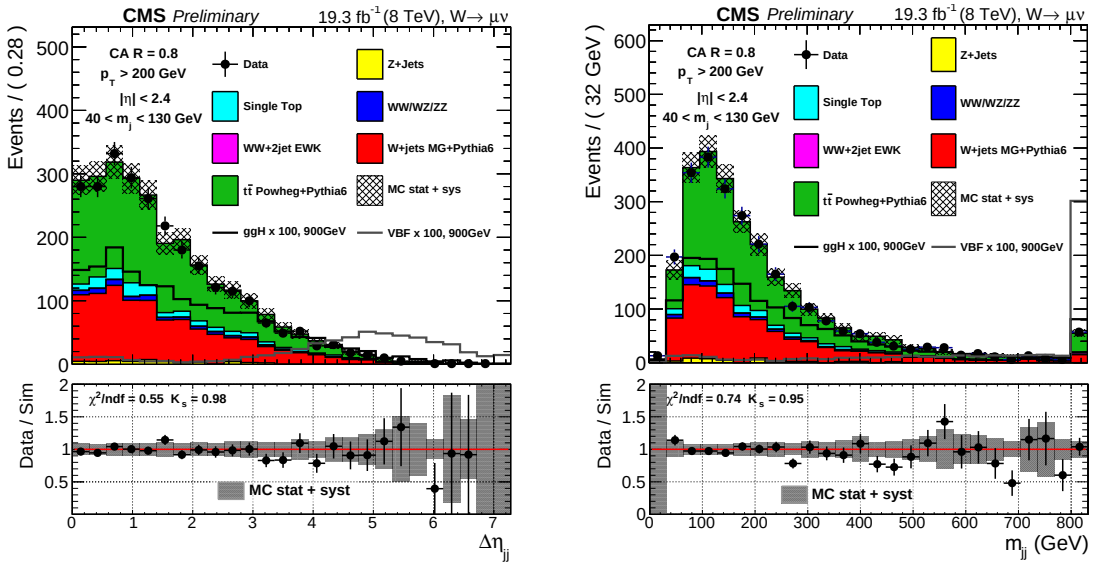

Figure 5.7: Absolute pseudorapidity difference $\Delta\eta_{jj}$ (left) and invariant mass m_{jj} (right) of the selected VBF tag-jet system in the 2-jet bin category, for events with an isolated muon in the final state, after the kinematic selection requirements previously described.

5.4 Signal modeling

The final analysis strategy relies on an unbinned fit to the reconstructed di-boson invariant mass spectrum (m_{WW}), where both the discovery potential and the exclusion power depend on an accurate description of the signal shape. The reconstructed gluon fusion lineshape, obtained in both the 0+1-jet and in the 2-jet category, after applying the event selection, is well modelled by a one-sided Crystal-Ball function, since interference effects are covered by the experimental resolution even for the highest mass points.
Figure 5.8 shows the Crystal-Ball fits to the reconstructed ggH m_{WW} spectrum predicted for a SM-like Higgs with $m_{\text{H}} = 800$ GeV in both 0+1-jet bin (left) and 2-jet category (right).
In contrast, for the VBF case, a composite model given by the product of a Crystal-Ball peak and a falling exponential function is more suitable especially at high mass, where interference

Selection	Value	Comments
`Tight` Lepton selection		
Electron p_T and η	$p_T > 35$ GeV, $\lvert\eta\rvert_{\text{SC}} < 2.5$ except [1.44, 1.56]	avoid the `ECAL` gap.
Muon p_T and η	$p_T > 30$ GeV, $\lvert\eta\rvert < 2.4$	
`Loose` Lepton selection		
Muon (Electron) p_T	$p_T > 10\,(20)$ GeV	
`CA8` jet selections		
Jet p_T and η	$p_T > 80$ GeV, $\lvert\eta\rvert < 2.4$	hadronic W reconstruction
`AK5` jet selections		
Jet p_T and η	$p_T > 30$ GeV and $\lvert\eta\rvert < 2.4$	used for b-tag and `VBF` tag-jet
$\not\!\!E_T$ selections		
$\not\!\!E_T$ μ (e) channel	$\not\!\!E_T > 70\,(50)$ GeV	
V-Boson selections		
Leptonic W p_T	$p_T > 200$ GeV	
Hadronic W p_T	$p_T > 200$ GeV	
Back-to-back topology	$\Delta R(\ell, \text{W}_{\text{had}}) > \pi/2$, $\Delta\phi(\text{W}_{\text{had}}, \not\!\!E_T) > 2$	
	$\Delta\phi(\text{W}_{\text{had}}, \text{W}_{\text{lep}}) > 2$	
Veto		
Number of `loose` leptons	0	besides the `tight` lepton
Number of b-tag jets	0	`CSV` medium w.p.
W-jet selections		
m_{pruned}	$65 < m_{\text{pruned}} < 105$ GeV	
N-subjettiness ratio	$\tau_2/\tau_1 < 0.5$	
Number of jets		
Number of jets	0+1-jet (≤ 1), 2-jet (≥ 2)	event categorization
Top mass veto		
$m(\text{W}_{\text{had}},\ j)$	$m(\text{W}_{\text{had}},\ j) > 200$ GeV	2-jet category only
$m(\text{W}_{\ell},\ j)$	$m(\text{W}_{\ell},\ j) > 200$ GeV	2-jet category only
`VBF` jet selections		
`VBF` tag-jets	at least two jets with $p_T > 30$ GeV	two highest p_T jets
`VBF` jet invariant mass	$m_{jj} > 250$ GeV	2-jet category only
`VBF` jet eta separation	$\Delta\eta_{jj} > 3$ GeV	2-jet category only

Table 5.2: Kinematic selections used in the 0+1-jet and 2-jet analyses for the final event categorization.

effects are stronger. Figure 5.9 reports the fits to the qqH m_{WW} distribution for a `SM`-like Higgs with $m_{\text{H}} = 800$ GeV in both 0+1-jet (left) and 2-jet channel (right).
The signal shape and normalization for the `SM`-like case are estimated from the `POWHEG` simulation, re-weighted to account for interference effects, as described in Section 5.2.1, while `BSM` scenarios are produced after the additional width re-weighting described in Section 5.2.2. Furthermore, data-to-simulation corrections are applied to account for discrepancies in the pruned mass scale, pruned mass resolution and W-jet tagging efficiency, evaluated in the $t\bar{t}$ control sample, as well as for lepton identification, b-tagging and trigger efficiency.

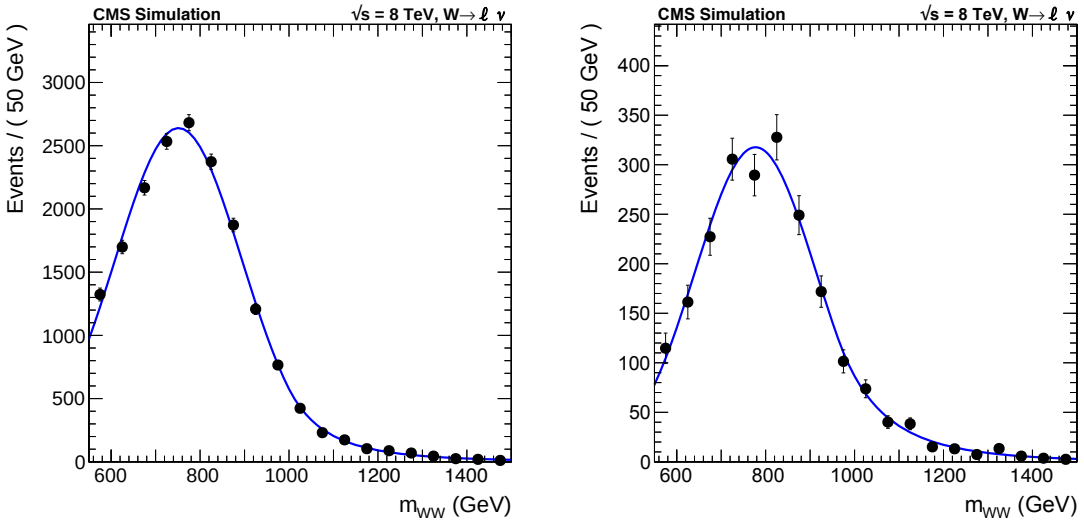

Figure 5.8: Crystal-Ball fit to the reconstructed lineshape of gluon fusion SM-like Higgs events, generated fixing $m_{\mathrm{H}} = 800$ GeV, after applying 0+1-jet category selections (left) and 2-jet bin ones (right). Events are re-weighted applying all the considered data-to-MC corrections, without normalizing them to the luminosity of data sample.

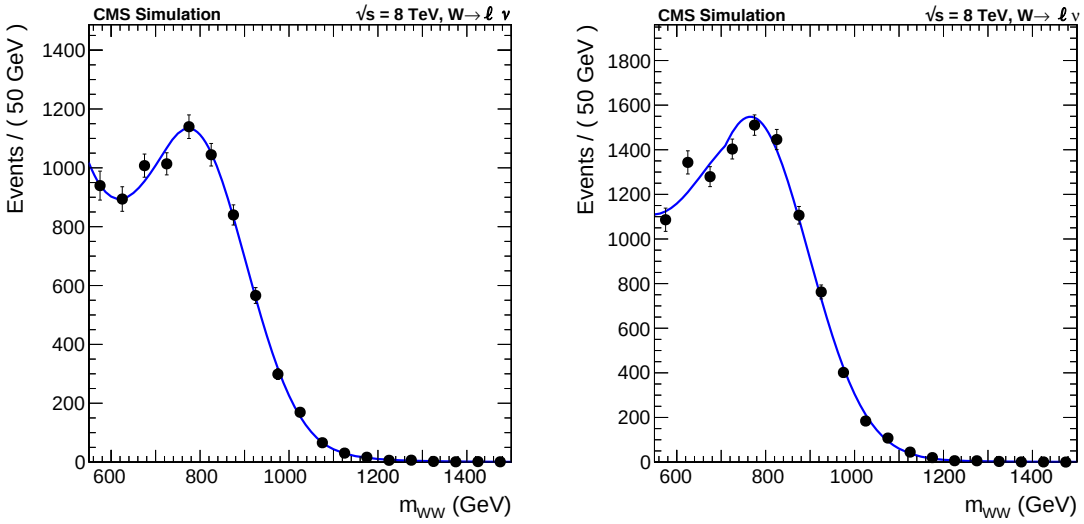

Figure 5.9: Fit to the reconstructed lineshape of VBF SM-like Higgs events, generated fixing $m_{\mathrm{H}} = 800$ GeV, after applying 0+1-jet category selections (left) and 2-jet bin ones (right) using a product between an exponential function and a Crystal-Ball shape. Events are re-weighted applying all the considered data-to-MC corrections, without normalizing them to the luminosity of data sample.

5.5 Background estimation

An important difference between this search and the one described in Chapter 4 stand in the online trigger strategy, which brings to lower p_T thresholds on leptons, and in different muon and electron identification criteria. This allows to move the kinematic acceptance turn-on to lower m_{WW} values, resulting in a smooth falling background shape starting from $m_{\mathrm{WW}} > 500 - 550$ GeV in all the analysis categories. This choice is motivated by the need of being

sensitive starting from a resonance with mass around 600 GeV, where boosted searches usually start to dominate.
Since jet substructure observables reproduced in simulated events crucially depend on the parton shower and underlying event model, a sample of hadronically decaying W-bosons is isolated in data to study the effect of jet selections. By applying the analysis selection strategy, but requiring at least one b-tagged jet in the event, a sample of boosted W-jets is isolated in a region of nearly pure $t\bar{t}$ events. The $t\bar{t}$ selection and the extraction of dedicated scale factors for pure W-jet sare performed following the same strategy described in the previous analyses, in particular in both Section 3.2.2 and Section 4.4.
In the 0+1-jet category, the $t\bar{t}$ normalization correction is measured to be 0.91 ± 0.08 (0.89 ± 0.09) in the muon (electron) channel, while, in the 2-jet one, the scale factor is 1.09 ± 0.25. The larger uncertainty in the 2-jet category comes from a reduced amount of statistics after applying `VBF` requirements.
Furthermore, looking at only events with $65 < m_{\text{pruned}} < 105$ GeV, data are compared to simulation considering two alternative $t\bar{t}$ generations, performed with `POWHEG` interfaced to `Pythia6` and `MC@NLO` interfaced to `Herwig++`, as a function of successive `VBF` selections:

- At least two additional `AK5` jets (`VBF` jets) outside the W-jet candidate cone, with one of them b-tagged $\rightarrow$ (**a**).
- (**a**) + top mass veto [$m(\text{W}_{\text{had}}, j)$ and $m(\text{W}_{\ell}, j) > 210$ GeV] $\rightarrow$ (**b**).
- (**a**) + `VBF` selections [$m_{jj} > 250$ GeV and $\Delta\eta_{jj} > 3$] $\rightarrow$ (**c**).
- (**a**) + top mass veto + `VBF` selections $\rightarrow$ (**d**).

Inside these regions, data-to-simulation normalization corrections are measured to be consistent, within the statistical uncertainty, with the scale factor derived in the 0+1-jet category and between alternative matrix-element or parton-shower simulations.
Once single-top and $t\bar{t}$ are corrected for to match yields in data, the reconstructed m_{WW} obtained from both `POWHEG`+`Pythia6` and `MC@NLO`+`Herwig++` simulations are compared to what observed in data, as shown in Figure 5.10. No significant discrepancies between the two generations are found, despite different matrix element and parton shower choices. Furthermore, as a function of the different event selections, the overall data-to-`MC` agreement is not significantly different and pretty much consistent within uncertainties. Thus, it is reasonable to assume that $t\bar{t}$ simulation is able to well model the m_{WW} distribution observed in data, once the normalization corrections are applied.
Finally, the simultaneous fit to the pruned mass distribution, used to extract the W-tagging scale factor as described in Section 3.2.7, is performed again to be consistent with the object selections applied in this analysis. It is measured to be 0.93 ± 0.09, while the jet mass scale needs to be shifted by +1.4 GeV and the resolution enlarged by 10%. These correction factors are also applied on other simulation based samples, like ggH and qqH productions and di-boson backgrounds (WW/WZ/ZZ).

5.5.1 W+jets background estimation

The W+jets contribution, which represents the leading background in the 0+1-jet channel and around 50% of the total background in the `VBF` category, is extracted from data in

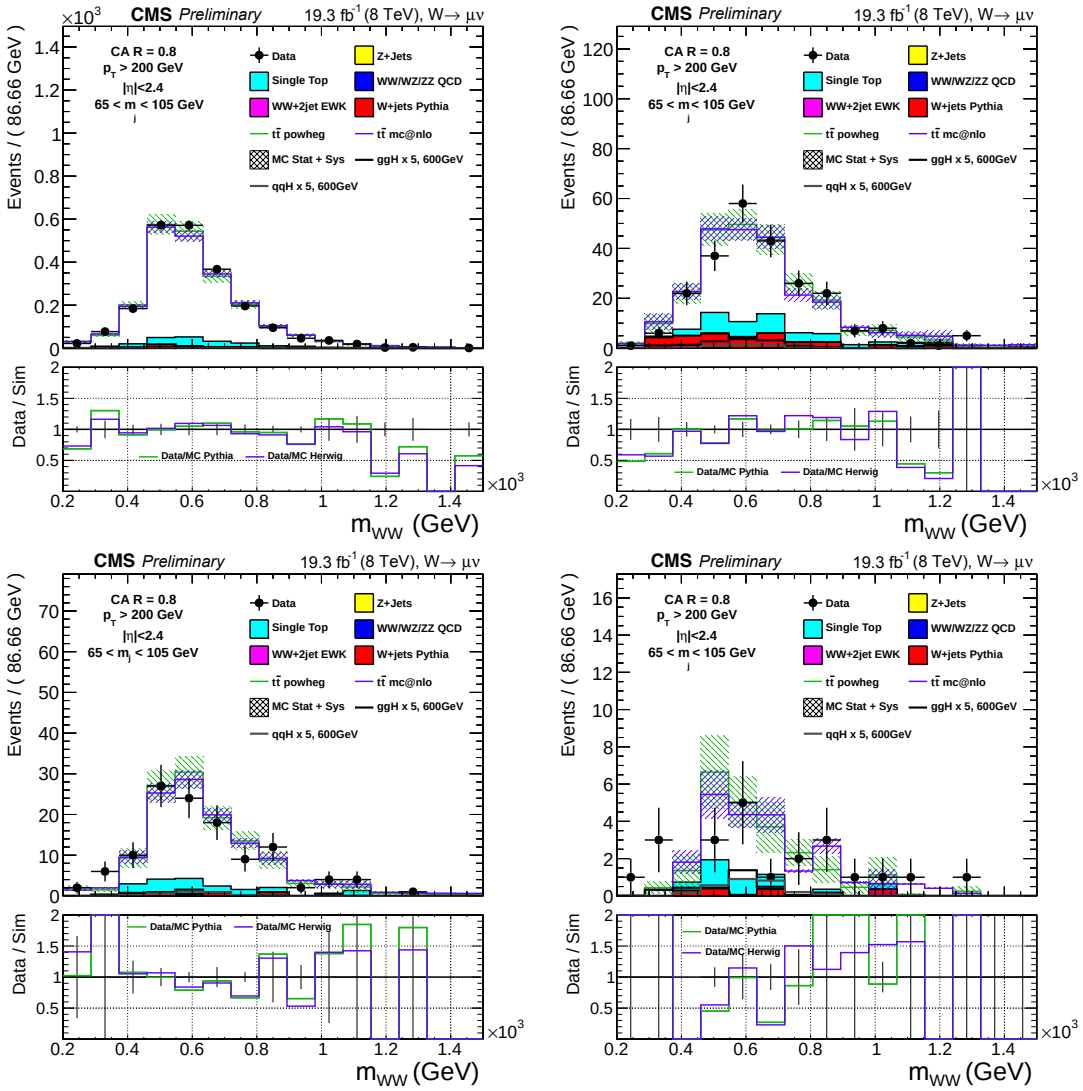

Figure 5.10: Di-boson invariant mass distributions (m_{WW}) for events with $65 < m_{\mathrm{pruned}} < 105$ GeV in the top enriched control region, evaluated after a different set of selections: basic $t\bar{t}$ sample definition (top left) (**a**), basic $t\bar{t}$ + top mass veto (top right) (**b**), basic $t\bar{t}$ + VBF selections (bottom left) (**c**) and basic $t\bar{t}$ + top mass veto + VBF selections (bottom right) (**d**).

both shape and normalization via a sideband method. The signal region (SR) is defined around the hadronic W-boson mass peak, requiring the pruned jet mass to fall in the range $65 < m_{\mathrm{pruned}} < 105$ GeV, while an event belongs to the lower sideband (LSB) when W-jet pruned mass lies in $40 < m_{\mathrm{pruned}} < 65$ GeV, or the upper one when $105 < m_{\mathrm{pruned}} < 130$ GeV. The same strategy is applied in both the 0+1-jet and the 2-jet category, with the only difference of combining electron and muon final states in the 2-jet case, to minimize the statistical uncertainty related to sideband fit extrapolation improving the analysis performance. Furthermore, both the W+jets sideband definition and the background estimation strategy are equivalent to the ones introduced in Section 4.4.

The W+jets rate is obtained from a sideband fit of data to the m_{pruned} spectrum, where data inside the signal region are not considered by the fit. The fit is unbinned and the functional

parametrization of each background component, namely di-boson, single top, $t\bar{t}$ and W+jets, are chosen via pre-fits on simulated events. Single top, di-boson and $t\bar{t}$ contributions are taken from simulation after applying the data-to-simulation corrections already described.
The set of empirical functions used to model each contribution are summarized in Table 5.3, where $\mathrm{F_{Pow2}}$ is defined to be:

$$\mathrm{F_{Pow2}} = \frac{[1-(m_{\mathrm{pruned}}/a)]^{p_0}}{(m_{\mathrm{pruned}}/a)^{p_1}}$$

Figure 5.11 and Figure 5.12 show, respectively, fits to the m_{pruned} distribution for simulated events of $t\bar{t}$, di-boson and W+jets processes in the 0+1-jet muon category and in the 2-jet one, where the uncertainty band is related to uncertainty on the fitted parameters describing both the shape and the normalization of each background.

Channel	m_{WW} (TeV)	W+jets	$t\bar{t}$	Single-t	Di-boson
0+1-jet μ+jet	[0.55-1.5]	$\mathrm{F_{ErfExp}}$,$\mathrm{F_{Pow2}}$	$\mathrm{F_{ErfExp2Gaus}}$	$\mathrm{F_{ErfExpGaus}}$	$\mathrm{F_{2Gaus}}$
0+1-jet e+jet	[0.55-1.5]	$\mathrm{F_{ErfExp}}$,$\mathrm{F_{Pow2}}$	$\mathrm{F_{ErfExp2Gaus}}$	$\mathrm{F_{ErfExpGaus}}$	$\mathrm{F_{2Gaus}}$
2-jet	[0.55-1.5]	$\mathrm{F_{ErfExp}}$,$\mathrm{F_{Pow2}}$	$\mathrm{F_{ErfExp2Gaus}}$	$\mathrm{F_{ErfExp}}$	$\mathrm{F_{2Gaus}}$

Table 5.3: Summary of the shapes used to fit the m_{pruned} spectrum of each background component, most of them already defined in Section 4.4, taking into account both the m_{WW} range and the final state topology (0+1-jet or 2-jet bin).

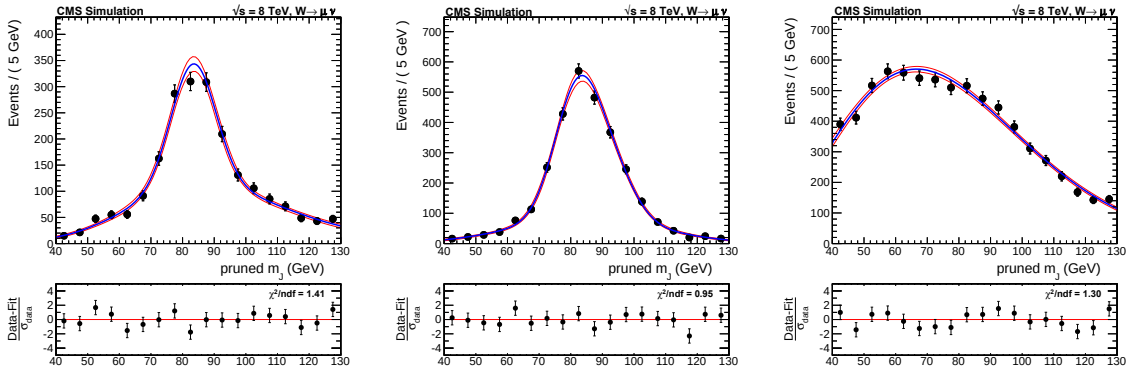

Figure 5.11: Fits to the pruned jet mass spectrum of simulated events with a reconstructed di-boson invariant mass (m_{WW}) in the range 0.55-1.5 TeV for the 0+1-jet $\mu\nu$+W-jet category. From left to right: $t\bar{t}$, di-boson (WW/WZ/ZZ) and W+jets ($\mathrm{F_{ErfExp}}$) processes, which are re-weighted applying all the considered data-to-MC corrections, without normalizing them to the luminosity of data sample.

Figure 5.13 shows the final result of the m_{pruned} sideband fit for the 0+1-jet muon category (left) and the 2-jet one (right). Two possible parametrizations, $\mathrm{F_{ErfExp}}$ and $\mathrm{F_{Pow2}}$, are chosen as possible candidates to model the W+jet m_{pruned} distribution, showing similar predictions inside the signal region on both data and W+jets simulation.
Possible biases in the W+jets normalization extraction, due to a mismodel of the pruned mass shape, are assessed by generating toys under one model and applying the full fitting procedure with the alternative one, using post-fit parameters from data to seed the toy generation. The best choice is represented by the parametrization showing the least bias and the uncertainty on the W+jets contribution is enhanced according to the difference between the injected and the fitted W+jets rate.

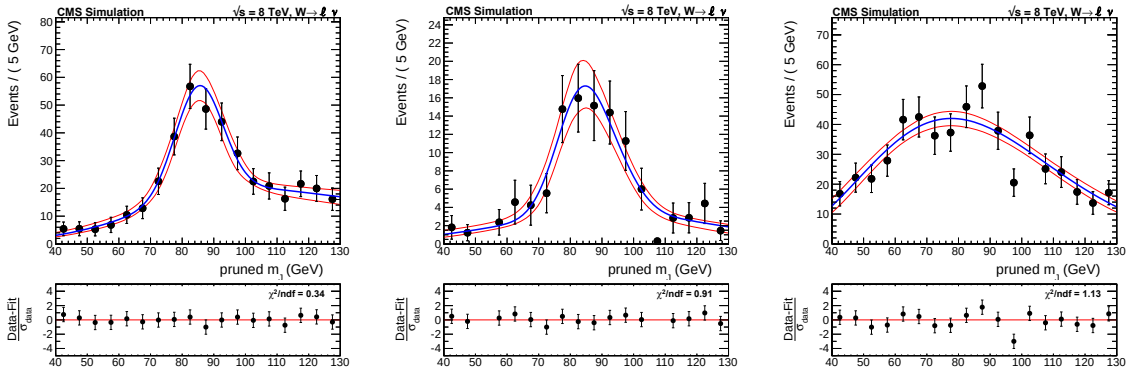

Figure 5.12: Fits to the pruned jet mass spectrum of simulated events with a reconstructed di-boson invariant mass m_{WW} in the range between 0.55-1.5 TeV for the 2-jet category. From left to right: $t\bar{t}$, di-boson (WW/WZ/ZZ) and W+jets ($\mathrm{F_{ErfExp}}$) processes, which are re-weighted applying all the considered data-to-MC corrections, without normalizing them to the luminosity of data sample.

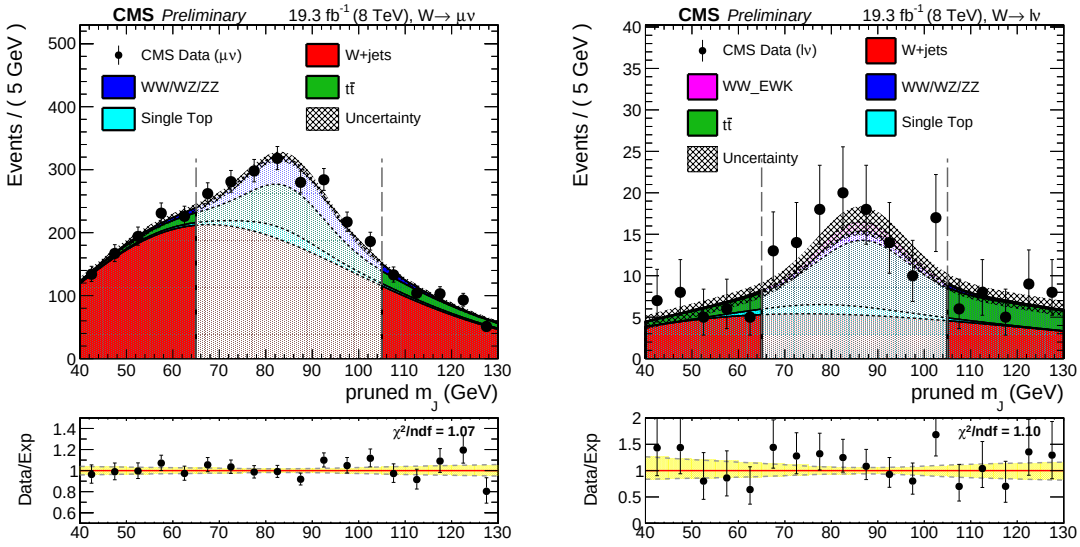

Figure 5.13: Distributions of the pruned jet mass (m_{pruned}) in the 0+1-jet muon category (left) and in the 2-jet one (right). The prediction of the non-resonant W+jets background comes from a fit excluding the signal region (marked through vertical dashed lines), while the minor background yields come from the simulation. The simulated resonant shapes are corrected using the differences between data and simulation in the W-boson mass peak position and width, which are measured in the $t\bar{t}$ control region. At the bottom of each plot, the ratio between event the observed rate in data and the predicted background is reported.

Thus, the background normalization uncertainty is increased by 15% in the 2-jet bin and by 9% in the 0+1-jet one.

The W+jets shape in the signal region is determined from data in the lower sideband since in the upper one, with $105 < m_{\mathrm{pruned}} < 130$ GeV, heavy resonances decaying into $\mathrm{X} \to h\mathrm{W}$ system can contribute biasing the shape extrapolation from data. In addition, the upper sideband suffers from a higher contamination from $t\bar{t}$ events, in particular for the 2-jet bin as displayed in Figure 5.13.

In both the low mass sideband and the signal region, background events fulfilling the selection criteria are characterized by a falling m_{WW} distribution over the range between 550-1500 GeV.

This falling spectrum can be parametrized by different class of functions, such as exponentials, power laws or polynomials. In each of these class of models, a representative candidate is chosen through an F-test, as described in Section 4.4. This test is performed using data in the W+jets sideband region, since it represents a pure sample of W+jets events in data as well as a good proxy for the final analysis signal region. According to a reference confidence level 4.4.1, the best models in the 0+1-jet category are represented by an exponential function and a power law with two parameters, while in the 2-jet one a single parameter power law is enough ($x = m_{\mathrm{WW}}$):

$$\mathrm{F_{Exp}}(x) = \mathrm{N} \cdot e^{c_0 \cdot x} \;;\; \mathrm{F_{Pow2}}(x) = \mathrm{N} \cdot x^{c_0 + c_1 \cdot \ln x} \;;\; \mathrm{F_{Pow}}(x) = \mathrm{N} \cdot x^{c_0} \tag{5.8}$$

In the W+jets sideband region, the non-dominant backgrounds contaminations are estimated fixing by their normalizations and shapes from simulation when fitting data. The fits to sideband data ($F^{\mathrm{W+jets}}_{\mathrm{Data,SB}}(m_{\mathrm{WW}})$), performed adopting an exponential function, are shown in Figure 5.14 for the 0+1-jet bin $\mu\nu$+W-jet channel (left) and the 2-jet category (right), where the uncertainty band reflects the fit parameter errors.

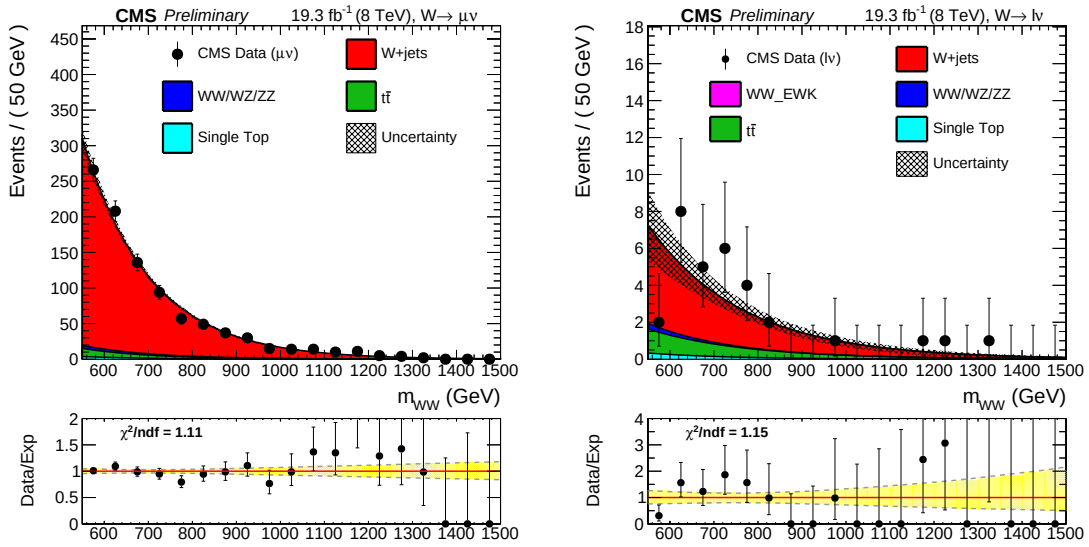

Figure 5.14: The fits for $F^{\mathrm{W+jets}}_{\mathrm{Data,SB}}(m_{\mathrm{WW}})$ for 0+1-jet $\mu\nu$+W-jet (left) and 2-jet (right) categories. Minor backgrounds ($t\bar{t}$, di-boson and single top) are estimated from simulation, while the W+jets shape is obtained from the fit to the data.

Finally, the W+jets shape in the signal region is obtained scaling $F^{\mathrm{W+jets}}_{\mathrm{Data,SB}}(m_{\mathrm{WW}})$ through the extrapolation function $\alpha_{\mathrm{MC}}(m_{\mathrm{WW}})$, derived from W+jets simulated events as defined in Equation 4.2. In this case, since no `Herwig++` W+jets exclusive samples are available, `MadGraph` + `Pythia 6` prediction is used as benchmark in the 2-jet category, while `Herwig++` is adopted in the 0+1-jet bin since it is known to better describe the pruned jet mass. The shapes of $\alpha_{\mathrm{MC}}(m_{\mathrm{WW}})$ are reported in Figure 5.15 for the 0+1-jet category (left) and the 2-jet one (right). The shaded regions correspond to the $1\,\sigma$ ($2\,\sigma$) bands of the fit parameters. For the 0+1-jet bin, the difference between the `Herwig` and `Pythia` determinations is taken as an additional systematic uncertainty, labelled as "alternate `PS`" in Figure 5.15. The systematic uncertainty associated to the chosen fit function is also considered and shown as "alternate function" in Figure 5.15, together with systematic effects due to uncertainties in the jet energy scale and resolution measurements.

The ultimate total background prediction in the signal region is obtained through Equation 4.3 and presented in Section 5.7, together with the statistical interpretation of the analysis.

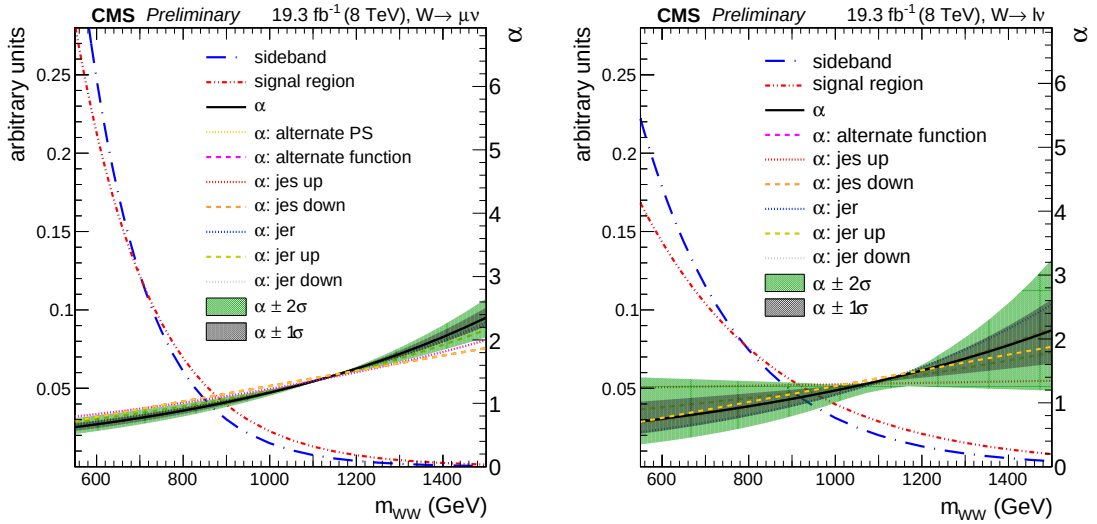

Figure 5.15: The W+jets extrapolation function $\alpha_{\mathrm{MC}}(m_{\mathrm{WW}})$ from the pruned mass sideband to the signal region is shown for the 0+1-jet (2-jet) category on the left (right). The shaded grey (green) band represents the $1\,\sigma$ ($2\,\sigma$) effect of the statistical uncertainty of fit parameters, while dashed colored lines refer to systematics related to parton shower, fitting function model, jet energy scale and resolution uncertainties. The original sideband and signal region fits for m_{WW} are also shown with dash-dot lines.

5.5.2 Bias tests for the m_{WW} background model

The described background estimation is based on the assumption of properly model each background contribution through an empirical continous shape. Since alternative parametrizations show potentially similar behaviors, a set of toy experiments is performed to select the best background model for the signal region, generating pseudodata to build a data sample corresponding to the expected statistics in the analysis.

As a first test, considering as benchmark background models an exponential function or a single parameter power law, background-only (`B`) and signal plus background (`S+B`) toys are generated and fitted adopting a chosen background model, fixing the nuisance parameters to their pre-fit value in the generation. Considering the fitted yields of each expected process, the pull distribution, defined as the difference between the fitted and the injected rate divided by the mean uncertainty of the fitted rate, is considered as figure of merit to estimate possible biases. The test is performed independently in each event category and for each considered m_{H} hypothesis, when a non null signal strength is injected. This self-bias is found to be negligible, less than 5% for each m_{H} hypothesis when `S+B` toys are performed, proving that the analysis is able to correctly detect a possible signal when the injection strength is comparable to the analysis sensitivity.

Then, the same test is performed to check the robustness of the background model, generating toys under one parametrization and fitting the obtained set of events with the alternative one (exponential or power law). The interesting parameter in these tests is represented by the fitted signal strength, where the considered signal model is represented by `SM`-like Higgs of various masses, in both toys with and without signal injection.

The measured bias is always below 12% for each mass point in both the 0+1-jet and the 2-jet category. As example, in Figure 5.16, the bias on the fitted signal strength (S+B fits) is shown for the 0+1-jet category (left) and the 2-jet one (right) as a function of the SM-like Higgs mass for background-only toys. In contrast, Figure 5.17 shows the same results in case of a sizable signal injection, three-times grater than the analysis exclusion sensitivity.

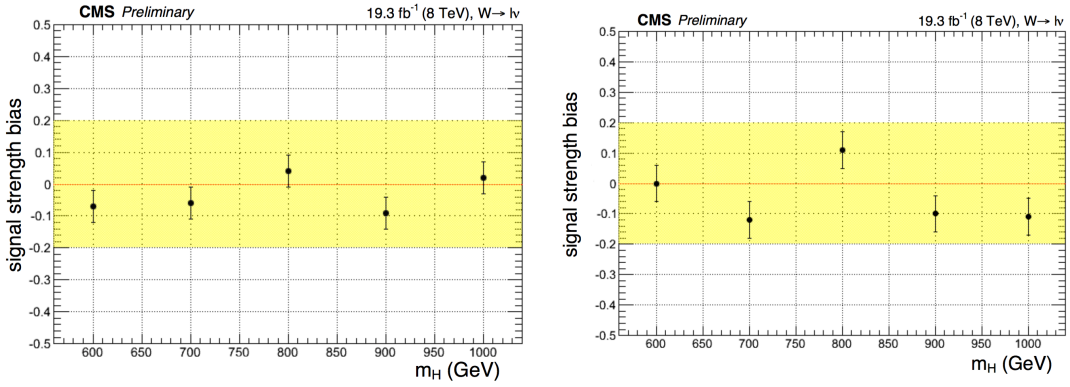

Figure 5.16: Bias estimation, as a function of the SM-like Higgs mass hypothesis, for events generated with a power law background model inside the signal region and fitted with an exponential function: 0+1-jet category (left), 2-jet one (right). S+B fits are performed on B-only toys. The bias is defined as the shift of the signal strength pull distribution.

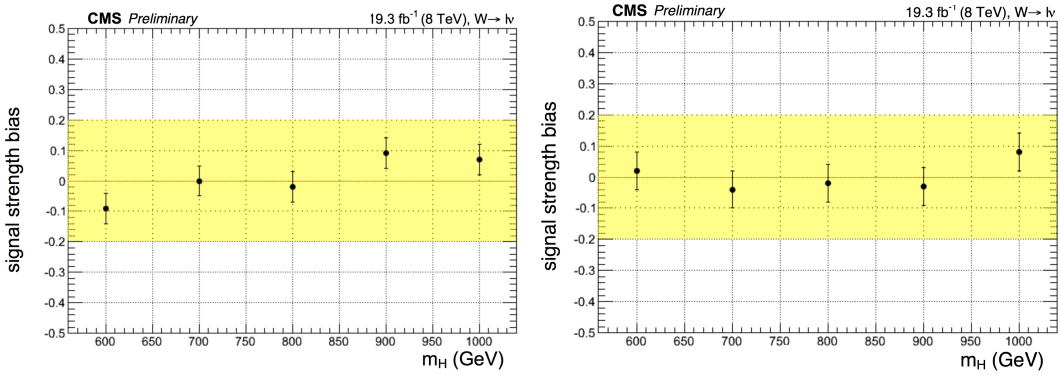

Figure 5.17: Bias estimation, as a function of the SM-like Higgs mass hypothesis, for events generated with a power law background model inside the signal region and fitted with an exponential function: 0+1-jet category (left), 2-jet one (right). S+B fits are performed on S+B toys, injecting three-times the signal rate expected to be excluded by the analysis. The bias is defined as the shift of the signal strength pull distribution.

5.6 Systematic uncertainties

In this section, systematic uncertainties affecting both the background prediction, in Section 5.6.1, and the signal one, in Section 5.6.2, are investigated and described.

5.6.1 Systematics on the background estimation

The systematic uncertainty on the W+jets background normalization is dominated by the pruned mass fit uncertainty, associated with the amount of data inside the sideband regions. A residual biase due to the adopted shape parametrization is also taken into account and

added in quadrature to the statistical term. This combined uncertainty is measured to be 5 (8)% in the 0+1-jet category for muon (electron) channel, while at the level of 22% in the VBF one.
The uncertainty on the $t\bar{t}$ normalization comes from the one on the data-to-simulation scale factors derived in the top-enriched control sample, which is estimated to be 6.5% in the 0+1-jet bin and 26.5% in the 2-jet one.
Then, uncertainties on the single lepton trigger and lepton identification efficiencies, derived through dedicated tag and probe studies on $Z \to \mu\mu$ and $Z \to ee$ events, are equal to 1% (1%) and 2% (2%) for μ (e) final states. The uncertainties coming from the lepton energy scale and resolution are found to be small, less than 1.5%, for all the background samples, as well as the uncertainty related to data-to-MC scale factors used for the b-tagging ($< 3.5\%$).
Systematics due to jet energy scale and resolution are summarized in Table 5.4 and are at the level of 1-7%, depending on the processes and the jet bin category (0+1-jet or 2-jet bin).
The uncertainty in the WW inclusive cross section in this boosted regime is assigned to be 20%, estimated from the uncertainty on the published CMS cross section measurement at $\sqrt{s} = 8$ TeV [5]. In addition, the uncertainty on the V-tagging scale factor, described in Section 5.5, is considered also for the WW background. The same uncertainties derived for WW production are also used for WZ, ZZ and WW electroweak processes ($O(\alpha_{ew}^6)$). The uncertainty related to the total integrated luminosity is 2.6% [108].
The systematic uncertainties on the W+jets m_{WW} shape in the signal region come from two independent contributions: the uncertainty on the extrapolation function $\alpha_{MC}(m_{WW})$ and the one on the fit to the m_{WW} spectrum in the low-mass sideband. The uncertainty on $\alpha_{MC}(m_{WW})$ derives from the ones on the m_{WW} fits defining the numerator and the denominator of the ratio, as shown in Figure 5.15, which is proportional to the statistics of simulated events surviving the analysis selections in both sideband and signal region.
The pink and the yellow dashed lines denote, respectively, the $\alpha_{MC}(m_{WW})$ functions derived from the alternative parton shower model and an alternative fit parametrization. To account for these additional shape variations, the Herwig-based uncertainties (shaded bands) are enlarged by a factor of two, so that the α parameters are weaklier constrained in the final fit. In the same way, to reduce potential bias in the final limit extraction due to a possible wrong description of the W+jet shape, the uncertainty on the W+jet shape parameters fitted in sideband region is increased by 20%, as studied in Section 5.5.2.

5.6.2 Systematics on the signal model

The systematic uncertainties on the signal cross section upper limits include the uncertainty on the integrated luminosity of the data sample, the theoretical description of the signal model as well as the ones related to the event reconstruction in the detector.
Uncertainties on the heavy Higgs production cross section arise from the combined choice of PDFs and α_s, as well as from the renormalization (μ_r) and factorization (μ_f) scales. They are typically around 6–7% and 7–12%, respectively, for the ggH production mechanism, 1–2% and 2–5% for qqH process.
Uncertainties on the signal lineshape due to the interference modeling varies for the ggH and qqH modes. For the gluon fusion, the prescription described in [35] is adopted, which considers the NNLO contribution to the signal interfering with the $gg \to VV$ background process. In contrast, for the VBF production, the uncertainty is assigned with LO precision from

Source	W+jets		$t\bar{t}$		Single-t		Di-boson		WW_{ewk}	
	0+1-j	2-j	0+1-j	2-j	0+1-j	2-j	0+1-j	2-j	0+1-j	2-j
Experimental sources										
Lumi	-	-	2.6%	2.6%	2.6%	2.6%	2.6%	2.6%	-	2.6%
$\varepsilon_{\text{trig}}$	-	-	1%	1%	1%	1%	1%	1%	-	1%
ε_{ℓ}^{id}	-	-	2%	2%	2%	2%	2%	2%	-	2%
$\varepsilon_{\text{btag}}$	-	-	1.7%	1.5%	3.3%	3%	0.6%	0.5%	-	0.7%
ℓ scale	-	-	0.4%	0.5%	<0.1%	<0.1%	1%	1.5%	-	1%
ℓ res.	-	-	<0.1%	<0.1%	<0.1%	<0.1%	<0.1%	<0.1%	-	<0.1%
Jet scale	2.5%	2%	4%	4%	4%	7%	3%	7.5%	-	4.6%
Jet res.	1%	1.9%	0.4%	3%	0.9%	8.3%	0.7%	4.3%	-	6.3%
V-tag	-	-	-	-	-	-	9%	9%	-	9%
Background estimation										
N_{top}	-	-	6.5%	26.5%	6.5%	26.5%	-	-	-	-
$N_{\text{W+jet}}$	5-8%	22%	-	-	-	-	-	-	-	-
N_{VV}	-	-	-	-	-	-	20%	20%	-	20%

Table 5.4: Summary of systematic uncertainties in the background yield, relative to the expected number of background events. All systematic sources in the list are treated as uncorrelated, with log-normal priors in the statistical interpretation. Systematic effects, that are varying the yield of a background source by less than 0.1%, are neglected.

renormalization and factorization scale variations in the `Phantom` generation, as described in Section 5.2. Another important source is represented by the knowledge of the ggH cross section inside the different exclusive jet bins in which the analysis is categorized, which is around 20-25%.

The experimental sources affecting physics objects properties are represented by jet energy scale and resolution, electron and muon energy scale and resolution and un-clustered particles, which contribute to the missing energy measurement. To evaluate them, the event selection is applied to signal events after varying leptons and jets four-momenta within one respective standard deviation (or applying an appropriate energy/momentum smearing in case of resolution uncertainties) of the corresponding uncertainty on energy or momentum scale and resolution. In this process, variations on the lepton and jet four-momenta are propagated consistently to the $\not\!\!E_T$ vector. The largest relative change in the signal efficiency, compared to the central value, is taken as a measurement of the systematic effect for that specific source. Uncertainties on data-to-simulation scale factors for W-jet tagging, lepton trigger, lepton identification and b-jet tagging are also considered.

Uncertainties on physics objects scale and resolution are affecting also the m_{WW} signal shape. In fact, after applying the analysis selection, the largest relative changes in the peak position and width of the fitted signal shape, compared to the default one, are taken as systematic uncertainty on the parameters of the adopted signal model. The dominant systematic source on the m_{WW} signal shape comes from the jet energy scale (resolution), approximately 3% (2%) on the peak position for the 0+1-jet category and 4% (3%) for the 2-jet one, 2% (5%) on the width for 0+1-jet bin, 5% (10%) on 2-jet one.

A summary of the impact of all the systematic sources on both signal rate and shape is presented in Table 5.5.

Source	ggH-Rate		qqH-Rate		ggH-Shape		qqH-Shape	
	0+1-jet	2-jet	0+1-jet	2-jet	peak	width	peak	width
Experimental sources								
Luminosity	2.6%	2.6%	2.6%	2.6%	-	-	-	-
ℓ scale	2%	3.5%	1.5%	1.8%	<0.1%	<0.1%	<0.1%	<0.1%
$\varepsilon_{\text{trig}}$	1%	1%	1%	1%	-	-	-	-
ε_{ℓ}^{id}	2%	2%	2%	2%	-	-	-	-
$\varepsilon_{\text{btag}}$	0.5%	0.2%	0.5%	0.2%	-	-	-	-
Jet scale	3-4%	4-5%	4-4.5%	4-4.5%	2-3%	1.5-2%	3.5-4%	3.5-5%
Jet resolution	2-2.5%	7-8%	3.5-5%	8.5-10%	2%	4-5%	2.5-3%	8.5-10%
V-tagging	9%	9%	9%	9%	-	-	-	-
Theoretical sources								
PDF+α_s ggH	9-12%	9-12%	-	-	-	-	-	-
PDF+α_s qqH	-	-	3.5-6%	3.5-6%	-	-	-	-
Interf. ggH	10%	10%	-	-	-	-	-	-
Interf. qqH	-	-	10%	10%	-	-	-	-
ggH 0+1-jet	26%	-	-	-	-	-	-	-
ggH 2-jet	6%	20%	-	-	-	-	-	-

Table 5.5: Summary of systematic uncertainties affecting signal yield, relative to the expected number of signal events, peak position and width after the analysis selection. All systematic sources in the list are treated as uncorrelated. Their values for each specific mass point is not indicated, only the range of variation as a function of the Higgs mass is reported.

5.7 Statistical interpretation

Combining the background estimate, described in Section 5.5, with the signal modelling, reported in Section 5.4, Figure 5.18 shows the final m_{WW} distributions in the signal region for the 0+1-jet $\mu\nu$+W-jet category (left), the $e\nu$+W-jet channel (middle) and the 2-jet bin (right). The uncertainty band takes into account both the fit uncertainty on the m_{WW} spectrum in the m_{pruned} sideband region, the uncertainty on the extrapolation function ($\alpha_{\text{MC}}(m_{\text{WW}})$) and the normalization uncertainties for all the background contributions, which are detailed in Table 5.4. These three categories are used simultaneously in an unbinned maximum likelihood fit to extract upper limits on the model parameters, as a function of different m_{H} hypotheses, through a modified frequentist method referred to as the CL_{S} method [149, 150], where a likelihood ratio test statistics is adopted and nuisance parameters are profiled. In the likelihood ratio, the total number of observed events is compared to the signal and the background predictions by means of a product of Poisson probabilities.
The predictions are subject to the multiple uncertainties described in Section 5.6, which are handled by introducing nuisance parameters with log-normal priors, that modify parametrically the expectations for both signal and background processes. A signal strength modifier (μ) is used to scale the Higgs boson cross section for all the production mechanisms by a common factor with respect to the SM predictions, keeping the decay branching fractions unchanged.

5.7.1 SM-like Higgs boson search

The combined results obtained for a heavy Higgs boson with SM-like couplings is shown in Figure 5.19 (left), where the expected and the observed exclusion limit at 95% CL are set on the signal strength modifier. Then, in Figure 5.19 (right), the expected and the observed

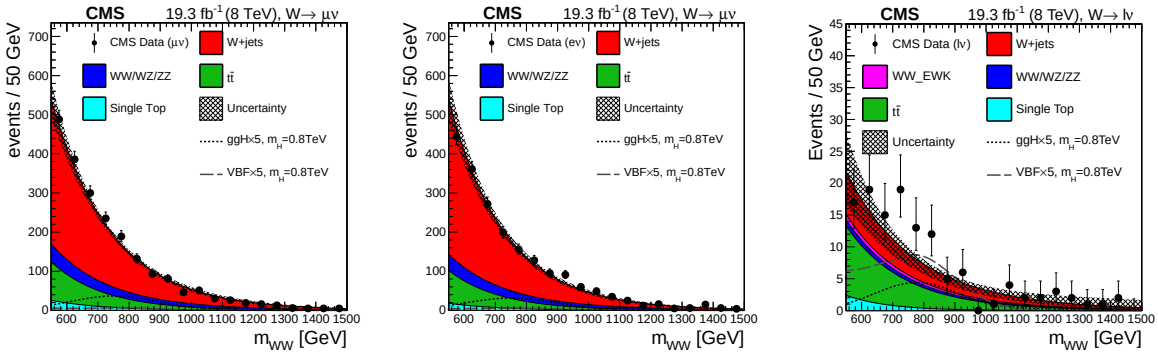

Figure 5.18: The final m_{WW} invariant mass distribution for the 0+1-jet bin category muon channel only (left), electron one (middle) and for the 2-jet bin (right). Points represent the observed data, shaded graphs represent the background and dashed graphs represent five times the expectation for a SM-like Higgs boson with $m_{\mathrm{H}} = 800$ GeV produced via ggH and qqH mechanisms, separately.

upper limits obtained combining the three analysis categories are compared to the individual inputs from 0+1-jet and 2-jet channels. The observed sensitivity is driven by 0+1-jet category for low m_{H} values, up to 800 GeV, then the VBF one becomes more performant due to the increase of qqH production cross section and a better signal-to-background ratio. The observed exclusion limit for the SM-like Higgs boson hypothesis varies from 1.1 times the SM cross-section for $m_{\mathrm{H}} = 600$ GeV to 5 times the SM expectation at 1000 GeV.

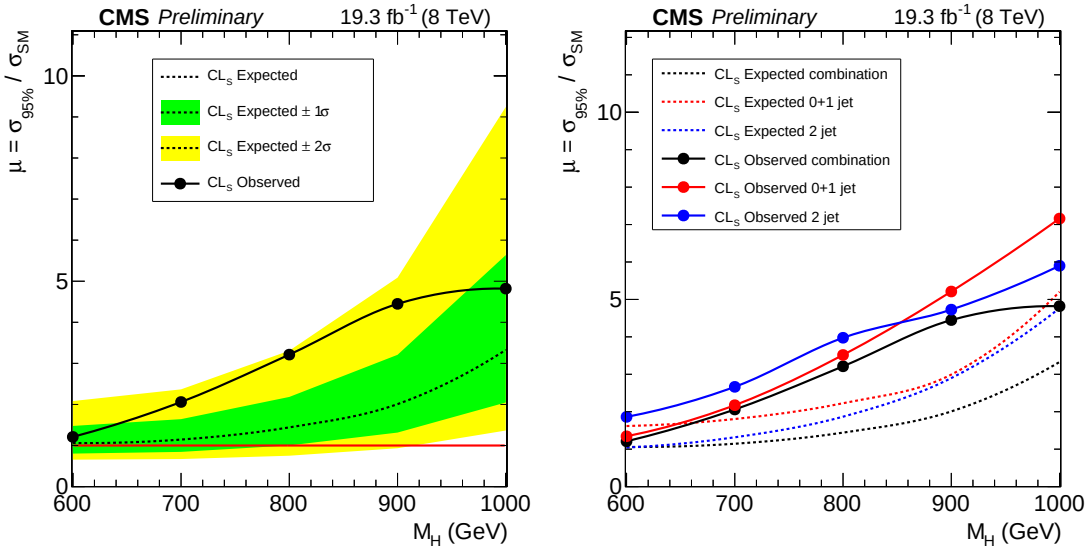

Figure 5.19: (Left) Observed (solid) and expected (dashed) 95% CL upper limit on $\mu = \sigma/\sigma_{\mathrm{SM}}$, obtained via the asymptotic CL_s technique, for a SM-like Higgs boson decaying into WW$\to \ell\nu q\bar{q}'$. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid horizontal red line at unity indicates the expectation for a SM-like Higgs boson. (Right) Comparison between esclusion limits for the single categories.

Since the observed limit reaches the edge of the $2\,\sigma$ band between 700-900 GeV, mainly driven by an excess in data with respect to the background-only prediction in the 2-jet category, the local significance has been computed as a function of m_{H}.

This observed p-value is defined as the probability for the background-only hypothesis to fluctuate up to the local excess measured in data. Low p-values indicate a large incompatibility with the background-only hypothesis, while a p-value equal to 0.5 is assigned when data underfluctuates with respect to the predicted amount of background.
Figure 5.20 shows the observed and the expected (for $\mu = \mu_{\mathrm{SM}}$) p-values for the 0+1-jet bin (left), 2-jet category (middle) and the ultimate combination (right). Local excesses of $2.6\,\sigma$ and $2.4\,\sigma$ are observed in the mass range between 700 and 800 GeV.

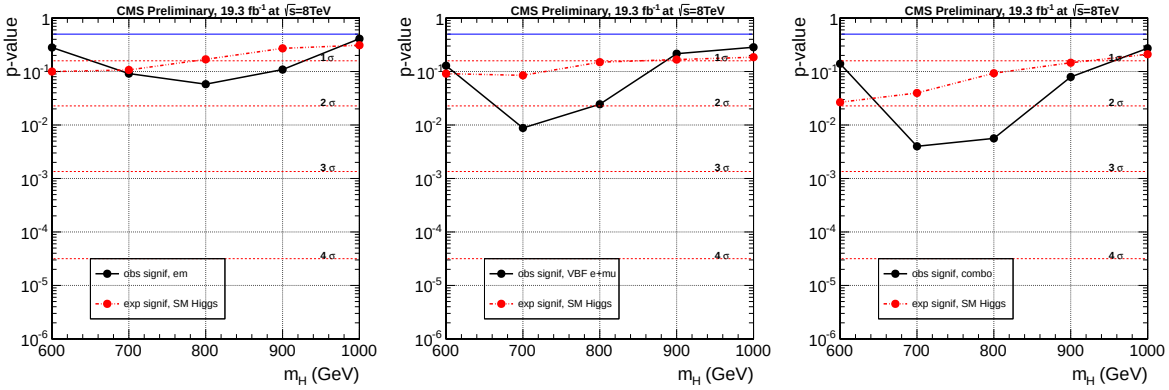

Figure 5.20: Expected and observed local p-values for a SM-like Higgs boson as a function of m_{H}: (left) 0+1-jet category, (middle) 2-jet one and their combination (right). The expected p-values are computed considering a possible signal with strength equal to the SM expectation.

5.7.2 Electroweak singlet interpretation

The re-interpretation of the SM-like Higgs results in the context of the electroweak singlet model is performed only where $C'^2 \leq (1 - \mathrm{BR}_{\mathrm{new}})$, that corresponds to the cases where the new state is expected to have an equal or narrower width with respect to the SM scenario. Results are presented distinguishing between $\mathrm{BR}_{\mathrm{new}} = 0$ and $\mathrm{BR}_{\mathrm{new}} > 0$ cases.
Under these hypotheses, signal samples with different Higgs boson widths are generated, scanning the two dimensional C'^2 and $\mathrm{BR}_{\mathrm{new}}$ space. For these discrete set of points, a limit and a local p-value are computed as a function of the new resonance mass. No significant excesses are observed and the local p-value remains always around $2.5\,\sigma$ in the m_{WW} between 700-800 GeV.
Figure 5.21 and Figure 5.22 show, respectively, the 95% CL observed and expected exclusion limits on the BSM signal strength as a function of m_{H} and C'^2, for three different $\mathrm{BR}_{\mathrm{new}}$ configurations ($\mathrm{BR}_{\mathrm{new}} = 0.0, 0.2, 0.5$). Since the sensitivity of this analysis is not enough to exclude the predicted signal strength, observed and expected exclusion contours for three and four times the expectation are reported.

5.8 Higgs high mass combination

The results presented in this Section are obtained by combining searches exploiting different Higgs boson production modes and final state topologies [159], which are listed in Table 5.6. All the final states are exclusive, thus there is no overlap of events among different channels. Individual analysis strategies are not detailed here, since this would go beyond the scope of

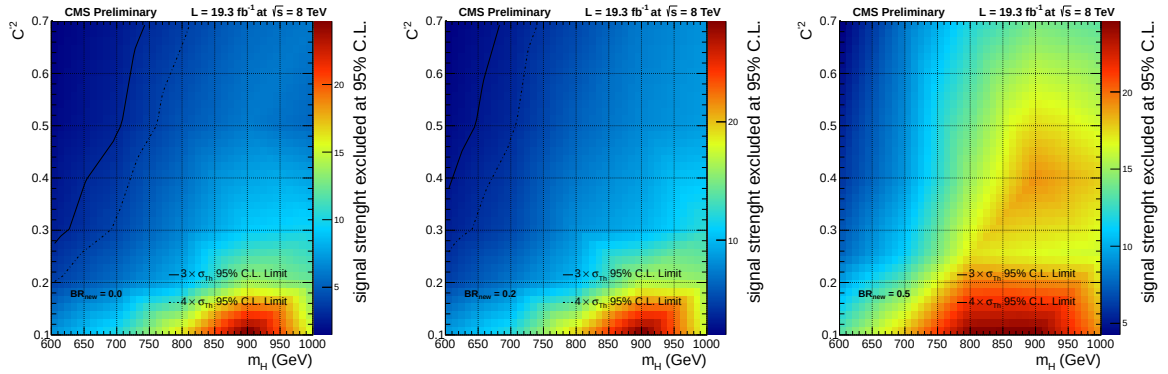

Figure 5.21: Observed upper limits at the 95% CL on the electroweak singlet model. Upper limits are displayed as a function of the heavy Higgs boson mass and the model parameter C'^2, for different values of $\mathrm{BR}_{\mathrm{new}}$. From left to right, limits for $\mathrm{BR}_{\mathrm{new}} = 0.0, 0.2, 0.5$. Observed exclusion contours for three and four times the expected signal strength are shown as solid and dashed black lines, respectively.

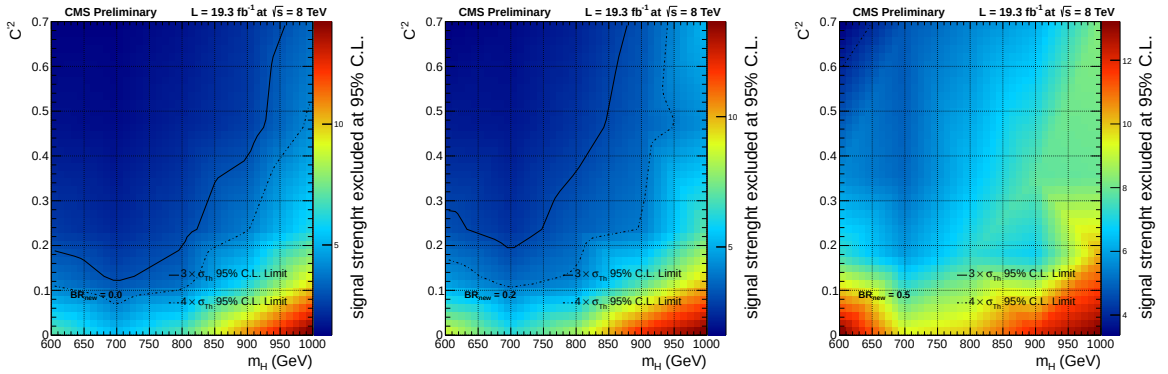

Figure 5.22: Expected upper limits at the 95% CL on the electroweak singlet model. Upper limits are displayed as a function of the heavy Higgs boson mass and the model parameter C'^2, for different values of $\mathrm{BR}_{\mathrm{new}}$. From left to right, limits for $\mathrm{BR}_{\mathrm{new}} = 0.0, 0.2, 0.5$. Expected exclusion contours for three and four times the expected signal strength are shown as solid and dashed black lines, respectively.

this thesis. The combination requires a simultaneous analysis of all the categories, accounting for both statistical and systematic uncertainties and their correlations. Also in this case, upper limits are obtained through the frequentist $\mathrm{CL_S}$ method.

The combined results for the SM-like Higgs boson search are displayed in Figure 5.23 (left), where the observed upper bound on the signal strength is shown for each final state channel. The expected combined 95% CL limit of the six channels is plotted as a dashed black line, while the yellow shaded region is the $\pm 2\,\sigma$ uncertainty on the expected limit. On the right hand side, the observed and expected performance are reported separately for each final state, exploiting WW (on top) and ZZ (on bottom) signatures.

In the lower mass region, the most sensitive channels are $\mathrm{H} \to \mathrm{ZZ} \to 4\ell$ and $\mathrm{H} \to \mathrm{WW} \to 2\ell 2\nu$. At the highest masses, the $\mathrm{H} \to \mathrm{ZZ} \to 2\ell 2\nu$ channel has the best sensitivity, while $\mathrm{H} \to ZZ \to 4\ell$, $\mathrm{H} \to \mathrm{WW} \to 2\ell 2\nu$, $\mathrm{H} \to \mathrm{WW} \to \ell\nu qq$ and $\mathrm{H} \to \mathrm{ZZ} \to 2\ell 2q$ contribute significantly. Features in the combined observed limit are due to corresponding ones in the limits of the individual channels. In fact, below 400 GeV, there are oscillations in the observed limit due to the high resolution of the most sensitive channel, $H \to \mathrm{ZZ} \to 4\ell$, and the narrow width of the Higgs boson candidate in this mass range. An excess in the combined limit around

Decay	Production	Final states	N_{ch}	m_H range [GeV]	σ_{m_H}/m_H
WW → $\ell\nu\ell\nu$	un-tag	$((ee, \mu\mu), e\mu)$ + (0/1-j)	4	145–1000 [a,b]	20%
	VBF tag	$((ee, \mu\mu), e\mu)$ + $(jj)_{VBF}$	2	145–1000 [ab]	20%
WW → $\ell\nu qq$	un-tag	$(e\nu, \mu\nu)$ + $(jj)_W$	2	180–600	5–15%
	un-tag	$(e\nu, \mu\nu)$ + $(J)_W$ + (0+1-j)	2	600–1000 [b]	5–15%
	VBF tag	$(e\nu, \mu\nu)$ + $(J)_W$ + $(jj)_{VBF}$	1	600–1000 [b]	5–15%
ZZ → $2\ell2\ell'$	un-tag	$4e$, 4μ, $2e2\mu$	3	145–1000	1–2%
	VBF tag	$(4e, 4\mu, 2e2\mu)$ + $(jj)_{VBF}$	3	145–1000	1–2%
	un-tag	$(ee, \mu\mu)$ + $(\tau_h\tau_h, \tau_\ell\tau_h, \tau_\ell\tau_\ell)$	8	200–1000	10–15%
ZZ → $2\ell2\nu$	un-tag	$(ee, \mu\mu)$ + (0 or ≥ 1 jets)	4	200–1000	7%
	VBF tag	$(ee, \mu\mu)$ + $(jj)_{VBF}$	2	200–1000	7%
ZZ → $2\ell2q$	un-tag	$(ee, \mu\mu)$ + $(jj)_Z^{0,1,2\text{ b-tag}}$	6	230–1000 [c]	3%
	un-tag	$(ee, \mu\mu)$ + $(J)_Z^{0,1,2\text{ b-tag}}$	6	230–1000 [c]	3%
	VBF tag	$(ee, \mu\mu)$ + $(jj)_Z^{0,1,2\text{ b-tag}}$ + $(jj)_{VBF}$	6	230–1000 [c]	3%
	VBF tag	$(ee, \mu\mu)$ + $(J)_Z^{0,1,2\text{ b-tag}}$ + $(jj)_{VBF}$	6	230–1000 [c]	3%

Table 5.6: Summary of the analyses included in the combination. The main contribution in the untagged categories is always from ggH, while the $(jj)_{VBF}$ refers to a di-jet pair consistent with the VBF topology. In contrast, $(jj)_{W(Z)}$ and $(J)_{W(Z)}$ refer to a di-jet pair and single merged jet with an invariant mass or a pruned mass consistent with a W (Z) hadronic decay. The superscript "0,1,2 b-tags" refers to the three possible categories of b-tag multiplicities.

[a] Electroweak singlet model interpretation starts at 200 GeV to avoid contamination from SM Higgs.
[b] 600-1000 GeV for $\sqrt{s} = 8$ TeV only.
[c] For $\sqrt{s} = 8$ TeV only.

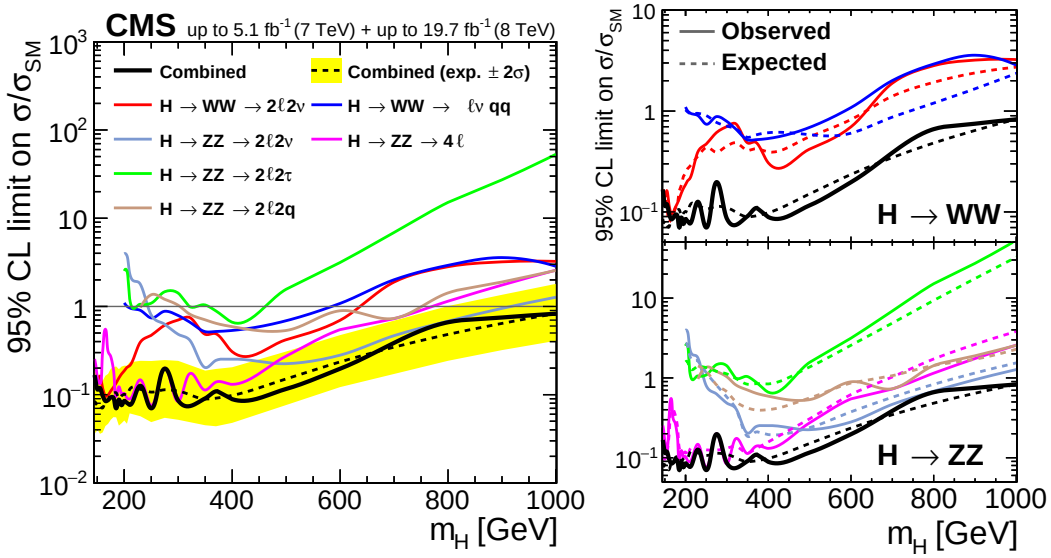

Figure 5.23: Upper limits at the 95% CL for each of the contributing final states and their combination (left). The observed and expected limits of the six individual channels are compared with each other and with the combined results (right), for H → WW channels (top right panel) and H → ZZ ones (bottom right panel) separately.

280 GeV is related to a simultaneous excess in both four leptons and H → WW → $2\ell2\nu$ channel. Moreover, the small excess of observed around 700 GeV in the boosted semi-leptonic H → WW analysis is not supported by the other channels and is reduced to less than 0.5 σ

in the combination. The combined upper limit, at the 95% CL, on the product of the cross section and branching fractions exclude a Higgs boson with SM-like couplings in the full search range $145 < m_{\mathrm{H}} < 1000$ GeV.
Finally, also this combined result is interpreted in the electroweak singlet model framework. Figure 5.24 shows the expected and observed contours at 95% CL on the singlet scalar cross section, with respect to its expectation, in the (m_{H},C'^2) plane for various values of $\mathrm{BR}_{\mathrm{new}}$. A large region of the parameter space foreseen has been excluded. In Figure 5.24, the indirect constraint from the direct measurement of the Higgs coupling to V-bosons is also reported, $C'^2 < 0.28$ for $\mathrm{BR}_{\mathrm{new}} = 0$, showing a significant improvement in the exclusion sensitivity through this combination of direct searches. Eventually, Figure 5.25 (right) shows the observed excluded cross section (in pb) as a function of the BSM resonance mass and its relative intrinsic width with respect to the SM expectation.

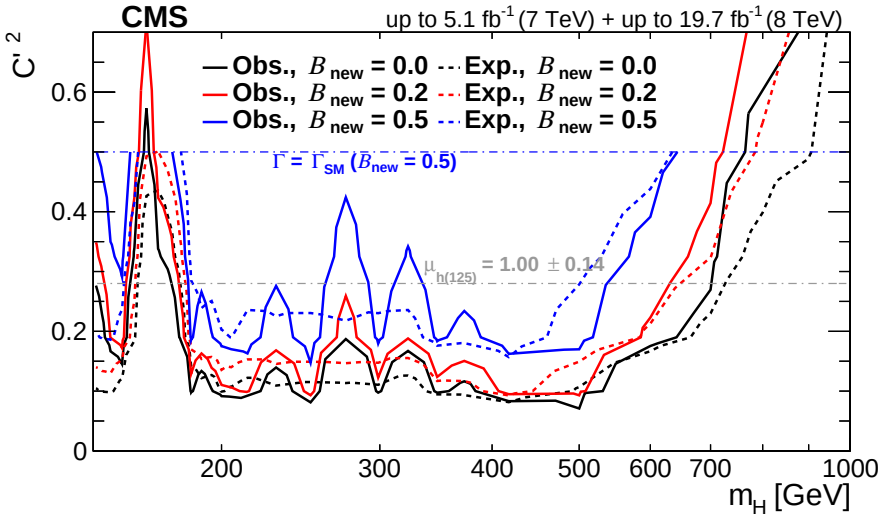

Figure 5.24: Upper limits at the 95% CL on the electroweak singlet extension of the SM from the combination of the di-boson final states (WW-ZZ), displayed as a function of m_{H} and the model parameter C'^2 for different values of $\mathrm{BR}_{\mathrm{new}}$. The upper dash-dotted line indicates where, for $\mathrm{BR}_{\mathrm{new}} = 0.5$, the width of the BSM Higgs boson reaches the width of a SM-like one. The lower dash-dotted line displays the indirect limit, at 95% CL, on C'^2 from the measurement of the SM Higgs coupling to V-bosons.

5.9 Conclusion

A search for a new SM-like heavy Higgs boson decaying into WW pairs has been presented, in which one of the W-bosons decays leptonically and the other one hadronically, considering a $\ell\nu$+W-jet ($\ell = \mu, e$) final state. This channel has the largest cross section times branching fraction predicted for a SM-like Higgs, with $m_{\mathrm{H}} > 180$ GeV, although the final state is contaminated by large background events from SM sources, mainly W+jets and $t\bar{t}$.
Jet substructures are used to reduce these backgrounds identifying high p_T jets containing the fragmentation products of a boosted hadronically decaying W-boson.
Events are then categorized according to the additional number of reconstructed jets with $p_T > 30$ GeV, to enhance the analysis sensitivity trying to disentangle signal candidates produced by gluon fusion or VBF mechanism. A crucial aspect is represented by the Higgs

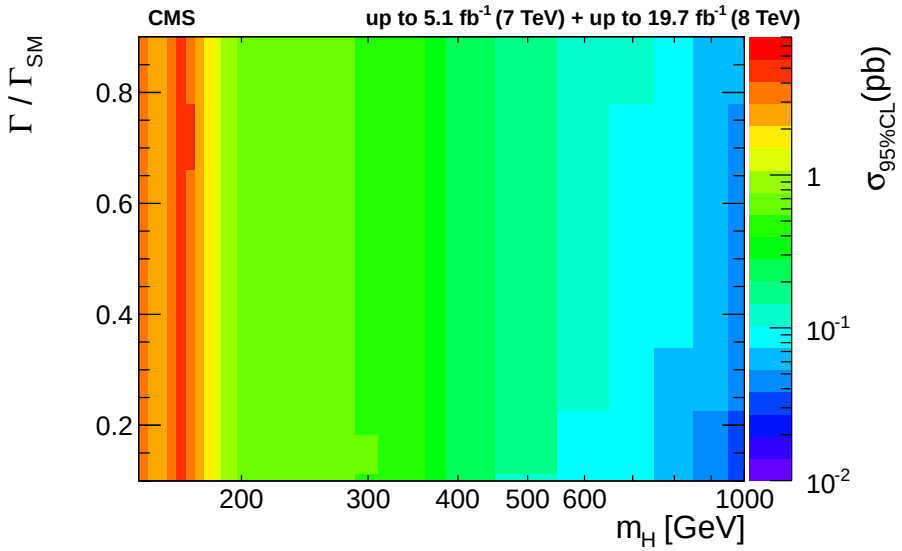

Figure 5.25: Upper bounds on the signal cross section (in pb) are reported as a function of Higgs mass and relative width of the resonance $\Gamma/\Gamma_{\mathrm{SM}}$.

lineshape model, which should account for the correct Higgs boson propagator description, known as Complex-Pole-Scheme (CPS), as well the interference effect between the heavy Higgs and the continuum di-boson production.

With no significant excess observed above the predicted background yield, an upper limit on the SM-like Higgs boson production cross-section is set through an unbinned fit to the reconstructed di-boson invariant mass spectrum (m_{WW}). A deviation with a local significance of $2.6\,\sigma$ is observed, mainly driven by an excess in the two-jet-bin (VBF) category in the range $700 < m_{\mathrm{WW}} < 800$ GeV.

The result is also interpreted in a BSM heavy Higgs scenario, called electroweak singlet model, based on an effective theory which predicts the existence of two Higgs-like scalar particles, with the mass of the lighter one being around 125 GeV.

Eventually, the analysis has been combined with other exclusive searches for a heavy Higgs boson, covering all the accessible WW and ZZ decay signatures. In the case of a heavy Higgs boson with SM-like couplings and decays, its existence has been excluded over the entire explored range of $145 < m_{\mathrm{H}} < 1000$ GeV. While, for the electroweak singlet partner of the SM Higgs, the large part of the C'^2 versus mass parameter space is excluded for various values of $\mathrm{BR}_{\mathrm{new}}$, giving tighter constraint than indirect limits from measurements at Higgs pole.

Chapter 6

Vector Boson Scattering in view of the CMS upgrade for the HL-LHC

In the physics collisions provided by the LHC during Run-I, an integrated luminosity of 25 fb^{-1} has been collected at $\sqrt{s} = 7, 8$ TeV. While no evidences for BSM physics has been seen yet, the results obtained so far have dramatically changed the landscape for these searches through the measurement of the Higgs boson properties and first direct or indirect exclusion for new physics. In fact, many models have been totally excluded, others have their parameters sharply limited. To enhance the sensitivity of CMS to new physics phenomena and to provide precise measurements in the Higgs sector, a powerful demand for higher instantaneous luminosity is mandatory.
The LHC machine group aims to achieve higher peak and integrated luminosity, well above those for which CMS was designed. Thus, the CMS detector requires upgrades (Phase-II) to preserve the optimal efficiency and resolution also in these conditions [166].
In this Chapter, the main goals of the high luminosity LHC programme (HL-LHC) are described in Section 6.1, while, in Section 6.2, a brief overview on the proposal for the CMS Phase-II upgrade is presented. Then, in Section 6.3, a description of the Delphes framework [167] is exposed together with its commissioning and tune to correctly reproduce the CMS detector performance in different scenarios. Section 6.4 gives an introduction to Vector Boson Scattering (VBS) measurements in view of the HL-LHC, as well as detailed information on the event simulation, while Section 6.5 describes the general event reconstruction and the selection strategy adopted in the analysis. Section 6.6 describes how backgrounds, arising from jets mis-identified as leptons, are estimated. Then, Section 6.7 and Section 6.8 are dedicated to the fully leptonic same sign WW and to the fully leptonic WZ bosons scattering analysis, respectively. Eventually, in Section 6.9, the ultimate expected sensitivity to the longitudinal scattering cross section is presented by combining same sign WW and WZ bosons channels.

6.1 High luminosity LHC physics programme

The current plan of the LHC operation is based on a set of long data taking periods interleaved with long shutdowns, as summarized in Figure 6.1. Run-II has just begun and it will operate at the instantaneous luminosity of $\mathscr{L} = 1 \times 10^{34}$ cm^{-2}s^{-1} with 25 ns bunch spacing, experiencing an average of about 25 pileup interactions per bunch crossing, which coincides

with the operating scenario for which CMS was designed. This data taking period will provide around 300 fb^{-1} of data and the CMS experiment will undergo a series of upgrades from LS1 to LS2, called Phase-I upgrade [168].
The high luminosity period that follows the LS3 is referred to as HL-LHC or Phase-II. In this operating scenario, the instantaneous luminosity will reach 5×10^{34} cm^{-2}s^{-1}, from a potential peak value of 2×10^{35} cm^{-2}s^{-1} at the beginning of fills, delivering 250 fb^{-1} per year for further 10 years of operation. Under these conditions the pileup will rise substantially to the point of becoming a major challenge for the experiments and the performance degradation due to the integrated radiation dose will need to be addressed.

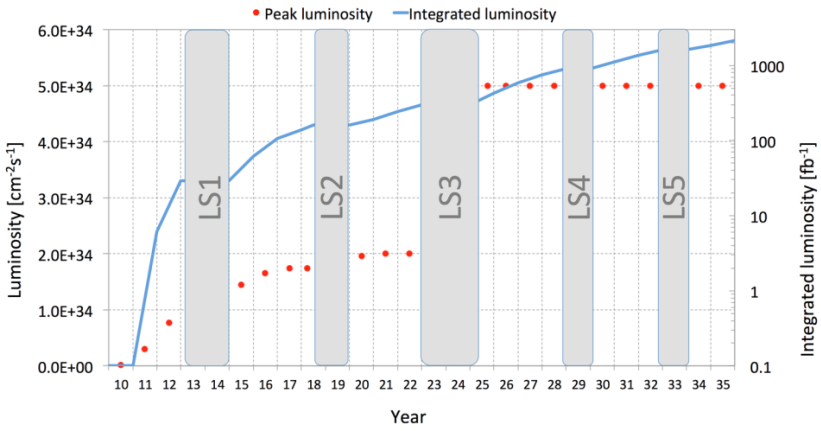

Figure 6.1: Schedule of the LHC operations towards 2035, showing preliminary dates for long shutdowns of the LHC and projected luminosities.

The CMS physics program at the HL-LHC will be based on the results obtained from the 300 fb^{-1} of data collected during Run-II, in terms of both precision measurements and searches for new physics. The study of the Higgs boson will be central in this program, through precise measurements of its couplings at percent level, as shown in Figure 6.2 (left), probing its tensor structure and measuring differential rates. In addition, the measurement of its coupling to the second generation of fermions will be possible via $h \to \mu\mu$ decays, as well as the study of Higgs self coupling through the di-Higgs production cross section.
The role of the Higgs boson in the EWSB sector will be tested studying the VBS process, which is also sensitive to new physics via anomalies in triple and quartic gauge couplings (aTGC and aQGC respectively).
New particles are expected at the TeV scale but have not yet been seen, thus the very high integrated luminosity collected at the HL-LHC will extend the reach of searches for new heavy bosons, very narrow width resonances predicted in models like extra dimensions, compositeness andlepto-quarks, up to 6 TeV or more.
Moreover, indirect searches of Dark-Matter (DM), by means of either searches for SUSY particles or exploring mono-object signatures (jet, photon, V-boson, top, Higgs), are expected to be competitive to direct experiments if the missing transverse energy resolution will br kept at similar level of the present data. New physics can also be detected through the study of rare B-meson decays, such as $B_s \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$, for which a high integrated luminosity is fundamental as proved in Figure 6.2 (right).

Finally, many interesting SM measurements will be performed during the HL-LHC, such as the measurement of the W-boson mass, the strong coupling constant α_s and the effective lepton mixing angle, as well as double differential cross sections of events with jets, top quarks, W and Z-bosons.

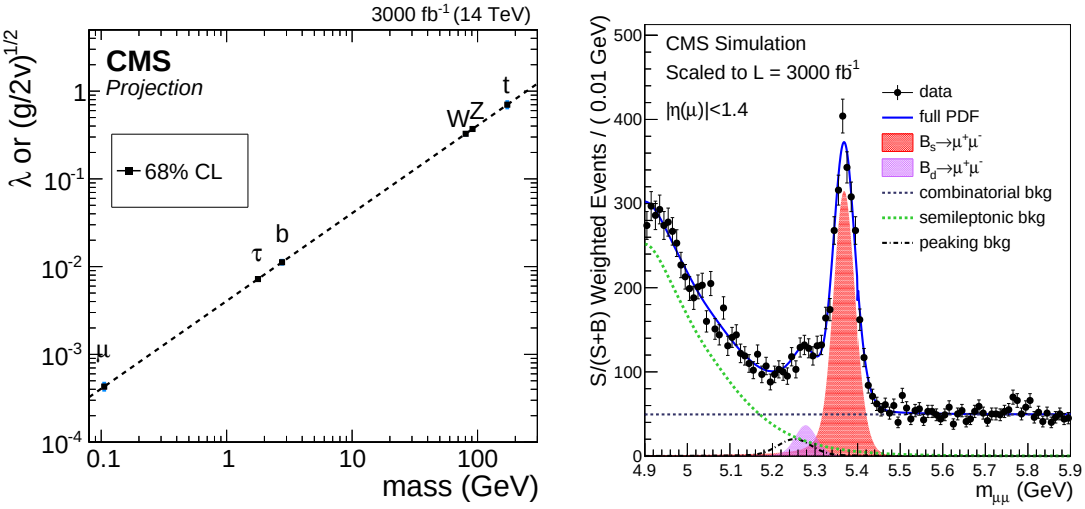

Figure 6.2: (Left) Projected precision on Higgs boson couplings as a function of boson or fermion masses after 3 ab^{-1} of data. (Right) Projections of the mass fits for the $B_{s,d} \to \mu^+\mu^-$ measurements to 3 ab^{-1} of integrated luminosity, assuming CMS Phase-II performances for barrel only.

6.2 The Phase-II upgrade of the CMS detector

The goal of the Phase-II upgrade is to keep unchanged the performances, showed by CMS during Run-I, in terms of reconstruction efficiency, energy resolution and background rejection for all the physics objects (electrons, muons, photons, taus, jets and missing energy).
The major challenge is represented by developing sensors able to overcome the radiation damage, due to the large integrated luminosity collected, measuring hard scattering products in a high pileup environment.
The LHC will produce collisions at a rate of about 5×10^9 Hz, where the flow of the particles through the detector will cause a significant damage on both the active materials and the on-board electronics, resulting in a progressive degradation of the performances.
To design a detector able to collect 3 ab^{-1} with stable performances, predictions of the dose rate and particle fluence are needed, as well as to know their effects on the different subdetectors. Test beams and Run-I data are used to tune a specific set of simulations. The damage produced in the detectors by particle radiation depends on both the sensor technology and the total amount of the absorbed dose, which increases as a function of η.
In silicon sensors, radiation damages the lattice increasing the dark current and forming trapping centers that reduce the charge collection efficiency.
For the $PbWO_4$ crystals of ECAL detector or the plastic scintillators used in HCAL, the effect is a loss in the trasmission efficiency, reducing the scintillation light especially in the endcaps, where the loss reaches more than 90%.
At the nominal luminosity foreseen for the HL-LHC, an average of 140 pileup interactions is

expected, which degrade triggers and offline reconstruction performances producing additional hits in the tracking system and energy depositions in the calorimeters. Figure 6.3 (left) shows the track multiplicity for an event recorded during Run-I with 78 reconstructed primary vertices. Figure 6.3 (right) gives an illustration of the out-of-time pileup, which affects pulses produced by scintillators (silicon devices are usually fast in response with respect to the buch spacing of 25 ns).

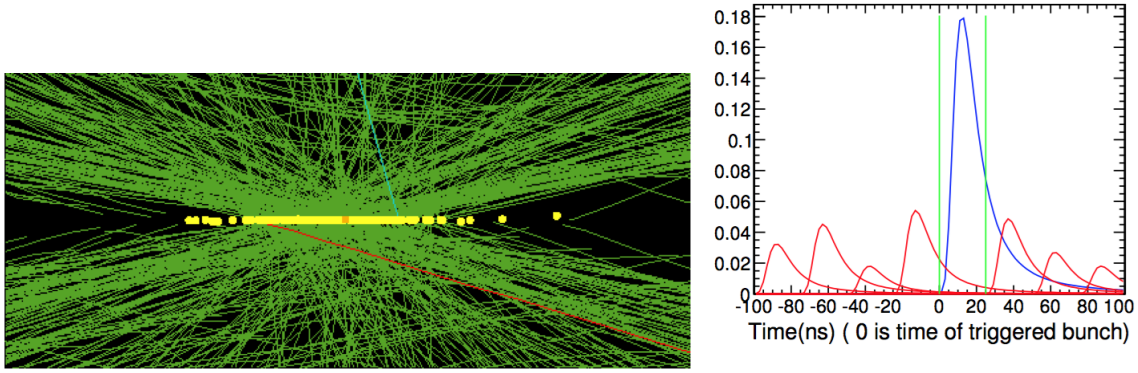

Figure 6.3: (Left) Track occupancy in an event with 78 reconstructed vertices recorded during 2012. (Right) Individual pulses from a calorimeter as a function of time. The signal from the triggered buch crossing is located at 0 ns, bounded by the green vertical lines, while in red are early (late) pulses falling inside the signal window. Pulses are not stacked one on top the other.

The nominal operating scenario consists of collisions with 25 ns bunch spacing, an average of 140 pileup interactions per bunch crossing and an instantaneous luminosity of $\mathscr{L} = 5 \times 10^{34}$ cm^{-2}s^{-1}. Considering these conditions and the expected radiation damage, both the tracker and the endcap calorimeters need to be replaced for the Phase-II.
In addition, the precise measurement the Higgs boson properties requires the ability of efficiently trigger low p_T leptons and photons. Thus, both the trigger electronics (L1 trigger) and the software one (HLT) must be upgraded too. In particular:

- **Tracker upgrade:** to maintain high track reconstruction efficiency in the new pileup conditions, the granularity of both pixel and outer strip tracker will be increased by a factor four. In the strip tracker, this goal will be achieved by shortening the sensor length without changing the pitch. Moreover, a set of design changes will lead to an important decrease of the material budget, providing better p_T resolution and a lower rate of photon conversions. Pixel layers will be replaced with smaller and thinner sensors, to improve the resolution on the impact parameter with direct consequences on b-tagging and high p_T tracks for jet substructure. The system coverage will be extended in the forward region up to $\eta = 4$, to better match the calorimetry coverage and to provide better reconstruction of forward jets produced by VBS processes, described in Section 1.6.

- **Endcap calorimeter:** ECAL and HCAL endcaps will be replaced by the High Granularity Calorimeter (HGCal), which has both electromagnetic and hadronic sections with an excellent transverse and longitudinal segmentation, providing a three-dimensional picture of the shower development. The electromagnetic section consists of 30 tungsten plates interleaved with active material of silicon sensors, corresponding to 25 X_0 and 1 λ_I.
The hadronic part has a front section of 12 brass and copper plates interleaved with sili-

con sensors for a total depth of 3.5 λ_I. After this, a more standard sampling calorimeter is placed, made by brass plates and plastic scintillators and readout by wavelength shifter fibers to reach 10 λ_I for the full calorimeter.

- **Endcap muon system:** to keep an optimal L1 trigger acceptance in the endcap region, four new muon stations will be added. Two innermost layers using Gas Electron Multiplier (GEM) chambers, which have a better spatial resolution, and two outer ones using RPC with an excellent time resolution to mitigate background effects.
  Additional GEM stations will be also added to increase the coverage up to $\eta = 3$.

- **Trigger upgrade:** the present L1 trigger deadtime is limited to 3.4 μs by the tracker readout. For the Phase-II operation, the deadtime will be increased up to 12.5 μs to allow track reconstruction at L1, aiming to reduce the combinatorial background due to pileup. The proposed L1 trigger acceptance will be of 500 kHz, thanks to an upgrade of the electronics in each sub-detector, allowing to keep similar thresholds to the ones adopted during Run-I.
  At the same time, the Data Acquisition (DAQ) will be upgraded to increase the total bandwidth and computing power to accomodate both larger event size and rate, since the HLT will run at about 5 kHz.

A full simulation of the upgraded detector geometry was implemented, based on GEANT4 [121], to evaluate and compare the proposed upgrade with the Run-I performance. In particular, three different configurations are considered:

- **Phase-I detector** operating at an average of 50 pileup events ($\mathscr{L} = 10^{34}$ cm^{-2}s^{-1}) without radiation aging. It is used to assess a default performance benchmark.

- **Phase-II detector** operating at an average of 140 pileup events ($\mathscr{L} = 5 \times 10^{34}$ cm^{-2}s^{-1}), in which detector upgrades are assumed not to degradate with radiation, while an intermediate aging after 1 ab^{-1} is adopted for the barrel calorimetric system.

- **Phase-I aged detector** operating at an average of 140 pileup events, where the effects of radiation damage after 1 ab^{-1} are simulated for the muon system, the barrel calorimeter and tracker, while a damage after 3 ab^{-1} is used for the outer tracker and the endcap calorimeter.

These setups are used to optimize physics object reconstruction and identification techniques, while feasibility studies for HL-LHC analyses are performed on simulated events based on Delphes, in which a simplified CMS geometry is implemented together with an analytical parametrization of efficiencies and resolutions that are applied on generator level particles. No particle-matter interactions are simulated in constrast with a GEANT4 based simulation.

6.3 Delphes: fast simulation for collider experiments

The Delphes framework [167] is designed for feasibility phenomenological studies, taking as input particles produced by a given generator (ME+PS) and performing a realistic simulation of a general purpose detector.
In fact, Delphes simulates the response to particles produced in a collision of a detector

composed by a inner tracking system surrounded by electromagnetic and hadron calorimeters, as well as a muon system. The detector simulation is obtained as follows:

- **Pileup simulation:** starting from a generated hard scattering, a set of minimum bias events are superimposed to obtain a realistic hadron collider event. The number of in-time pileup events is generated according to a Poisson distribution, which central value is defined by the chosen luminosity scenario.
  After this, the hard scattering is placed, by definition, at the nominal interaction point $(x, y, z) = (0, 0, 0)$, while additional pileup events are randomly spread in space according to a Gaussian resolution function with a width defined by the beam spot resolution $(x, y, z) = (0, 0, \text{Gaus}(0, \sigma_z^{\text{BS}}))$. All these additional pileup events are generated in-time with the hard scattering, thus no out-of-time contribution is simulated via `Delphes`. Finally, if the distance between two vertices is less than the expected spatial resolution ($\Delta z < \sigma_z^{vtx}$), they are merged into a single one which position, along the z-axis, is given by a p_T weighted average of the two input ones.

- **Tracking simulation:** long lived particles are propagated through an ideal uniform solenoidal magnetic field, parallel to the beam direction and placed inside the tracker volume only. The trajectory of a charged particle is assumed to be helicoidal up to the surface of the calorimetric system. Charged tracks (muons, electrons and stable charged hadrons) have a defined probability of being reconstructed inside the tracker volume (reconstruction efficiency) and their generated transverse momentum is smeared according to a parametric description of the tracker resolution. Efficiencies and resolutions are usually parametrized as a function of the particle type, p_T and η. Finally, no fake tracks are simulated.

- **Vertexing simulation:** each track, independently whether it comes from the hard scattering or from a pileup interaction, is always associated to the right generated vertex. Thus, vertexes are not fitted from the single tracks and both vertex reconstruction efficiency and track-vertex wrong assignment are completely neglected.

- **Calorimeter simulation:** the electromagnetic calorimeter is responsible for the measurements of electron and photon energy, while the hadronic one measures the energy of stable charged and neutral hadrons. In both cases, these particles are completely stopped in these sub-detectors and only muons are allowed to reach the outer muon system. The calorimeter granularity is finite and defined to be exactly the same between the hadronic and the electromagnetic systems.
  The (η, φ) dimension of each cell can be easily defined and changed. Since no hadronic and electromagnetic showers are simulated, each particle is assumed to be fully contained by a single calorimeter cell (calo-tower). The coordinates of an energy deposit in the calorimeter are obtained by a uniform smearing along the (η,φ) tower surface.
  According to the particle type, particles interacting with the calorimeters deposit a fixed fraction of their energy in the corresponding `ECAL` or `HCAL` cell. By default, electrons and photons release their energy inside `ECAL`, stable hadrons like charged pions deposit their energy completely in `HCAL`, while unstable particles like kaons and lambdas leave 30% of their energy in `ECAL` and 70% in `HCAL`. Finally, muons and neutrinos are not releasing energy inside the calorimetric system.

Energy resolutions for ECAL and HCAL are independently parametrized according to the following empiric formula [169]:

$$\frac{\sigma_E}{E}(\eta) = \sqrt{\frac{a(\eta)}{\sqrt{E}} \oplus \frac{b(\eta)}{E} \oplus c(\eta)} \tag{6.1}$$

where $a(\eta)$ is called stochastic term, $b(\eta)$ is the noise term and $c(\eta)$ is the constant one, which rules the calorimeter resolution for high energetic objects. Therefore, ECAL and HCAL energies are independently smeared through a log-normal distribution with a width equal to the expected energy resolution.
The ultimate energy of a calorimetric tower is given by:

$$E_{\mathrm{CT}} = \sum_{\mathrm{i} \in \mathrm{CT}} \Big[\log(E_i \cdot f_{\mathrm{ECAL}}, \sigma_{\mathrm{ECAL}}(E_i, \eta_i)) + \log(E_i \cdot f_{\mathrm{HCAL}}, \sigma_{\mathrm{HCAL}}(E_i, \eta_i)) \Big] \tag{6.2}$$

where the sum runs over all the particles that are pointing to the same calorimetric tower, while f_{ECAL} and f_{HCAL} are the corresponding particle energy deposit fractions.

- **Particle-flow simulation:** a simplified PF approach is implemented in which tracking and calorimeter information are combined together to perform particle identification. Starting from the transverse momentum of each track as measured by the tracker system, ECAL and HCAL energies of the associated calorimetric tower are also computed. At this point, the track momentum is compared to the calo-tower energy and if the energy measured by ECAL is significantly larger than the one associated to the track:

$$E_{\mathrm{ECAL}} - E^{tk}_{\mathrm{ECAL}} > 2 \cdot \sigma_{\mathrm{ECAL}}(E, \eta) \tag{6.3}$$

where $E^{tk}_{\mathrm{ECAL}} = f_{\mathrm{ECAL}} \cdot p^{tk}_T$, a photon candidate pointing to that tower is created. In a similar way, if an excess is measured in HCAL a new neutral hadron is generated with energy equal to the difference between the calorimeter and the tracking measurements. If no excesses are found either in ECAL or in HCAL, only a charged track is kept. Finally, if no tracks are pointing to a calorimetric tower and if the energy of that tower measured by ECAL (HCAL) is above a minimum threshold, usually called zero-suppression placed at $3\,\sigma_{\mathrm{noise}}$, a photon (neutral hadron) candidate is added to the event.

- **Mitigation of pileup from charged particles:** the easiest strategy to reduce in-time pileup effects at hadron colliders is represented by the so called Charged Hadron Subtraction (CHS), already described in Section 2.4.4, which is also implemented in Delphes.

- **Mitigation of pileup from neutrals:** the average energy density related to neutral particles from pileup interactions is estimated starting from CHS particles and used to apply residual corrections on jet and lepton isolation. Two main approches are possible: jet area method [85] and grid median estimation [170], as implemented in FastJet [122]. In the former case, the event is clustered according to a jet algorithm, for this analysis k_t with $R = 0.4$, where the pileup density is evaluated as the median value of the ratio between the jet transverse momentum and its area, skipping jets with very low p_T or with a null area. In the latter one, the (η,φ) plane is divided into a fixed number of rectangular cells, so that the median value of the ratio between the total momentum

inside the cell and its area is considered a measurement of the pileup energy density. The estimations from these alternative methods are usually compatible and the second one is adopted since it requires less computing resources.

- **Electron** and **muon reconstruction** and **identification** are parametrized in `Delphes` by means of a set of defined probabilities dependent on the particle kinematics, which are derived from the full `GEANT4` based `CMS` simulation. This strategy is also used to properly smear their momentum/energy to account for detector effects, which are usually described by continous functions in the (p_T, η) plane.
  In addition, for each identified lepton (μ or e), energy deposits related to particles falling inside a chosen isolation cone around its direction are calculated, dividing them into three types: energy from charged hadrons (ch) pointing to the primary vertex, energy from photons (ph) and energy from neutral hadron candidates (nh). Residual pileup contributions associated to neutral particles are subtracted according to:
  $$I = \Big[\sum_{i=ch,\mathrm{PV}} p_T^i + \max(0, \sum_{i=ph} p_T^i + \sum_{i=nh} p_T^i - \rho \cdot \pi R^2) \Big] / p_T^\ell \tag{6.4}$$
  where $i = (ch, nh, ph)$ indicates different classes of particles inside the isolation cone, while ρ is the average pileup energy density.

- **Jet reconstruction:** jets are obtained clustering the `Delphes PF` particles according to a given clustering algorithm using the `FastJet` package. Jets are cleaned by identified leptons, according to a geometric ΔR matching.
  Pileup jet occupancy is reduced by considering `CHS` jets, while residual pileup effects in jet are corrected by the `L1` subtraction, as defined in Section 2.4.4. In `Delphes`, although every reconstructed particle has a response close to unity, since the energy/momentum smearing is, in most of the cases, simply Gaussian, the average reconstructed momentum of a jet is not perfectly matching the generated one due to reconstruction inefficiencies and a partial pileup subtraction. Thus, a residual calibration is still possible, but the typical size of these corrections are pretty small. For this reason, no additional `L2` corrections has been applied in the following `Delphes` based studies.
  In addition, the `MC` flavor of a jet is assigned depending on the flavor and the energy of partons clustered inside the jet by the reconstruction algorithm. Only simulated partons not coming from the matrix element, with at least one parton daughter are considered. If a b-quark or a c-quark is associated to the jet, the jet is assigned to be a b-jet or a c-jet, respectively, while in the other cases the flavour is assigned to be the one of the associated parton with the highest p_T. The probability for a jet with a certain flavor i to be b-tagged is obtained through a parametric description, as a function of jet η and p_T, of b-tagging efficiency and mis-tagging rate.

- **Missing energy** is obtained by the negative vector sum of all the `Delphes PF` candidates. No further corrections are applied, since jet energy corrections are expected to have a limited contribution on the missing energy performance in `Delphes`.

6.3.1 Delphes tune for CMS Phase-I detector

The performances of the `Delphes` simulation have already been tuned and validated looking at 8 TeV `CMS` full simulation, considering the `Phase-I` geometry and an average number of

pileup interactions $\langle N_{PU} \rangle = 20$. To properly compare the two detector descriptions, a common sample of events generated with `MadGraph` interfaced with `Pythia6` is used as input. Figure 6.4 (left) shows the muon transverse momentum resolution as measured in $Z/\gamma^* \to \mu\mu$ events, where the resolution is estimated as the variance of the relative difference between the reconstructed and the generated muon transverse momentum $(p_T^{reco} - p_T^{gen})/p_T^{reco}$, considering all the reconstructed muons matched to generated ones within $\Delta R < 0.2$. At the same time, electron and photon resolutions are shown in Figure 6.4 (right).

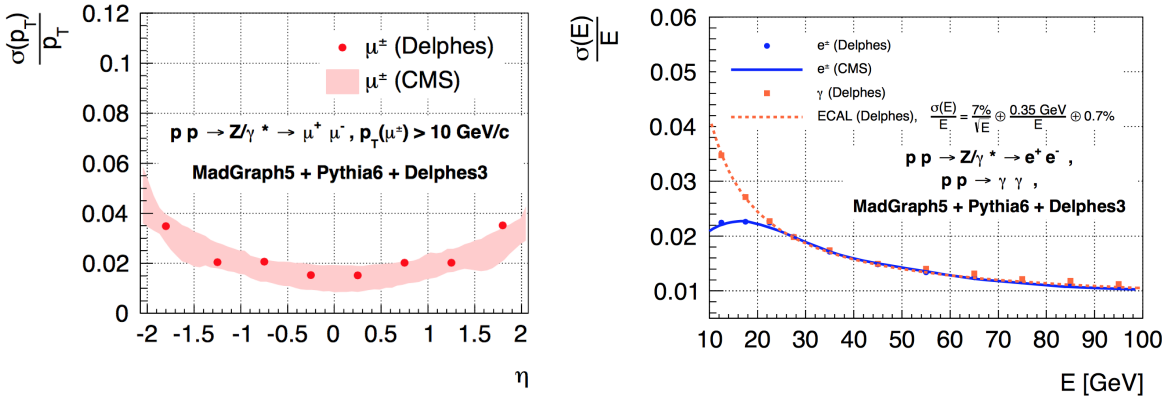

Figure 6.4: Comparison between `CMS Phase-I` simulation and the related `Delphes` tune for muon momentum resolution (left) and electron energy one (right), considering $Z/\gamma^* \to \mu\mu$ and $Z/\gamma^* \to ee$ events, respectively [167].

The validation of jet performances is performed on `QCD` events, comparing `PF+CHS` jet energy resolution for `AK5 jets` as a function of p_T, as reported for jets in the central barrel in Figure 6.5 (left). The agreement is good for jets with $p_T > 30$ GeV, while in the low p_T region some discrepancies are visible.
Moreover, missing energy performance is studied in events with genuine $\not{E}_T$, where neutrinos are produced by fully leptonic $t\bar{t}$ events. Figure 6.5 (right) shows the `PF` missing energy resolution as a function of the $\not{E}_T$ value, where the generator level missing energy is defined as the vector sum of the final state neutrinos.

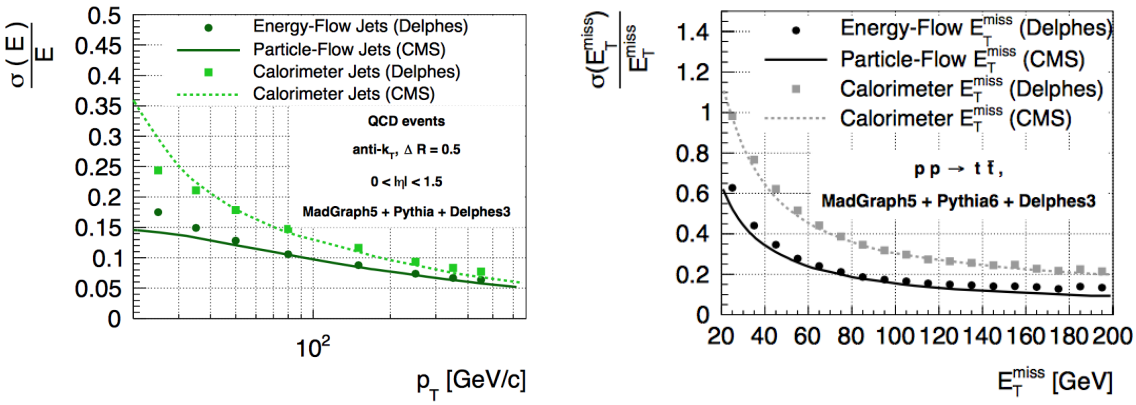

Figure 6.5: Comparison between `CMS Phase-I` simulation and the related `Delphes` tune for jet energy resolution (left) and missing energy one (right), considering `QCD` and fully leptonic $t\bar{t}$ events, respectively [167].

An extensive validation of Delphes Phase-I simulation is performed considering pp collisions at $\sqrt{s} = 14$ TeV with an avergage of 50 pileup interactions per bunch crossing. Two benchmark processes are considered: $gg \to h \to 4\ell$ events produced via POWHEG interfaced with Pythia 6 and the electroweak VBS of same sign W-bosons decaying into a fully leptonic final state ($qq' \to \mathrm{W}^{\pm}\mathrm{W}^{\pm}(\tilde{q}\tilde{q}') \to 2\ell^{\pm}2\nu(\tilde{q}\tilde{q}')$).
Figure 6.6 (top) shows the $h \to \mathrm{ZZ}^* \to 4\mu$ signal efficiency as a function of the different selections adopted in the analysis to define the signal region, comparing the Delphes Phase-I prediction to the official CMS simulation. The, in Figure 6.6 (bottom), the four lepton invariant mass ($m_{4\mu}$) distributions are compared between Delphes and the CMS simulation, normalizing both signal and background (SM $\mathrm{ZZ}^* \to 4\ell$) to an integrated luminosity of 3 ab^{-1}.

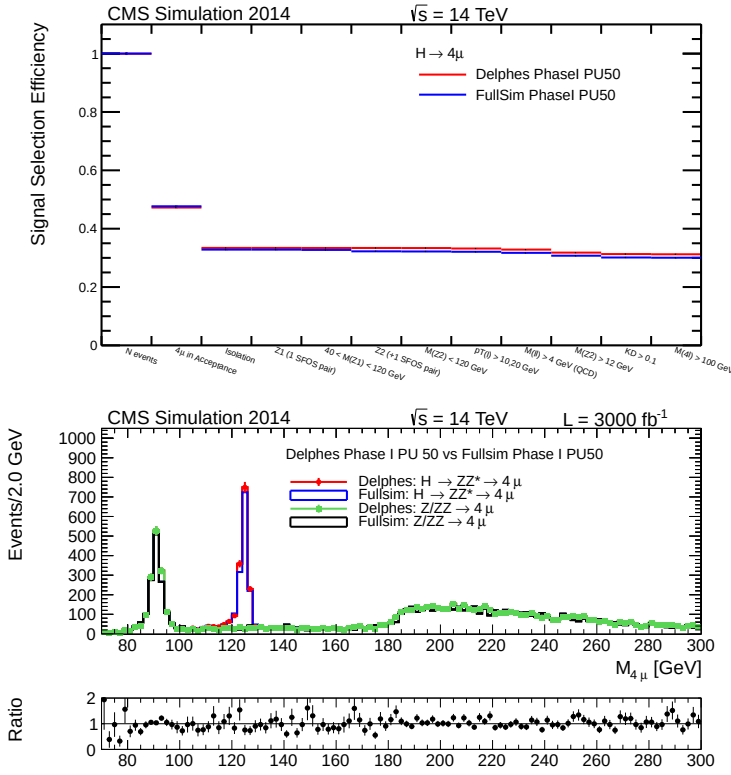

Figure 6.6: (Top) Signal efficiency calculated at each step of $h \to \mathrm{ZZ}^* \to 4\mu$ event selection, comparing CMS simulation with Delphes Phase-I tune. (Bottom) Four muon invariant mass distribution ($m_{4\mu}$) normalized to 3 ab^{-1} for $h \to \mathrm{ZZ}^* \to 4\ell$ (blue line and red dots) and $ZZ \to 4\ell$ background (black line and green dots). Dots refer to Delphes, while solid lines to the CMS simulation at $\langle \mathrm{N_{PU}} \rangle = 50$ interactions.

In the VBS case, CMS full simulated events are produced at LO $\mathrm{O}(\alpha_{ew}^6)$ starting from MadGraph interfaced with Pythia 6. Then, they undergo the GEANT 4 based simulation of the CMS detector. In contrast, the events considered as inputs for the Delphes Phase-I simulation are obtained from Phantom LO generation $O(\alpha_{ew}^6)$ interfaced with Pythia 8
Minimum bias events used to simulate pileup interactions are obtained via Pythia 6 with Tune Z2* and Pythia 8 with Tune CUETP8S1 [171] based on CTEQ6L1 [117], respectively. Besides some differences in both hard scattering and pileup generation, Figure 6.7 (left) shows the absolute selection efficiency, evaluated as a function of typical requirements used to ex-

plore the WW same sign scattering [48], between Delphes Phase-I (dashed red) and full CMS simulation (solid blue) merging di-muon and di-electron final states. Since the signal extraction strategy involves both leptons (μ or e) and VBS jets reconstruction and identification, as well as specific kinematic requirements, the reasonable agreement, around 10%, observed on the "inclusive" signal efficiency proves the reliability of using Delphes Phase-I simulation for feasibility studies in this scenario.
The largest discrepancy is observed when reconstructed jets are matched to generated ones, which is related to a slightly different pileup simulation and jet resolution between full simulation and Delphes.
Figure 6.7 (right) shows the combined efficiency of matching the two leading reconstructed jets with generator level ones as a function of their η seperation ($\Delta\eta_{jj}$). A flat difference of about 5% is visible, except for the high $\Delta\eta_{jj}$ region where Delphes efficiency drops by 15%.

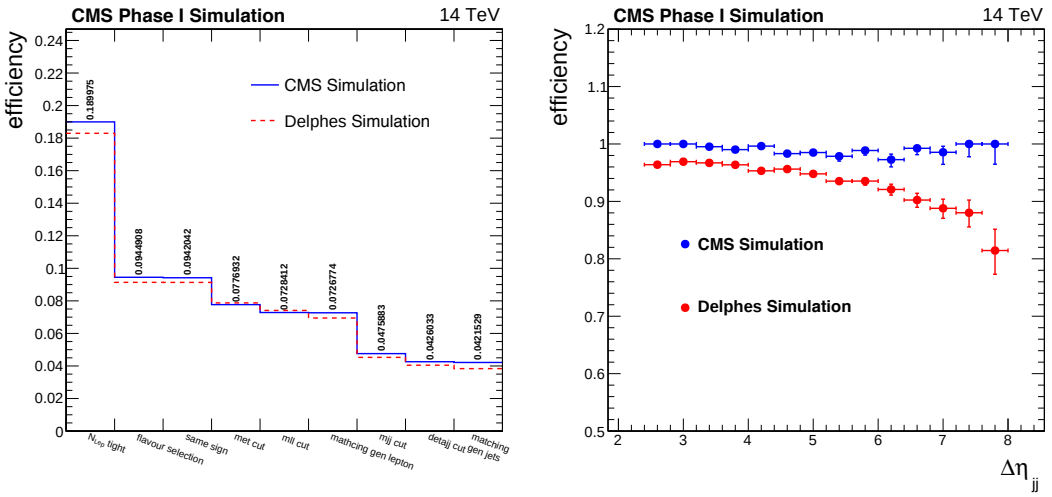

Figure 6.7: (Left) Absolute efficiency evaluated on a dedicated WW same sign fully leptonic scattering sample as a function of typical VBS selections, comparing Delphes Phase-I tune to CMS full simulation in events with two muons or two electrons in the final state. (Right) Efficiency for matching the two leading reconstructed jets with a pair of generator level jets as a function of $\Delta\eta_{jj}$ after the CMS simulation (blue dots) or the Delphes Phase-I approximation (red dots).

Finally, the distributions of the most interesting observables are compared for events surviving VBS-like selections in the di-muon final state. Figure 6.8 shows such comparison for the WW same sign scattering in terms of normalized shapes, where both Delphes Phase-I and CMS simulation are considered. Generator level quantities are displayed together with the corresponding ones after detector simulation and reconstruction. The main differences are located already at generator level, since slightly different setups have been considered.
Anyway, given the scope of the feasibility study described in the next Sections, discrepancies at the level shown here will not alter the outcome of the analysis.

6.3.2 Delphes tune for CMS Phase-II detector

A dedicated Delphes Phase-II scenario has been setup implementing the geometry of an upgraded CMS detector close to the one described in Section 6.2. Considering the same VBS event generation used in the previous Section, the GEANT 4 based CMS Phase-II simulation and reconstruction is compared to a specific Delphes Phase-II tune.

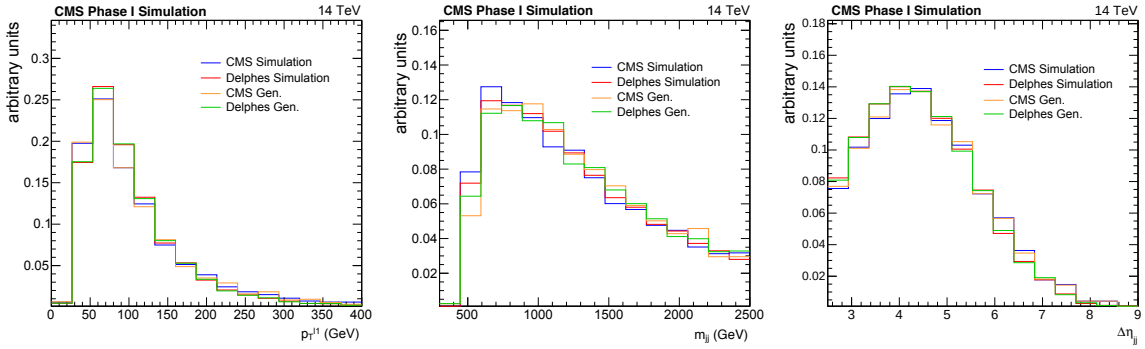

Figure 6.8: Distribution of relevant observables after VBS-like selections showing a comparison between Delphes Phase-I and CMS full simulation for WW same sign scattering events. From left to right: leading lepton p_T, invariant mass of the selected VBS tag-jets (m_{jj}) and $\Delta\eta_{jj}$. Generator level quantities are shown with green (Delphes) and yellow (CMS) solid lines, while reconstructed objects with red (Delphes) and blu (CMS) lines.

Figure 6.9 (left) shows the absolute signal efficiency obtained after applying a set of selections used to isolate a VBS WW same sign enriched region. The two simulations are in good agreement since the efficiencies differ at most by 7%.
In contrast with the Delphes Phase-I tune, differences in the reconstructed VBS tag-jet kinematics are less pronounced since the discrepancy in matching reconstructed to generated jets is limited. In fact, Delphes Phase-II predicts a more correct VBS jet reconstruction efficiency in each bin of $\Delta\eta_{jj}$, as reported in Figure 6.9 (right).

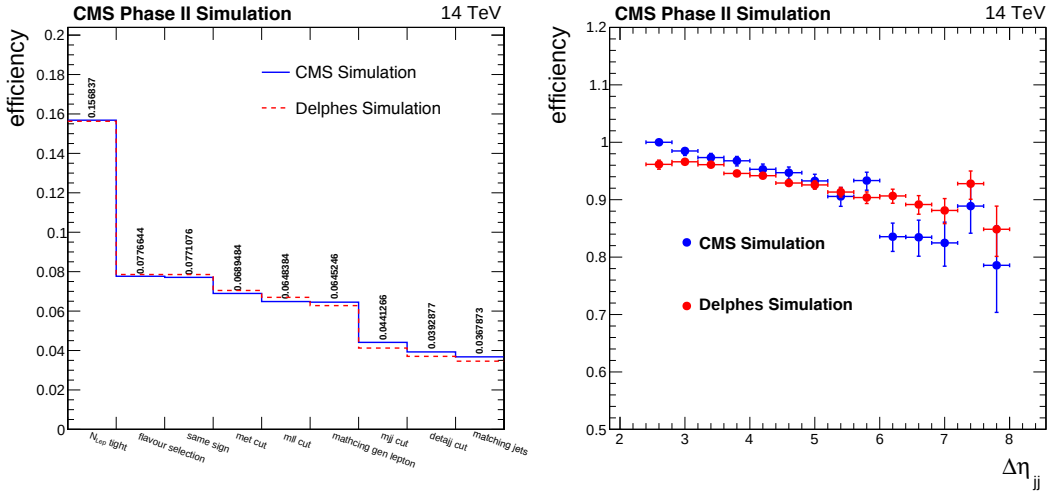

Figure 6.9: (Left) Absolute efficiency evaluated on a dedicated WW same sign fully leptonic scattering sample as a function of typical VBS selections, comparing Delphes Phase-II tune to CMS Phase-II full simulation in events with two muons or two electrons in the final state. (Right) Efficiency for matching the two leading reconstructed jets with a pair of generator level jets as a function of $\Delta\eta_{jj}$ after the full CMS simulation (blue dots) or the Delphes Phase-II approximation (red dots).

The reliability of Delphes Phase-II tune in correctly describing the kinematics of VBS events is proved by Figure 6.10, where normalized distributions of the leading lepton p_T, m_{jj} and $\Delta\eta_{jj}$ of the VBS tag-jets are compared between the CMS Phase-II simulation and the Delphes one.

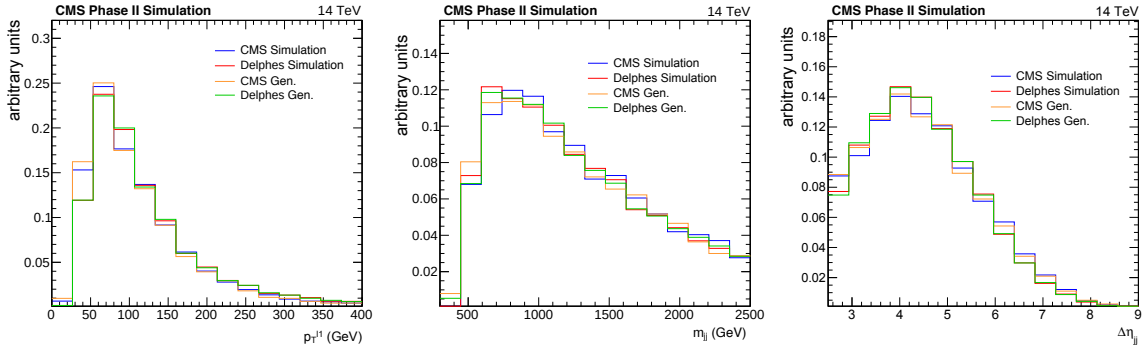

Figure 6.10: Distribution of relevant observables after VBS-like selections, showing a comparison between Delphes Phase-II and CMS full simulation for the electroweak WW same sign scattering. From left to right: leading lepton p_T, invariant mass of the selected VBS tag-jets (m_{jj}) and $\Delta\eta_{jj}$. Generator level quantities are shown with green (Delphes) and yellow (CMS) solid lines, while reconstructed objects with red (Delphes) and blu (CMS) lines.

Eventually, no Delphes tune for the CMS Phase-I Aged scenario has been studied yet, thus, in the next Sections, the performance expectations for this case are obtained from the Phase-II results re-scaling the signal yields by the relative loss in the selection efficiency measured on CMS full-sim samples.

6.4 Vector Boson Scattering at the HL-LHC

In Section 1.6, a brief introduction about the role of the VBS for testing the EWSB sector of the SM was given. Here, a dedicated study on the electroweak scattering of massive vector bosons is reported in view of the HL-LHC data taking.

In pp collisions, VBS can only be detected in the topology $pp \to \mathrm{VV}\,jj$, where vector bosons are irradiated by a pair of incoming quarks, which get deflected originating two hard jets in the final state referred to as VBS tag-jets. The electroweak scattering interaction happens through a variety of possible mechanisms, $\mathrm{O}(\alpha_{ew}^6)$ at LO, which reflect the SM gauge structure including triple and quartic gauge couplings (TGC,QGC), t-channel Higgs boson exchange, s-channel Higgs production, also known as VBF production, as well as a variety of non-scattering processes producing the same final state like $t\bar{t}$, single-top, Vγ, etc.

Individually the amplitude of the electroweak scattering interactions rises quickly with the energy, violating unitarity at the TeV scale. The effect of mutual strong interferences results in a finite total cross section. Thus, the scattering topology is particularly sensitive to new physics inside the EWSB sector, since any additional diagram or difference with respect to the SM coupling structure would alter this cancellation at high scattering center of mass energy ($\sqrt{\hat{s}}$). This is detectable through observables sensitive to $\sqrt{\hat{s}}$, such as the di-boson invariant mass, the vector bosons transverse momenta and the VBS jet kinematics. In parallel, angular correlations between the scattering products are sensitive to the polarization of the outgoing vector bosons, which may give fundamental information about the new physics involved.

Finally, di-boson signatures are also explored for direct searches of new resonances, in particular looking for massive neutrals, singly or doubly charged Higgs bosons predicted in BSM models like 2HDM or Higgs-Triplet-Model [172, 173].

The scattering phenomenology is characterized by the properties of the VBS tag-jets, which

are highly energetic, with a large η separation ($\Delta\eta_{jj}$) and invariant mass (m_{jj}). The two reconstructed jets with the highest transverse momentum are usually adopted to identify the fragmentation products of the deflected quarks.
Because of the scattering topology, the η coordinates of the two tag jets (η_1, η_2) determine a relevant metric in the analysis, since a limited activity is expected in the geometrical region between the VBS jets due to the absence of color flow among the vector boson decay products [174].
In this study, only fully leptonic decays of the vector bosons are considered to select signatures with an expected low background contamination, avoiding also ambiguities in the separation between tag-jets and jets produced by hadronically decaying vector bosons. In particular, two alternative final states are explored:

- **Same sign WW scattering** ($pp \to \mathrm{W}^{\pm}\mathrm{W}^{\pm}\,jj \to 2\ell^{\pm}\,jj$), where the two vector bosons decay into two same charged leptons (μ or e), either with the same flavor (SF) or a different one (OF). Backgrounds come from irreducible sources, as same sign production of W-bosons from $\mathrm{O}(\alpha_{ew}^4\alpha_s^2)$ processes and WZ+jets events at $\mathrm{O}(\alpha_{ew}^4\alpha_s^2)$ or $\mathrm{O}(\alpha_{ew}^6)$, when a lepton from $\mathrm{Z} \to \ell\ell$ is out of acceptance or not reconstructed.
  In addition, reducible backgrounds are represented by semi-leptonic $t\bar{t}$ and leptonic W+jets events when one jet is mis-identified as a lepton. Fully leptonic decays of opposite sign WW pairs, $t\bar{t}$ and DY+jets ($Z/\gamma^* \to \ell\ell$) events contribute as backgrounds when the charge of one of the two leptons is wrongly reconstructed. Finally, this final state is expected to be particularly sensitive to the scattering of longitudinally polarized vector bosons, since the contaminations to the pure $\mathrm{W}_{\mathrm{L}}^{\pm}\mathrm{W}_{\mathrm{L}}^{\pm} \to \mathrm{W}_{L}^{\pm}\mathrm{W}_{L}^{\pm}$ component arising alternative spin correlations ($\mathrm{W}_{\mathrm{T}}^{\pm}\mathrm{W}_{\mathrm{L/T}}^{\pm} \to \mathrm{W}_{\mathrm{L}}^{\pm}\mathrm{W}_{\mathrm{L}}^{\pm}$) and non scattering diagrams are expected to be particularly small [175].

- **WZ boson scattering** ($pp \to \mathrm{W}^{\pm}\mathrm{Z}\,jj \to 3\ell\nu\,jj$), where the final state is characterized by three leptons in the detector, two of them with the same flavor, opposite charge and invariant mass compatible with a $\mathrm{Z} \to \ell\ell$ decay. Besides the major background represented by irreducible $pp \to \mathrm{W}^{\pm}\mathrm{Z}\,jj$ $O(\alpha_{ew}^4\alpha_s^2)$ production, Drell-Yan lepton pairs produced in association with jets and fully leptonic $t\bar{t}$ contaminate the signal region when one jet is mis-identified as a lepton.
  Then, ZZ pair production arising from both $O(\alpha_{ew}^6)$ and $O(\alpha_{ew}^4\alpha_s^2)$ interactions shows the same signature of the signal when one lepton is outside the detector acceptance or is not reconstructed. Even if the signal-to-background ratio disfavors this channel with respect to the WW same sign one, a kinematic constraint allows the reconstruction of the full event kinematics, adding separation power once the angular correlations between decay products are exploited.

The cross sections for both same sign WW and WZ electroweak scattering are predicted with NLO accuracy [176, 177].
Studies of same sign W-boson scattering have been performed by both ATLAS [47] and CMS [48] experiments with Run-I data, while only prospects for the observation of the electroweak WZ scattering, as well as aQGC sensitivity, were previously studied by CMS [178].

6.4.1 Analysis benchmarks

The determination of the VBS cross section is an independent confirmation of the SM nature of the Higgs boson recently discovered, where any deviations from the expected predictions are indications of new BSM phenomena. To assess the sensitivity of the upgraded CMS experiment to the VBS, several benchmarks are considered:

- **Inclusive scattering cross section:** the expected uncertainty on the inclusive electroweak scattering cross section is estimated for both same sign WW and WZ bosons channels.
- **Longitudinal scattering component:** looking at the VBS kinematics and topology, a set of observables sensitive to final states with two longitudinally polarized vector bosons is used to predict the expected sensitivity on the longitudinal component in both WW and WZ bosons channels.
- **Anomalous QGC measurement:** the SM Lagrangian can be modified through the addition of high order operators predicted in an EFT framework (8-dimensional operator for aQGC), as described in Section 1.6.2. Selecting an observable sensitive to the anomalies, same sign WW events are fitted to a BSM model parametric in the additional operators.
- **Partial unitarization scenario:** different BSM models predict the possibility that the spontaneous breaking of the electroweak symmetry is realized via a mechanism which involves more than one Higgs boson, presenting a phenomenology in which the discovered Higgs only partially fulfills to the unitarization of the VBS. Assuming that any of these new resonances will appear along the same sign WW bosons spectrum, the VBS measurements would show an intermediate behaviour between the SM case, where the EWSB is granted by the Higgs boson, and a Higgs-less scenario.

6.4.2 Simulated samples

Signal samples have been produced with LO generators $O(\alpha_{ew}^6)$. In particular, Phantom [160], able to generate events with two partons in the initial and six objects in the final state through an exact LO calculation in α_{ew}^6 or $\alpha_{ew}^4\alpha_s^2$, is adopted to simulate both signals, same sign WW and WZ bosons scattering at $O(\alpha_{ew}^6)$, and irreducible backgrounds (ZZ scattering, same sign WW+jets and WZ+jets at $O(\alpha_{ew}^4\alpha_s^2)$, fully leptonic $t\bar{t}$ as well).
Phantom is also used to generate partial unitarized scenarios in which the SM Higgs boson partially contributes to the unitarization, through rescaling its couplings to vector bosons.
In contrast, MadGraph is used to produce same sign WW and electroweak WZ bosons events up to un-decayed V-bosons providing, as additional information, their polarization state V-bosons. To preserve the polarization information, vector bosons are decayed using the Decay package instead of MadSpin. This represents an approximation since the interference is calculated up to the V-boson level, instead of considering the stable particles produced in the final state.
When signal or irreducible background events are generated via Phantom, a set of preselections are applied on generator level particles (quarks, leptons and neutrinos), as listed in Table 6.1. For MadGraph generation a set of similar selections are considered too.
Figure 6.11 shows a comparison, performed considering matrix element particles, between MadGraph and the complete Phantom calculation for the same sign $W^{\pm}W^{\pm}$ scattering after

applying the preselections listed in Table 6.1. The agreement is found to be reasonable, within 10%, for both the same sign WW and the WZ bosons electroweak scattering, where the Wγ contribution under the Z-pole and its related interference should be added to the event generation.

Observable	Selection
minimum jet p_T	20 GeV
maximum jet η	6.5
minimum jet energy	20 GeV
minimum lepton p_T	20 GeV
maximum lepton η	4.0
minimum di-boson invariant mass (m_{VV})	130 GeV
minimum di-jet invariant mass (m_{jj})	300 GeV
minimum di-jet rapidity difference ($\Delta\eta_{jj}$)	2
minimum di-lepton invariant mass ($m_{\ell\ell}$)[a]	4 GeV

Table 6.1: List of selections applied during `Phantom` generation for signals and irreducible backgrounds.

[a] Applied only in the WZ and ZZ bosons generations.

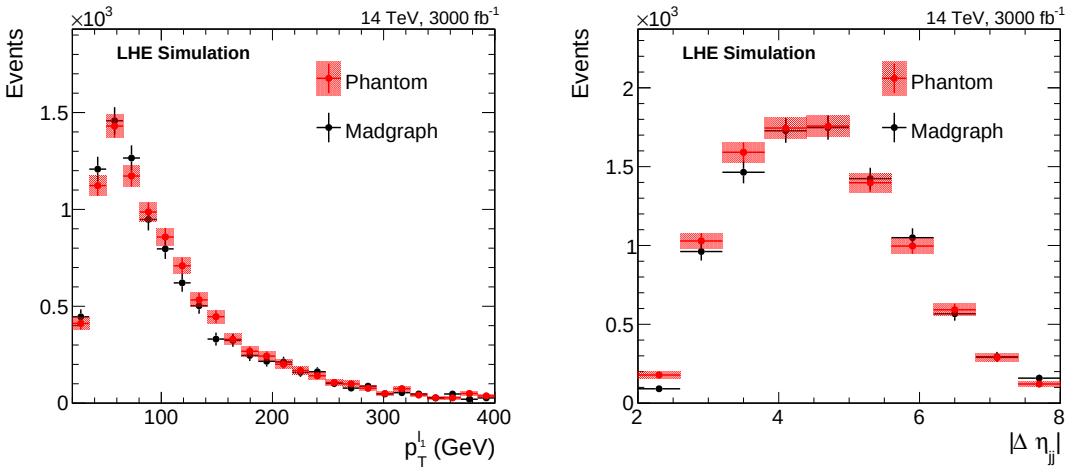

Figure 6.11: Comparison between `Phantom` and `MadGraph` predictions at matrix element obtained after applying the selections listed in Table 6.1 for the same sign WW scattering: (left) leading lepton p_T, (right) $\Delta\eta_{jj}$.

The gauge bosons anomalous couplings are parametrized using the effective field theory approach described in Ref. [179], as implemented in `MadGraph` that calculates matrix element coefficients used to weight each event of a single generation to reproduce different settings of the `EFT` operators.

Semi-leptonic $t\bar{t}$ events are generated with `POWHEG` at `NLO QCD`, while W+jets and `DY`+jets samples, up to three jets at matrix element, are generated via `MadGraph`. Finally, electroweak $\mathrm{O}(\alpha_{ew}^6)$ and `QCD` $\mathrm{O}(\alpha_{ew}^4\alpha_s^2)$ opposite sign WW+2 jets events are simulated with `Phantom`, which are used to evaluate the mis-charge contribution from fully leptonic $t\bar{t}$ and $\mathrm{W}^+\mathrm{W}^-$ productions in the same sign analysis. `MadGraph` and `Phantom` generations adopt `CTEQ6L1` [117] as `PDF`, while `POWHEG` uses `CT10` [119].

A summary of all the generated samples with the related cross sections are listed in Table 6.2. In both signal and background simulations, only decays into muons or electrons are allowed, thus exclusive decays of the vector bosons to τ-leptons are discarded. The contribution from leptonic τ decays is expected to be small, altering the selected signal and background rates by few percent without changing the results of this feasibility study.
The interference between electroweak $O(\alpha_{ew}^6)$ and QCD $O(\alpha_{ew}^4\alpha_s^2)$ processes shows a small impact, around 5% of the signal cross section, for both same sign WW and WZ boson cases after applying the preselections listed in Table 6.1. For this reason, to have a more efficient event production, electroweak and QCD events are generated independently negleting mutual interference effects.
Eventually, all the generated events are then interfaced with Pythia8 [115] for parton showering, hadronization and underlying event simulation.

Process	Generator	Cross Section (fb)
WW_{ewk} SS $O(\alpha_{ew}^6)$ $\kappa_V^2 = 1$ [a]	Phantom	8.252 ± 0.004
WW_{ewk} SS $O(\alpha_{ew}^6)$ $\kappa_V^2 = 0.9$ [a]	Phantom	8.288 ± 0.004
WW_{ewk} SS $O(\alpha_{ew}^6)$ $\kappa_V^2 = 0.8$ [a]	Phantom	8.331 ± 0.004
WW_{ewk} SS $O(\alpha_{ew}^6)$ $\kappa_V^2 = 0.7$ [a]	Phantom	8.404 ± 0.004
WW_{ewk} SS $O(\alpha_{ew}^6)$ $\kappa_V^2 = 0.6$ [a]	Phantom	8.457 ± 0.004
WW_{ewk} SS $O(\alpha_{ew}^6)$ $\kappa_V^2 = 0.5$ [a]	Phantom	8.528 ± 0.004
WW_{ewk} SS $O(\alpha_{ew}^6)$ $\kappa_V^2 = 0$ [a]	Phantom	8.984 ± 0.004
WW_{QCD} SS $O(\alpha_{ew}^4\alpha_s^2)$	Phantom	2.126 ± 0.004
WW_{ewk} OS $O(\alpha_{ew}^6)$ $\kappa_V^2 = 1$	Phantom	57.88 ± 0.03
WW_{QCD} OS $O(\alpha_{ew}^4\alpha_s^2)$	Phantom	1419.2 ± 0.7
WZ_{ewk} $O(\alpha_{ew}^6)$ $\kappa_V^2 = 1$	Phantom	7.848 ± 0.004
WZ_{ewk} $O(\alpha_{ew}^6)$ $\kappa_V^2 = 0$	Phantom	8.044 ± 0.004
WZ_{QCD} $O(\alpha_{ew}^6\alpha_s^2)$	Phantom	32.52 ± 0.02
ZZ_{ewk} $O(\alpha_{ew}^6)$ $\kappa_V^2 = 1$	Phantom	0.281 ± 0.001
ZZ_{QCD} $O(\alpha_{ew}^4\alpha_s^2)$	Phantom	1.265 ± 0.001
WW_{ewk} SS $O(\alpha_{ew}^6)$ $\kappa_V^2 = 1$ EFT	Madgraph	11.24
WW_{ewk} SS $O(\alpha_{ew}^6)$ $\kappa_V^2 = 1$ Polarized	Madgraph	11.24
WZ_{ewk} $O(\alpha_{ew}^6)$ $\kappa_V^2 = 1$ Polarized	Madgraph	8.71
$t\bar{t} \to WWb\bar{b} \to \ell\nu\, jj\, b\bar{b}$	POWHEG [b]	$(242.5 \pm 0.2)10^3$
W+3 jets [c]	Madgraph	$6.35 \cdot 10^3$
DY+2 jets [c]	Madgraph	$28.82 \cdot 10^3$
Z+3 jets [c]	Madgraph	$0.61 \cdot 10^3$

Table 6.2: List of the simulated samples used for the analysis together with the generator and the corresponding cross section in fb. Specific selections are applied in each generation to enriched the analysis phase space. The uncertainty on the cross section, when indicated, corresponds to the the accuracy of the generator for a chosen configuration of scales and PDF. SS stands for same sign lepton pairs, OS for opposite sign.

[a] κ_V^2 is the strength of the Higgs couplings to vector bosons with respect to the SM prediction.
[b] Preselections listed in table 6.1 not applied in this case.
[c] Specific preselections are adopted to generate these background samples, to boost their acceptance efficiency in the analysis phase space.

6.5 Event reconstruction

Based on the studies performed for the HL-LHC [166], events are assumed to be collected by double-lepton triggers which require thesholds on the lepton transverse momentum of 20 and 10 GeV, respectively, assuming a trigger efficiency close to 100%. No corrections for trigger inefficiencies are applied on simulated events as a theshold of $p_T > 20$ GeV is required on all the reconstructed leptons in the final state (e or μ).
The CMS full simulated samples are used to parametrize efficiencies and resolution of leptons, jets, missing energy as well as b-jet tagging efficiency, which are implemented in Delphes [167] for each different detector scenario, as described in Section 6.3.1 and Section 6.3.2.
Leptons coming from electroweak interactions are expected to be isolated from additional hadronic or electromagnetic activity in the detector, while leptons coming from QCD sources, such as semi-leptonic b-quark decays and fake leptons (jets mis-identified as leptons), are often accompanied by charged and neutral particles. In Delphes, both muons and electrons are required to be isolated according to Equation 6.4, where the isolation calculated considering $R = 0.3$ is required to be less than 0.25 (0.6) in case of 50 (140) in-time pileup scenario.
To veto additional leptons with respect to the ones expected from the decay of WW or WZ boson pairs, "loose" leptons are defined adopting the same identification working point but relaxing the isolation requirement to 0.35 (0.75). Combining the identification and the isolation criteria, the selection efficiency for muons (electrons) is stable around 80% (75%) for $p_T > 20$ GeV and $|\eta| < 1.5$. Then, it drops in the endcap region up to 65% (55%) at $|\eta| = 2.5$. In the Phase-II scenario, these efficiencies propagate quite stably in the extended η coveraged region up to $|\eta| = 4$.
Jet reconstruction starts from all the Delphes PF candidates removing the charged ones that are not associated to the primary vertex, partially mitigating pileup effects in the tracker covered region up to $|\eta| = 2.5$ (4.0) for the Phase-I (Phase-II) detector. Remaining particles are clustered via the anti-k_t algorithm with $R = 0.4$. To mitigate residual pileup effects from photons and neutral hadrons, a jet median area subtraction [85] is applied independently in three η bins: $|\eta| < 2.5$, $2.5 < |\eta| < 3.0$ and $|\eta| > 3.0$.
After this, only jets with $p_T > 30$ GeV, $|\eta| < 4.7$ and not in geometrical overlap to identified leptons inside a cone of $\Delta R = 0.3$ are considered in the analysis. To identify jets produced by the hadronization of b-quarks, the properties of the Combined Secondary Vertex (CSV) algorithm [125] are parametrized in Delphes, considering a working point that provides about 70% efficiency for b-quark jets and a mistag rate of 1% for light flavours and gluon ones.

6.6 Mis-identified lepton background

A major source of background in fully leptonic VBS searches is represented by the mis-identification of jets as leptons (e or μ). In the same sign WW analysis, this is mostly due to semi-leptonic $t\bar{t}$ production ($t\bar{t} \to b\bar{b}\,\mathrm{W}^+\mathrm{W}^- \to b\bar{b}\,\ell\nu jj$), where b-quark jets are not identified by the CSV algorithm. Each jets produced by the showering of the matrix element event could be mis-identified as a lepton. A smaller contribution is also coming from the W-boson production in association with at least three jets (W+jets).
In contrast, in the WZ boson analysis this could happen when a Z-boson is produced in association with at least three jets or in fully leptonic $t\bar{t}$ pairs in association with at least one jet,

when two leptons in the event have same flavor, opposite charge and invariant mass compatible with the Z-boson one.
To quantify this background, the probability for a jet with a given kinematics (p_T,η) and flavor (f) to be identified as a lepton (e or μ) with (p'_T,η') has to be estimated (fake lepton probability). The mis-identified lepton kinematics is not necessarily the same of the jet one, as the lepton transverse momentum p'_T is usually expected to be lower of than the jet one. This poses the problem of the universality of the fake rate, since it introduces a dependence on the jet p_T spectrum used for the mis-identification probability measurement.
The fake rate is measured starting from a high statistics sample, showing a similar kinematics to the real fake lepton background properties expected in the VBS analysis, through a full scan of η and p_T phase space. As a baseline, a large inclusive $t\bar{t}$ sample, simulated at a center-of-mass energy $\sqrt{s} = 13$ TeV with an average of 20 in-time pileup interactions, is considered. Events are selected requiring the presence of at least one identified and isolated lepton (μ or e), called tag-lepton (ℓ_{tag}), which should be also matched to a generator level lepton produced by a W-boson decay. Jets overlapping to the tag-lepton within a cone of $\Delta R = 0.3$ are discarded. At the same time, the remaining jets are required to come from the primary interaction by matching reconstructed AK4 jets with the ones obtained by clustering generator level particles, to assess to the jet parton flavor information. In this way, jets from pileup interactions, for which is not possible to associate a flavor from generator level information, are not considered in the measurement of the mis-identification probability.
The "fake rate" is defined as follows:

$$\mathscr{P}(p_T^i, \eta_k, f_m) = \frac{N(\ell_{tag}, p_T^i(j,\text{M}), \eta_k(j,\text{M}), f_m(j,\text{M}))}{N(\ell_{tag}, p_T^i(j), \eta_k(j), f_m(j))} \tag{6.5}$$

where $N(\ell_{tag}, p_T^i(j), \eta_k(j), f_m(j))$ is the number of times in which, in an event with a tag-lepton, a reconstructed jet with (p_T,η) and associated parton flavor (f) falls in the i-th, k-th and m-th bin, respectively, while $N(\ell_{tag}, p_T^i(j,\text{M}), \eta_k(j,\text{M}), f_m(j,\text{M}))$ is the number of jets belonging to the same bin matched to an identified lepton within a cone of $\Delta R = 0.3$.
This mis-tag probability is measured independently for two jet flavors, b-jets and non b-jets (light quark or gluon jets) as they show different properties due to the possibility for a b-quark to decay semi-leptonically ($b \to c\ \ell\nu$) during its fragmentation. In addition, jets from b-quarks are also more collimated on average than light flavor or gluon jets due to the different structure of the QCD radiation emitted by the initial quark.
Figure 6.12 (top) shows the measurement of the probability in bins of ($|\eta|$, p_T) for a b-jet to be identified as a "tight" muon, which is of the order of 5×10^{-4}. The corresponding projections are reported in the bottom part of Figure 6.12. The bulk of the probability for a b-jet to be identified as a muon arises when the b-quark decays semi-leptonically. Since the jet is more collimated for large transverse momentum, the fake rate decreases as a function of the jet p_T because the muon becomes less isolated. At the same time, the fake rate increases along $|\eta|$ moving from the barrel towards the endcap region.
The same trends are observed measuring the b-jet fake rate to electrons, where the magnitude of the effect is a factor two weaker than the muon case thanks to ECAL cluster shapes and H/E requirements that are more agressive in rejecting electrons produced inside jets.
Figure 6.13 shows the (η,p_T) dependent measurement of the light-quark and gluon jets fake rate into electrons. Here, the mis-identification probability raises as a function of the jet

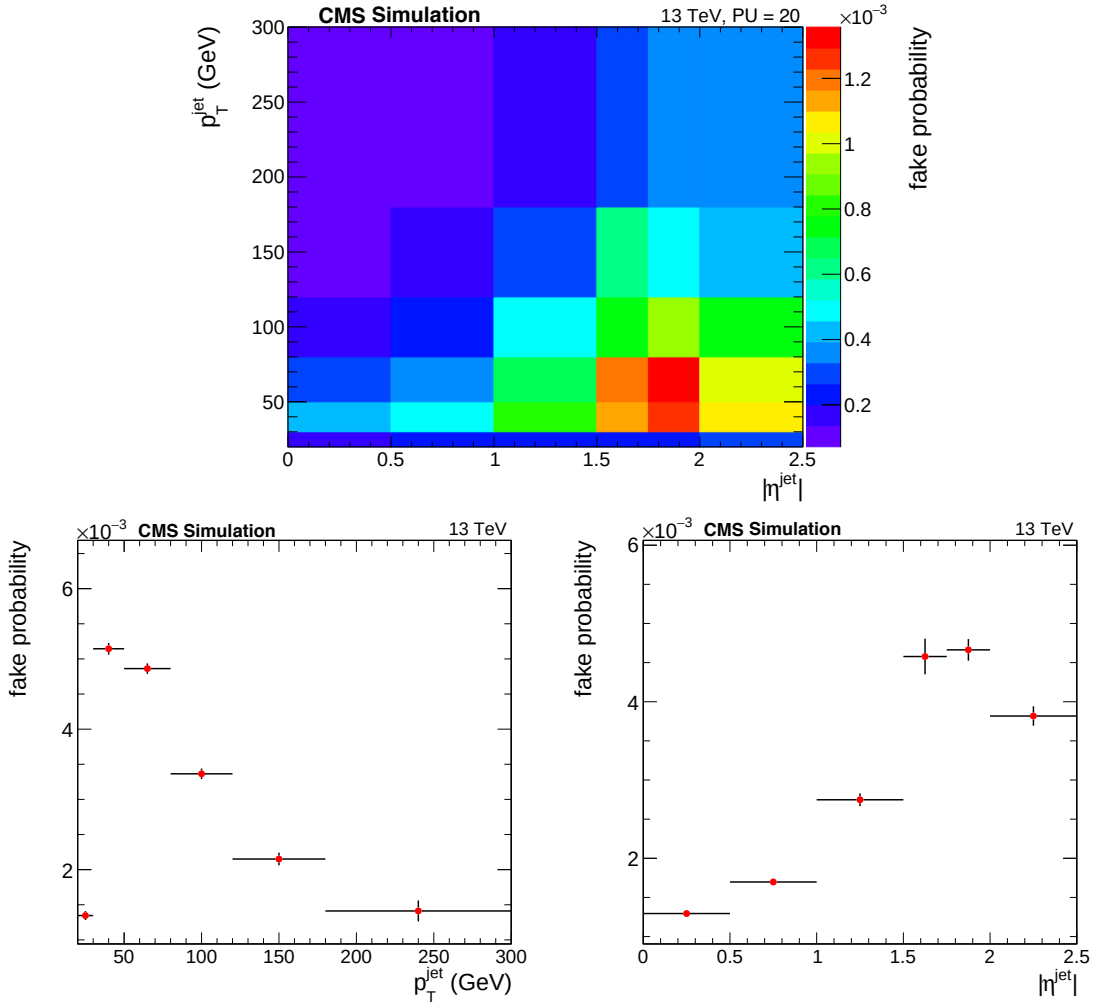

Figure 6.12: Probability to mis-identify a jet, originating from the fragmentation of a b-quark, as a muon as a function of the reconstructed jet η and p_T (top). The projection of the two-dimensional map along p_T ($|\eta|$) is shown in the the bottom left (bottom right) plot.

transverse momentum since, even if the electron track candidate appears to be less isolated at higher p_T, the properties of ECAL and HCAL deposits become more electron-like. In addition, the fake rate is stable in the barrel region, around 3.5×10^{-4}, while it decreases in the endcap one. Similar properties are also shown by the fake rate of the light-quark and gluon jets into muons, which is around ten times smaller than the electron one.
Since the probability for a b-jet to mis-identified as a lepton (e or μ) is larger than the one for non b-jets, the semi-leptonic $t\bar{t}$ background is expected to be the largest contribution to the fake lepton background in the same sign WW analysis, while the W+jets contamination is expected to be more important in final states with electrons.
The direction of the matched jet and the fake lepton are checked to be almost identical, while this, as expected, is not the case for the transverse momentum. For this reason, the p_T migration between the jet and the fake lepton has been evaluated as a correlation matrix:

$$M_f(i,k) = \frac{p_T^i(\ell)}{p_T^k(j)} \quad \eta_\ell = \eta_j \quad f = \text{b-jet , non b-jet} \tag{6.6}$$

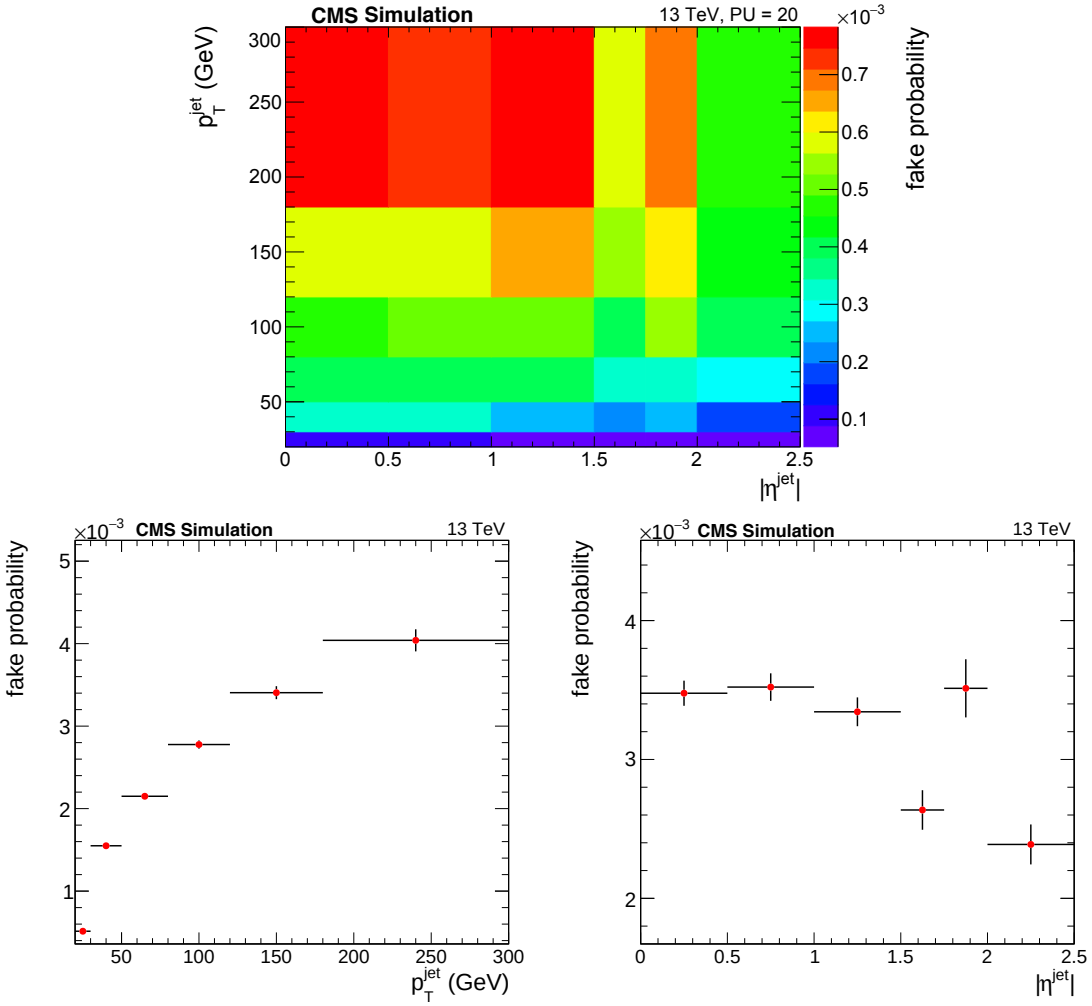

Figure 6.13: Probability to mis-identify a jet, originating from the fragmentation of a light-quark or a gluon, as an electron as a function of the reconstructed jet η and p_T (top). The projection of the two-dimensional map along p_T ($|\eta|$) is shown in the bottom left (bottom right) plot.

Figure 6.14 shows the jet-to-lepton p_T migration as determined in the $t\bar{t}$ sample for b-jets mis-identified as muons (left) and light-flavor quark/gluon jets to electrons (right).
As a second step, the fake rates and the migration matrixes are applied to all jets, before any selection, in $t\bar{t}$, W+jets and Z+jets events according to their kinematics and flavor to predict the expected fake lepton background in the VBS analyses.
To validate these values and estimate their possible dependence with $\sqrt{s}$ and the detector or the reconstruction scenario, the same strategy is applied on a large $t\bar{t}$ sample produced at $\sqrt{s} = 8$ TeV, with similar pileup conditions as in the 13 TeV one.
The fake rate trends as a function of jet p_T and η are in agreement between the two cases, while the normalizations are within at most a factor two, which can be explained by the different lepton reconstruction and identification strategies adopted in the two scenarios. This suggests that the fake rate estimate does not strongly depend on $\sqrt{s}$ and that the values obtained at 13 TeV can be used in the analysis, without a dedicated extrapolation to 14 TeV.
In addition, to further check if the fake rate estimate can be used also for the different up-

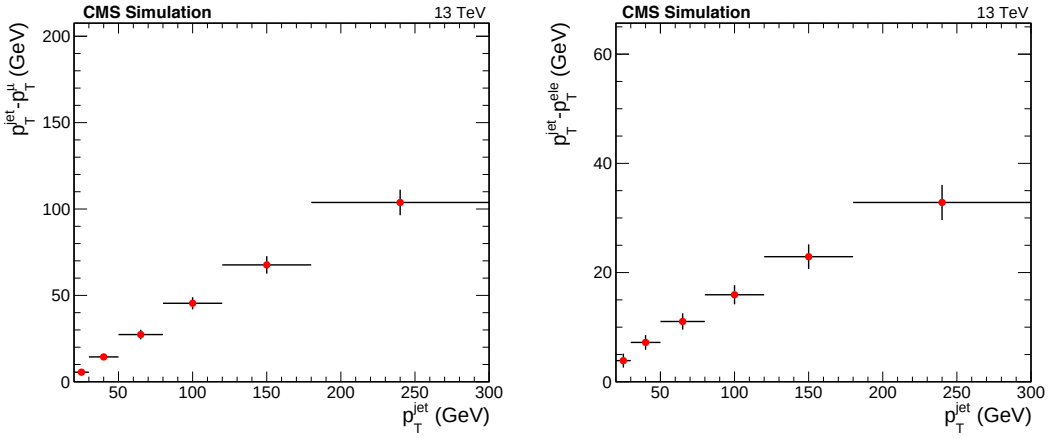

Figure 6.14: Profile of the p_T migration matrix obtained from $t\bar{t}$ events for b-jets to muons (left) and non b-jets to electrons (right). The error bars are set as the uncertainty on the mean value of the difference between the lepton and the jet p_T, where empty bins are discarded.

grade scenarios, the same quantities were calculated using smaller $t\bar{t}$ productions for `Phase-I` (`Phase-II`) detector at 50 (140) pileup interactions in larger bins of p_T and η. Also in these cases, shapes are found to be in agreement within the large statistical uncertainties. Thus, it looks reasonable to use a common estimate for the fake rate performed at 13 TeV, where the largest `MC` sample was available to minimize statistical uncertainties.
Finally, since discrepancies on the fake rate normalization have been observed during `Run-I` between data and simulation, results will be presented as a function of a global fake rate scale factor to account for an unknown data-to-simulation shift, assuming to have at least a correct prediction for the fake lepton background shape.

6.7 The same sign $W^{\pm}W^{\pm}$ bosons scattering

To select with the best efficiency fully leptonic same sign WW bosons scattering events, reducing the background contamination as well as enhancing the electroweak scattering component relatively to the `QCD` $O(\alpha_{ew}^4\alpha_s^2)$ one, a dedicated set of selections is optimized.
First, two same-sign identified and isolated leptons, either electrons or muons, with $p_T > 20$ GeV are required to be present in the detector. Additional loose leptons in the event are vetoed to reduce $pp \to WZjj \to 3\ell\nu jj$ contamination, according to a looser isolation criterion introduced in Section 6.5. This background, together with events due to Z/γ^*+jets associate production, is further reduced by a Z-boson mass veto, requiring $m_{\ell\ell}$ to be outside the window $m_Z \pm 10$ GeV for the same flavor final state.
To suppress the possible presence of `QCD` multi-jet background arising from double fake leptons, the di-lepton mass ($m_{\ell\ell}$) has to be larger than 40 GeV, as well as the missing energy ($\not{E}_T$), which are also useful to reduce W+jets and Z/γ^*+jets contributions, respectively.
Furthermore, a sizable reduction of the same sign WW $O(\alpha_{ew}^4\alpha_s^2)$ background is achieved requiring the η separation between the two identified leptons to be less than 2 units ($\Delta\eta_{\ell\ell}$).
To suppress the $t\bar{t}$ contribution, no jet with $p_T > 30$ GeV should be identified as coming from a b-quark by the `CSV` algorithm, as already described in Section 6.5. Events are also discarded if a muon with $p_T > 5$ GeV is found inside a reconstructed jet, with $p_T > 20$ GeV, within $\Delta R < 0.4$. Considering `AK4` jets with $p_T > 1$ GeV obtained clustering only the tracks associ-

ated to the primary vertex, except for the identified leptons, the scalar sum $H_T = \sum_j p_T^j$ is adopted as an estimator of the total hadronic activity of the event, showing a good separation against the $t\bar{t}$ production. The H_T value is required to be less than 125 (150) GeV for `Phase-I` (`Phase-II`) scenario, according to the tracker coverage along η.
Then, at least two `CHS` jets with $p_T > 30$ GeV are required, where the two leading p_T one are selected as `VBS` tag-jets. Their pseudorapidity separation ($\Delta\eta_{jj}$) has to be larger than 2.5 units, while the invariant mass of the tag-jet system (m_{jj}) should be larger than 850 GeV. In contrast, since events with longitudinally polarized vector bosons show lower lepton and jet transverse momenta, as well as lower $\Delta\eta_{jj}$ and m_{jj}, a relaxed selection $m_{jj} > 650$ GeV allows to optimize the sensitivity to the longitudinal scattering, estimated as $S/\sqrt{S+B}$ where S and B are the expected signal and background yields surviving the analysis selections, respectively. Eventually, the distance between the di-lepton and the tag-jet systems $\Delta R(\ell\ell, jj)$ should be less than 6 units, as well as the leptons are required to be within the tag-jets along the η direction through a selection on the Zeppenfeld variable ($Z_\ell < 0.5$), which is defined as follows:

$$Z_\ell = \frac{1}{\Delta\eta_{jj}}\left(\eta_\ell - \frac{\eta_{j,1} + \eta_{j,2}}{2}\right) \tag{6.7}$$

To increase the total sensitivity of the analysis, selected events are categorized depending on both lepton flavor ($\mu\mu$, μe, $e\mu$ and ee), where leptons are sorted in p_T, and the final state charge ($++$ or $--$) yielding to eight independent categories. The reason of such classification is related to a different background composition among the flavor categories, for example the fake lepton probability is not the same between muons and electrons, and to take advantage of a charge asymmetry in the signal production. In fact, since the `LHC` is a pp collider, the cross section of the same sign W^+W^+ production is larger than W^-W^- one of about a factor seven, while the probability for a jet to be mis-identified as a positive or a negative charged lepton are consistent within the systematic uncertainty of the fake rate method. A summary of the selections and the event categorizations adopted in this analysis is reported in Table 6.3.
Figure 6.15 shows examples of distributions for several observables obtained merging all the event categories after the full analysis selection, considering the `CMS Phase-II` detector and an integrated luminosity of 3 ab^{-1}.
The mis-identified lepton background results from the application of the fake rate and the migration matrix to semi-leptonic $t\bar{t}$ and W+jets simulated events, while the "wrong charge" contribution is obtained applying the mis-charge probability measured in 8 TeV data [180] to Z+jets, $W^{\pm}W^{\mp}$ and $t\bar{t}$ fully leptonic decays. The signal is displayed including (full blue line) or excluding (dotted red line) the Higgs boson in the `VBS` stacked on top of the backgrounds, while the full red line shows the difference between these two cases.
Eventually, merging the event categories into a single one, Table 6.4 reports the event yields for both signal and each background contribution obtained after the analysis selections normalized to 3 ab^{-1} for each detector scenario (`Phase-I`, `Phase-II` and `Phase-I Aged`).
Since no `Delphes` simulation is available for the aged detector, the `Phase-I Aged` prediction is obtained by rescaling the `Phase-II` signal yields with the inclusive signal efficiency loss measured with fully simulated events, assuming no difference in the shapes.
Even if it is reasonable to expect that the fake lepton background would be larger with an aged detector, no specific rescaling of the fake lepton background is implemented. Anyway, the degradation of the analysis performance with the increase of the mis-identified lepton

Selection	Value	Comments
`Tight` Leptons		
Electron p_T and η	$p_T > 20$ GeV, $\lvert\eta\rvert < 2.5$ except [1.44, 1.56]	avoid `ECAL` gap.
Muon p_T and η	$p_T > 20$ GeV, $\lvert\eta\rvert < 2.4$	
Isolation as Eq. 6.4	$I_{rel} < 0.25$ (0.65)	`Phase-I` 50PU (`Phase-II` 140PU)
`Loose` Leptons		
Lepton p_T, Isolation	$p_T > 20$ GeV, $I_{rel} < 0.35$ (0.75)	no loose leptons
Jet selection		
Jet p_T and η	$p_T > 30$ GeV, $\lvert\eta\rvert < 4.7$	`CHS` + jet area median correction
Event selections		
Num. `Tight` leptons	2	same charge
Num. `Loose` leptons	0	
Missing energy $\not{E}_T$	$\not{E}_T > 40$ GeV	reduce multi-jet and Z/γ^*+jets
Di-lepton invariant mass $m_{\ell\ell}$	$m_{\ell\ell} > 40$ GeV and $m_{\ell\ell} \neq m_Z \pm 10$ GeV	reduce multi-jet, W and Z/γ^*+jets
Lepton Separation $\Delta\eta_{\ell\ell}$	$\Delta\eta_{\ell\ell} < 2$	reduce $W^{\pm}W^{\pm}$ `QCD`
Num. jets $p_T > 30$ GeV	2	
Num. b-jets	0	reduce $t\bar{t}$ using `CSV`
Num. soft-μ $p_T > 5$ GeV	0	reduce $t\bar{t}$
$H_T = \sum_{jet} p_T^j$	$H_T < 125$ (150) GeV for `Phase-I` (`II`)	reduce $t\bar{t}$
$\Delta R(\ell\ell,\ jj)$	$\Delta R(\ell\ell,\ jj) < 6$	
Zeppenfeld (Z_ℓ)	$Z_{\ell_1} < 0.5$ and $Z_{\ell_2} < 0.5$	
`VBS` jet selections		
Jet separation $\Delta\eta_{jj}$	$\Delta\eta_{jj} > 2.5$	reduce $t\bar{t}$ and W+jets
Di-jet invariant mass m_{jj}	$m_{jj} > 850$ (650) GeV	reduce $t\bar{t}$ and W+jets
Event categories		
Lepton flavor	$\mu\mu$, μe, $e\mu$, ee	leptons are sorted in p_T
Lepton charge	(++), (−−)	WW $O(\alpha_{ew}^6)$ mostly (++)

Table 6.3: List of the selections and the event categories used in this analysis to enhance the sensitivity to the same sign WW bosons scattering signal against reducible and irreducible backgrounds.

background is studied for each detector scenario, in fact the final results are quoted as a function of a global fake rate scale factor.

Scenario	WW α_{ew}^6	WW $\alpha_{ew}^4\alpha_s^2$	WZ α_{ew}^6	WZ $\alpha_{ew}^4\alpha_s^2$	fake lepton	mis-charge
`Phase-I`	2911 ± 54	64 ± 8	115.4 ± 10.7	208.1 ± 14.4	1270 ± 382.6	60.5 ± 19.7
`Phase-II`	3040 ± 55	71.2 ± 8.4	118.7 ± 10.9	175.6 ± 13.2	1328 ± 400.1	85.0 ± 27.1
`Phase-I A`	2128 ± 46	49 ± 7	118.7 ± 10.9	175.6 ± 13.2	1328 ± 400.1	85.0 ± 27.1

Table 6.4: Signal and background yields after the full selection where only Poissonian statistical uncertainties are reported, except for fake lepton and mis-charge backgrounds for which a systematic uncertainty of 30% is added in quadrature.

6.7.1 Systematic uncertainties

As the final results will be derived through a binned likelihood template fit, shape and normalization uncertainties for each process involved must be estimated. Shape uncertainties are evaluated by re-applying the analysis selections and reproducing the distribution of a given observable after considering the systematic variations for each source of uncertainty.

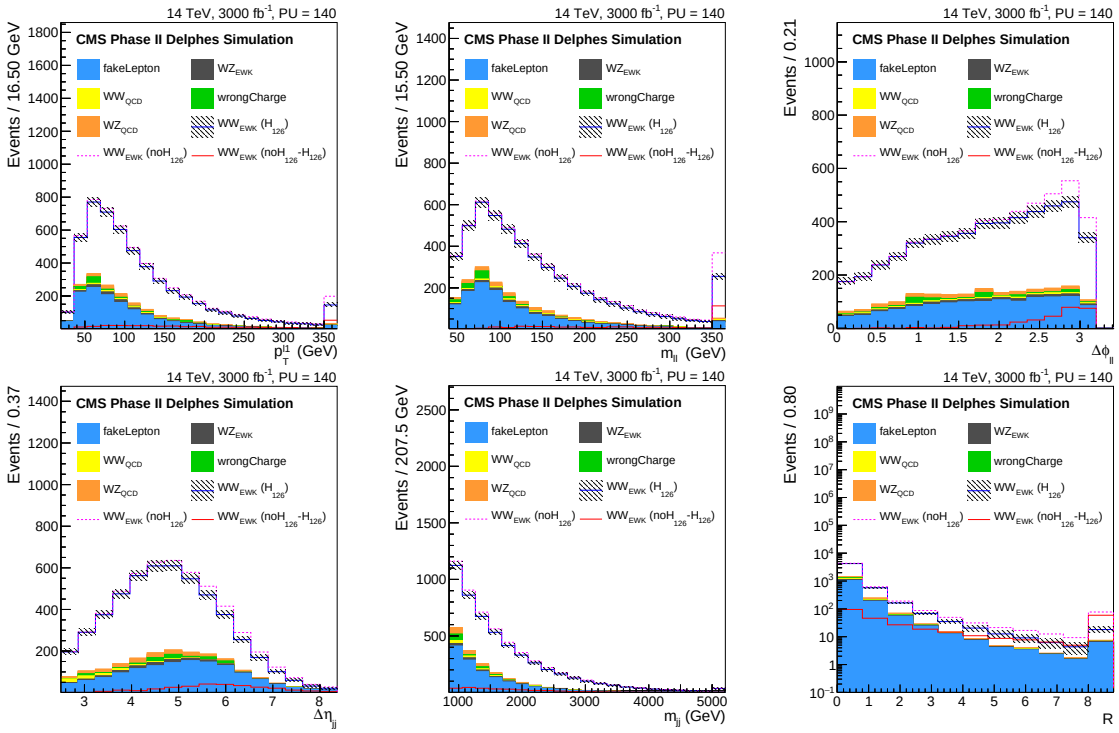

Figure 6.15: Distributions of different observables obtained merging all the event categories after the final selections. Signal and background expectations are reported normalized to an integrated luminosity of 3 ab^{-1} for the CMS Phase-II detector. (Top line) from left to right: leading lepton transverse momentum ($p_T^{\ell_1}$), di-lepton invariant mass ($m_{\ell\ell}$) and $\Delta\varphi_{\ell\ell}$. (Bottom line) from left to right: pseudo-rapidity separation between VBS tag-jets ($\Delta\eta_{jj}$), VBS jets invariant mass (m_{jj}) and R-variable defined as $\mathrm{R} = (|\vec{p}_T^{\,\ell_1}| \cdot |\vec{p}_T^{\,\ell_2}|)/(|\vec{p}_T^{\,j_1}| \cdot |\vec{p}_T^{\,j_2}|)$.

The lepton trigger, reconstruction, and selection efficiencies are usually measured comparing data and simulation in a $Z/\gamma^* \to \ell\ell$ enriched sample. The estimated uncertainty is taken from the 8 TeV analysis [48] measured to be 2% per lepton (e or μ). The uncertainties due to the momentum scale and resolution for electrons and muons are also taken into account by shifting or smearing lepton momenta before re-applying the analysis selection. The uncertainty on the knowledge of the muon (electron) energy scale and resolution is obtained from Phase-I simulation. The uncertainties on the jet energy scale and resolution vary from 1% to 3% as a function of the jet η and p_T. In this process, variations on leptons and jets momenta are propagated consistently to the $\not{E}_T$ vector. The largest relative change in each template bin, compared to the nominal prediction, is taken as a measurement of the systematic effect for that specific source. The systematic related to the b-veto selection is determined to be 4%, while the uncertainty on the luminosity is equal to 2.6% as measured during 8 TeV data taking [108]. The statistical uncertainty in the yield of each bin and is taken into account in the statistical interpretation.

These systematic sources related to the CMS Phase-I detector are also adopted for the Phase-II scenario, since the properties of physics objects are similar between Phase-I 50 PU and Phase-II 140 PU simulations. In contrast, for the Phase-I Aged scenario, several systematics are degradated according to comparisons between the Phase-II and Phase-I Aged full simulations [166].

For the mis-identified lepton background, the normalization uncertainty is assigned to be 30%, as measured in 8 TeV data [181].
The theoretical uncertainty related to the partial knowledge of the cross section, due to missing higher order corrections in the perturbative expansion, is estimated by varying renormalization and factorization scale up and down by a factor of two with respect to the nominal value used in the simulation. It is found to be around 3-5% for both WW (WZ) electroweak scattering and QCD production. The uncertainty on the normalization of the same sign WW boson signal and WZ processes related to the PDF is assigned to vary between 5-7%, as estimated in the 8 TeV VBS search [48]. Finally, an acceptance uncertainty of 2% on the signal normalization is also considered.
The systematic sources affecting the same sign $W^{\pm}W^{\pm}$ scattering analysis are summarized in Table 6.5 for each detector scenario and process.

Source	WW α_{ew}^6		WW $\alpha_{ew}^4\alpha_s^2$		WZ α_{ew}^6 & $\alpha_{ew}^4\alpha_s^2$		fake lep	mis-ch.
	P-I/P-II	P-I A	P-I/P-II	P-I A	P-I/P-II	P-I A		
Experimental Sources								
Jet scale	1-3%	1.5-4%	1-3%	1.5-4%	1-3%	1.5-4%	-	-
Jet res.	1-3%	1.5-4%	1-3%	1.5-4%	1-3%	1.5-4%	-	-
μ scale	1%	2%	1%	2%	1%	2%	-	-
μ res.	1%	2%	1%	2%	1%	2%	-	-
e scale	2%	4%	2%	4%	2%	4%	-	-
e res.	2%	4%	2%	4%	2%	4%	-	-
Lep. eff.	2%	2%	2%	2%	2%	2%	-	-
b-tag eff.	4%	5.5%	4%	5.5%	4%	5.5%	-	-
Fake lept	-	-	-	-	-	-	30%	-
Mis-ch.	-	-	-	-	-	-	-	30%
Lumi	2.6%	2.6%	2.6%	2.6%	2.6%	2.6%	-	-
Theoretical Sources								
Acceptance	2%	2%	-	-	-	-	-	-
QCD scale	3%	3%	3%	3%	5%	5%	-	-
PDF	7%	7%	7%	5%	5%	5%	-	-

Table 6.5: The systematic sources considered for the $W^{\pm}W^{\pm}$ scattering analysis divided between experimental and theoretical sources and quoted for each detector scenario. Systematics on energy scale/resolution of leptons and jets, as well as the fake lepton and the mis-charge ones, are also affecting the shape of the observables used in the statistical interpretation.

6.7.2 Same sign $W^{\pm}W^{\pm}$ scattering result

Once the final event selection is applied, the expected analysis performances for each detector scenario are evaluated by likelihood template fits, where the different templates are derived from simulation treating the systematics as nuisance parameters. Starting from a template with N_{bins} bins, the adopted likelihood function is defined as follows:

$$\mathcal{L}(\text{Data}|\text{S},\text{B}) = \prod_{i}^{\text{N}_{\text{bins}}} \Big[\text{Poisson}(\text{N}_i|\mu\cdot\text{S}_i(\vec{\vartheta}) + \text{B}_i(\vec{\vartheta}))\Big] \times \prod_k \text{Gaus}(\vartheta_k|0,1) \times \prod_j \text{LgN}(\vartheta_j|1,\sigma_{\vartheta_j})$$

where N_i, S_i and B_i represent the observed, expected signal and background rates in the bin i. The number of events observed in each bin is assumed to be Poissonian-distributed with an expected rate given by $\mu \cdot S_i(\vec{\vartheta}) + B_i(\vec{\vartheta})$, where $\vec{\vartheta}$ represents the set of nuisance parameters while μ is the signal strength which rules the overall normalization of the signal with respect to the pre-fit prediction.
Nuisance parameters (systematic uncertainties) affecting only the global normalization of signal or background processes are added to the model through log-normal priors ($\mathrm{LgN}(\vartheta|1, \sigma_\vartheta)$), where the widths are fixed to the values described in Section 6.7.1.
At the same time, shape uncertainties are accounted for introducing a vertical morphing of each bin content, using a Gaussian prior on each morphing parameter [182] ($\mathrm{Gaus}(\vartheta|0, 1)$).
The fit is performed in each of the eight di-lepton categories, then results are combined following the same method developed by ATLAS and CMS collaborations in the context of the LHC Higgs combination group [183, 184]. Templates from different observables are considered and listed in Table 6.6, both one and two-dimensional likelihood fits are performed exploring all the possible pairwise combinations of variables.

Description	Definition
p_T of leading and trailing leptons	$p_T^{\ell_1}$, $p_T^{\ell_2}$
p_T of leading and trailing jets	$p_T^{j_1}$, $p_T^{j_2}$
R-variable	$R = (\lvert p_T^{\ell_1}\rvert \cdot \lvert p_T^{\ell_2}\rvert)/(\lvert p_T^{j_1}\rvert \cdot \lvert p_T^{j_2}\rvert)$
η separation between tag jets	$\Delta\eta_{jj}$
Invariant mass of tag jets system	m_{jj}
Invariant mass of di-lepton system	$m_{\ell\ell}$
Missing transverse energy	$\not{E}_T$
Angular separation along φ	$\Delta\varphi_{\ell\ell}$, $\Delta\varphi_{jj}$, $\Delta\varphi_{\ell\ell,\not{E}_T}$, $\Delta\varphi_{\ell\ell,jj}$
Jet asymmetry	$\mathrm{asim}_j = (\lvert p_T^{j_1}\rvert - \lvert p_T^{j_2}\rvert)/(\lvert p_T^{j_1}\rvert + \lvert p_T^{j_2}\rvert)$
Lepton asymmetry	$\mathrm{asim}_\ell = (\lvert p_T^{\ell_1}\rvert - \lvert p_T^{\ell_2}\rvert)/(\lvert p_T^{\ell_1}\rvert + \lvert p_T^{\ell_2}\rvert)$
Reconstructed invariant mass	$m_{\ell\ell,jj,\not{E}_T}$
Razor mass	$m_{\mathrm{R}} = \sqrt{(E_{\ell_1} + E_{\ell_2})^2 - (p_z^{\ell_1} + p_z^{\ell_2})^2}$
Transverse mass	$m_{\mathrm{T}} = \sqrt{2 p_T^{\ell\ell} \not{E}_T \cdot (1 - \cos(\Delta\varphi_{\ell\ell,\not{E}_T}))}$
Transverse Razor mass	$m_{\mathrm{T}}^{\mathrm{R}} = \sqrt{[\not{E}_T(p_T^{\ell_1} + p_T^{\ell_2}) - \vec{\not{E}}_T \cdot (\vec{p}_T^{\,\ell_1} + \vec{p}_T^{\,\ell_2})]/\sqrt{2}}$

Table 6.6: Main observables considered in the same sign $W^{\pm}W^{\pm}$ scattering analysis.

Figure 6.16 (left) shows the expected uncertainty on the inclusive electroweak scattering cross section measurement, obtained combining all the event categories, when different observables are considered. Both $1\,\sigma$ and $2\,\sigma$ confidence belt on the fitted signal strength are reported. To evaluate the expected performances a signal plus backgrund fits are adopted, where pseudo-data are generated injecting a signal rate compatible with SM prediction, while the nuisance parameters are fixed to the pre-fit estimation.
The inclusive cross-section is determined by fitting the two-dimensional distribution of $(R, m_{\ell\ell})$, where R is defined according to Table 6.6. Figure 6.16 (right) shows the impact of the most relevant nuisance parameters on the estimated cross section uncertainty, when signal and background yields are normalized to 3 ab^{-1} for the Phase-I scenario. The dashed red line refers to the uncertainty as obtained by the complete likelihood fit (6%), empty dots are the

uncertainties obtained by removing one nuisance parameter, indicated by the bin label, while full black dots are the ones coming from a simplified model in which only one nuisance is considered. Finally, the solid red line refers to the fit without any systematic source.
The result is mainly affected by systematic sources involving the signal normalization, such as b-tagging, acceptance, luminosity and jet energy scale, then by the uncertainty on the mis-identified lepton rate.
Finally, the combined expected uncertainty for each detector scenario, evaluated by fitting the $(R, m_{\ell\ell})$ template, is shown in Figure 6.17 as a function of several possible scale factors for the mis-identified lepton background (left) and the integrated luminosity (right), fixing the fake rate scale factor to unity. After 3 ab^{-1} of data, this measurement appears to be dominated by the systematic uncertainties, where the cross section is measured with a precision of about 6% for both `Phase-I` and `Phase-II`. This degrades to 8% when the `Phase-I Aged` detector is considered.

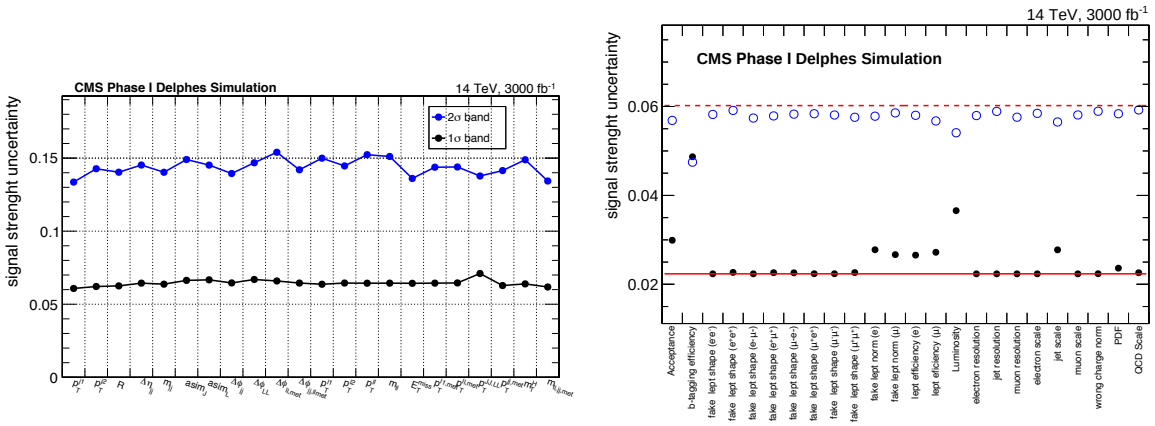

Figure 6.16: (Left) The expected total uncertainty for the same-sign WW scattering cross section measurement after 3 ab^{-1}, as obtained by a template fit on different observables. (Right) A detailed study of the impact of the most relevant systematic sources on the cross section measurement. Empty dots refer to the uncertainties obtained removing one nuisance parameter, full black dots are obtained adding a single nuisance parameter to the fit model. Dash (solid) red line refers to the uncertainty obtained including all (no) the systematics in the final fit.

6.7.3 Sensitivity to the longitudinal $W^{\pm}W^{\pm}$ scattering component

The inclusive `VBS` cross section is composed by three terms depending on the possibile polarization combinations of the final-state vector bosons: $W_L^{\pm}W_L^{\pm}$, $W_L^{\pm}W_T^{\pm}$ and $W_T^{\pm}W_T^{\pm}$.
The discovery significance is adopted as figure of merit to asses the sensitivity of the analysis to the $W_L^{\pm}W_L^{\pm}$ scattering with respect to the transverse modes ($W_L^{\pm}W_T^{\pm}$ and $W_T^{\pm}W_T^{\pm}$) and the other backgrounds listed in Table 6.4. The significance is defined from the following test statics:

$$q_0 = -2\ln\frac{\mathcal{L}(\text{Data}|\mu=0, \text{B}(\hat{\vartheta}_0))}{\mathcal{L}(\text{Data}|\hat{\mu}\cdot\text{S}(\hat{\vartheta})+\text{B}(\hat{\vartheta}))} \quad \hat{\mu}\geq 0 \tag{6.8}$$

where a signal-like excess ($\hat{\mu} > 0$) corresponds to a positive value of q_0, while if data are compatible with the background-only hypothesis (null hypothesis) $q_0 \to 0$. The likelihood functions have the same form of the one in Section 6.7.2.
In the asymptotic limit, this test statistics follows a χ^2 distribution which allows for the

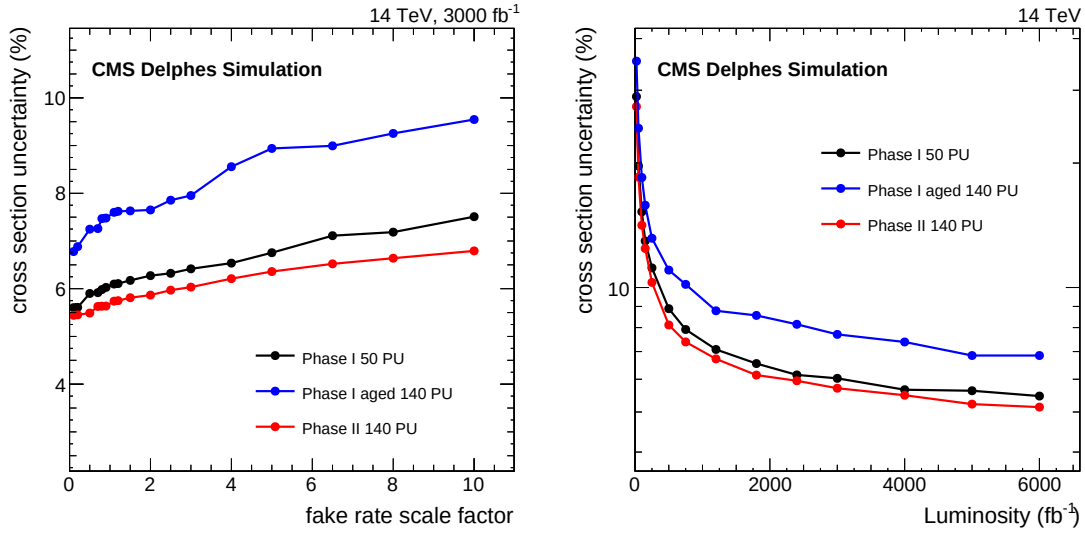

Figure 6.17: The expected uncertainty on the inclusive electroweak scattering cross section for the different detector scenarios as a function of: (left) several scale factors to the mis-identified lepton background after 3 ab^{-1} of data, (right) integrated luminosity of the sample for a unity scale factor.

evaluation of a local p-value as:

$$p_0 = P(q_0 \geq q_0^{obs}|\mathrm{B}) = \int_{\sqrt{q_0^{obs}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{2}\Big(1 - \mathrm{erf}(\sqrt{q_0^{obs}}/2)\Big) \tag{6.9}$$

which measures the probability to obtain a q_0 value larger than the observed one (q_0^{obs}) under the background-only hypothesis.
Particular care is dedicated to optimize the analysis selections, as described in Section 6.7, as well as to identify the kinematic variables which provide the best separation with respect to the transverse components. The polarized samples produced by `MadGraph` are adopted and Table 6.7 reports the expected yields for the signal, breakdown into the different polarization modes, and the background processes after the selections foreseen for this analysis.

Scenario	W_LW_L	W_LW_T	W_TW_T	WW QCD	WZ	fake lepton	mis-charge
Phase-I	149.5	1061.7	1981.6	116.1	462.8	2247.8	61.6
Phase-II	173.9	1196.1	2183.3	175.6	412.3	2759.7	103.2
Phase-I A	121.6	837.3	1528.3	122.9	412.3	2759.7	103.2

Table 6.7: Signal and background yields after the dedicated selections foreseen for the longitudinal scattering analysis. The scattering yields, predicted by `MadGraph` and normalized to 3 ab^{-1}, are subdivided among the different polarizations of the vector bosons in the final state.

The production of two longitudinally polarized vector boson is a rare process also in the acceptance region of the analysis, as it represents only 7% of the selected W$^{\pm}$W$^{\pm}$ events.
Figure 6.18 shows the performances in terms of discovery significance extracted adopting a binned likelihood analysis for a number of different observables (left) and different two-dimensional templates (right). The best performance is achieved through a fit to the $(\Delta\varphi_{jj}, p_T^{\ell_1})$ distribution. Figure 6.19 shows the shapes of these two observables for the $W_L^{\pm}W_L^{\pm}$, $W_L^{\pm}W_T^{\pm}$ and $W_T^{\pm}W_T^{\pm}$ components of the `VBS` process.

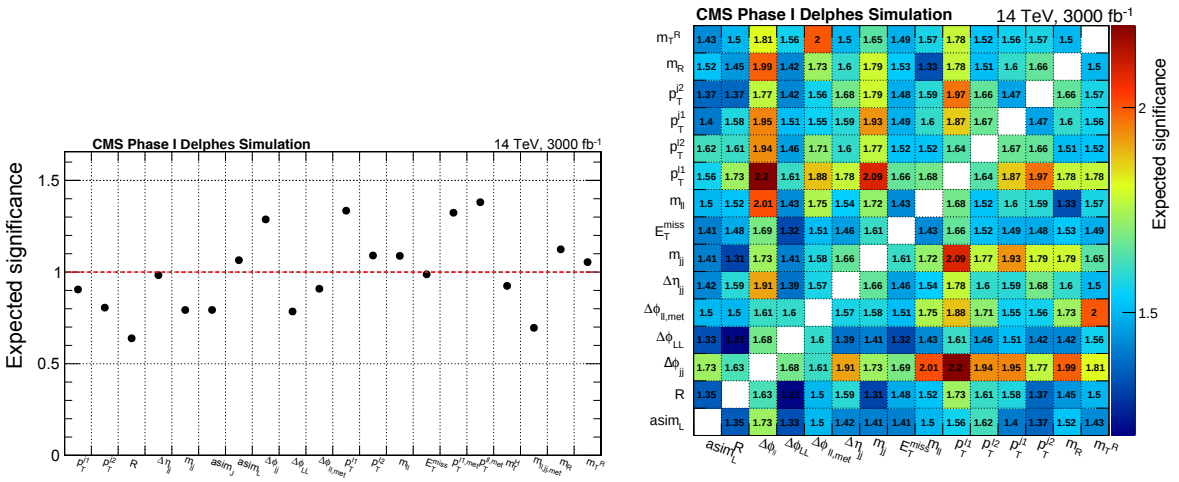

Figure 6.18: Determination of the best observables sensitive to the longitudinal component of the electroweak scattering for the `Phase-I` detector scenario. (Left) Expected significance obtained from single observables, (Right) expected significance for different pairs of variables. The maximum significance is reached by a two-dimensional fit to $(\Delta\varphi_{jj}, p_T^{\ell_1})$.

Eventually, the expected significance estimated combining all the di-lepton categories is reported in Figure 6.20 for each detector scenario. On the left, the expectation after 3 ab^{-1} of data is shown as a function of a global scale factor for the fake lepton background, while, on the right, it is reported as a function of the collected luminosity. Considering the `CMS Phase-II` detector, a unity scale factor for the fake rate and 3 ab^{-1} of integrated luminosity, an expected significance of 2.4 σ is reached, while a value lower by 25-30% would be obtained with the `Phase-I Aged` detector.

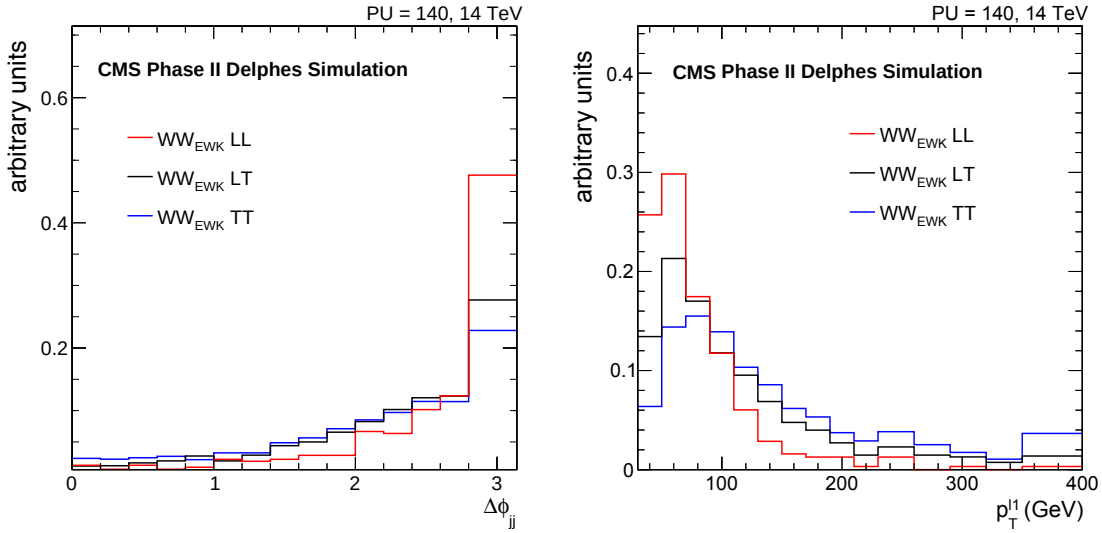

Figure 6.19: Shapes of the observables adopted to asses the sensitivity to the longitudinal scattering component as obtained for the `Phase-II` detector in the $\mu^+\mu^+$ category: (left) azimuthal angular difference between the two `VBS` tag-jets $(\Delta\varphi_{jj})$, (right) leading lepton transverse momentum (p_T^{ℓ}).

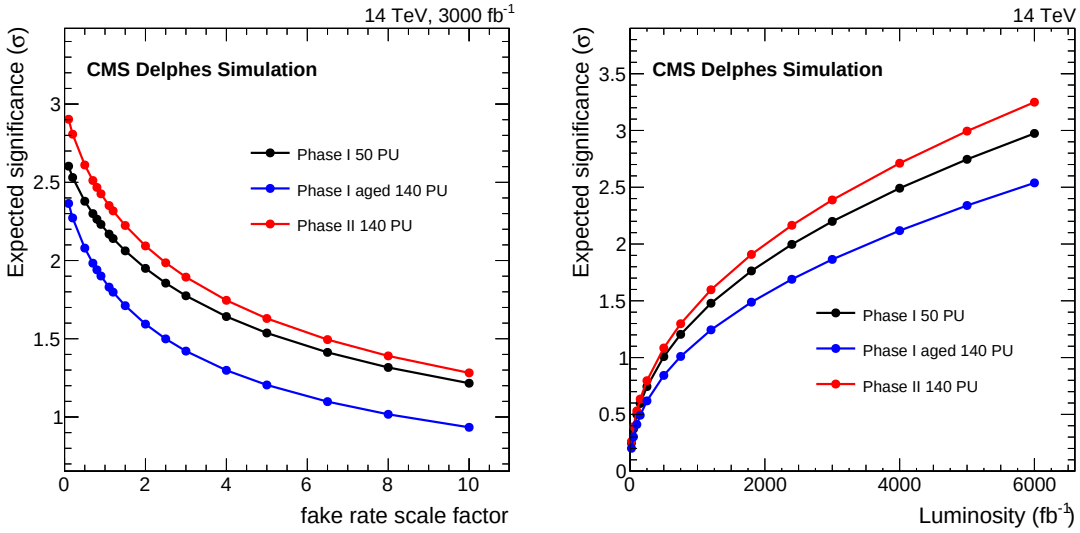

Figure 6.20: On the left, the expected discovery significance for the longitudinal vector bosons scattering in the W$^{\pm}$W$^{\pm}$ channel for the different detector scenarios, as a function of a possible scale factor to the fake rate after 3 ab^{-1} of data. On the right, the evolution of the discovery sensitivity, for a unity fake rate scale factor, as a function of the integrated luminosity.

6.7.4 Partial unitarization scenario

As described in Section 6.4.1, partial unitarized scenarios are simulated, without introducing any additional resonance along the same sign W$^{\pm}$W$^{\pm}$ spectrum, by changing the coupling strength of the SM Higgs to vector bosons (κ_V). A discrete scan is performed as a function of the coupling strength, as indicated by the samples listed in Table 6.2, where $\kappa_V^2 = 1$ refers to the SM scenario where the unitarization is granted by the SM Higgs boson, while $\kappa_V = 0$ refers to the Higgs-less case.

To study qualitatively the effect of κ_V^2 on the VBS spectrum, the distribution of the main observables considered in the analysis are compared, in both shape and normalization, after applying the analysis selection. Taking the SM case as a reference benchmark, Figure 6.21 shows the difference between the distributions obtained in a partial unitarized framework ($\kappa_V^2 = 0.0, 0.5, 0.9$) and the SM expectation for the leading lepton $p_T^{\ell_1}$ (left), $m_{\ell\ell}$ (middle) and m_{jj} (right). The predicted difference on the normalization is roughly quadratically dependent on $\Delta\kappa_V^2$ and the shapes look similar as a function of the κ_V^2 value.

The analysis has been optimized adopting the Higgs-less case as reference, since small variations in the discrimination power of kinematic observables are expected when different κ_V^2 are considered. In particular, after applying the analysis selections, the templates obtained through the difference between the Higgs-less scenario and the SM one are considered as a potential signal (noH-H), while the W$^{\pm}$W$^{\pm}$ scattering predicted by the nominal couplings is considered as part of the background. To assess the most sensitive kinematic variables to this non unitarized scenario, the 95% CL exclusion limit for the Higgs-less hypothesis, calculated adopting the asymptotic approximation of the CL$_S$ method [152] and the LHC test statistics [184], is used as figure of merit.

Figure 6.18 shows the performances in terms of 95% CL exclusion limit obtained via a binned likelihood fit for a number of different observables (left) and different two-dimensional tem-

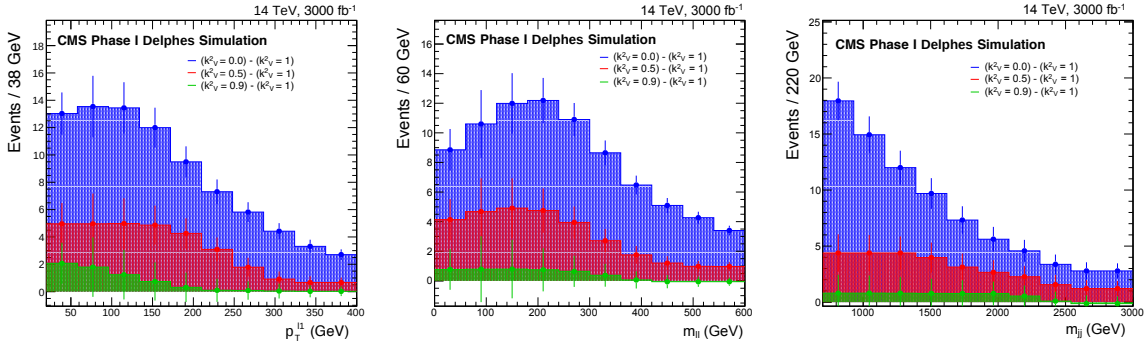

Figure 6.21: Difference between partial unitarized scenarios in which the Higgs coupling to vector bosons is varied ($\kappa_V^2 = 0.0, 0.5, 0.9$) and the SM case ($\kappa_V^2 = 1$) for three different observables: leading lepton $p_T^{\ell_1}$ (left), di-lepton invariant mass $m_{\ell\ell}$ (middle) and the VBS tag-jet invariant mass m_{jj} (right).

plates (right), where the eight categories of the analysis are combined and the Phase-I detector scenario is considered. The optimal result is obtained by fitting a two-dimensional template represented by (R,$m_{\ell\ell}$).

Figure 6.23 shows the ultimate expected 95% CL exclusion limit for the Higgs-less scenario. The result is reported in terms of the strength modifier μ of the noH-H hypothesis as a function of a global scale factor applied to the jet-to-lepton mis-identification rate (left) and of the total integrated luminosity (right).

When the exclusion reaches the unity value, after about 100 fb^{-1}, the analysis is expected to become sensitive to partial unitarized scenarions, in which the existing Higgs boson does not completely fulfill the VBS unitarization. From these projections, the upgraded CMS detector is more sensitive than the Phase-I Aged one and recovers similar performances one would get with the same luminosity recorded with the LHC Run-I data conditions.

To study the sensitivity to partial unitarized scenarios induced by variations of κ_V, a ded-

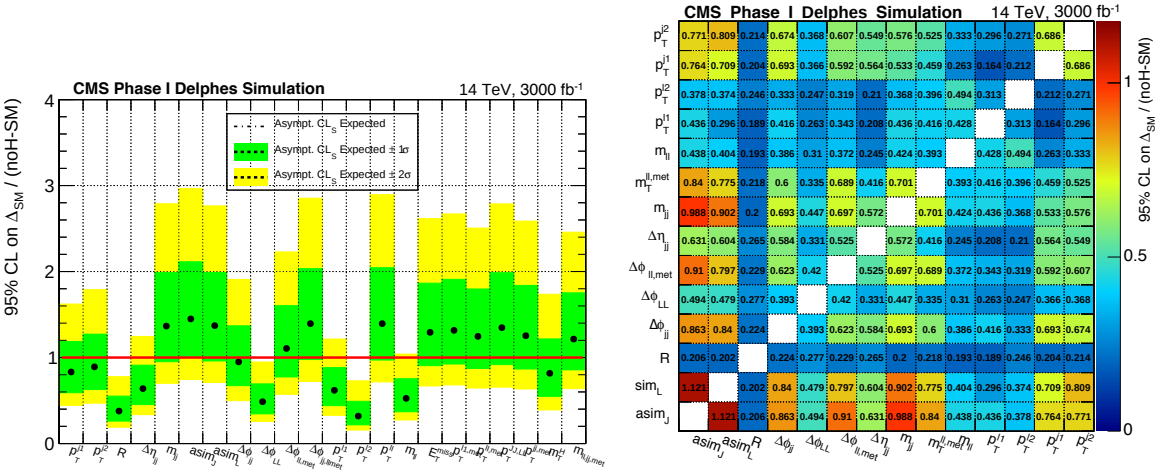

Figure 6.22: Determination of the best observables sensitive to the Higgs-less hypothesis considering the CMS Phase-I detector. Signal and backgrounds yields are normalized to 3 ab^{-1}. (Left) Expected 95% CL exclusion limit obtained from single observables, (right) the exclusion limits evaluated for different pairwise combinations. The lowest limit is reached by a two-dimensional fit to R-$m_{\ell\ell}$.

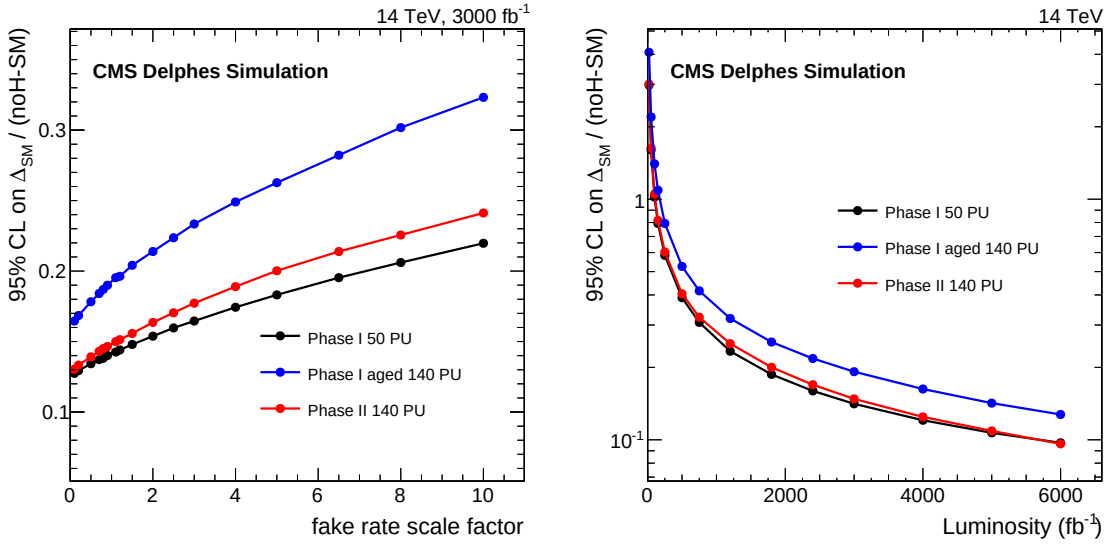

Figure 6.23: The expected 95% CL exclusion limits for the Higgs-less scenario after 3 ab^{-1} of integrated luminosity, as a function of the jet-to-lepton mis-identification rate scale factor (left), as well as a function of the integrated luminosity for a unity fake rate scale factor (right). The limit is expressed as a deviation from the SM divided by the difference of the Higgs-less case from the SM itself.

icated hypothesis test is adopted. In particular, considering the SM case as null hypothesis (H_0), alternative ones (H_a) are tested through the following test statistics:

$$q = -2 \ln \frac{\mathcal{L}(\text{Data}|\hat{\mu}_a \cdot \text{S}(\kappa_V^2 \neq 1, \hat{\vartheta}_a) + \text{B}(\hat{\vartheta}_a))}{\mathcal{L}(\text{Data}|\hat{\mu}_0 \cdot \text{S}(\kappa_V^2 = 1, \hat{\vartheta}_0) + \text{B}(\hat{\vartheta}_0))} \tag{6.10}$$

where $(\hat{\mu}_a,\hat{\vartheta}_a)$ and $(\hat{\mu}_0,\hat{\vartheta}_0)$ indicate the fitted signal strength and the fitted nuisance parameters under the alternative and the null hypothesis, respectively. Toy experiments are generated according to either the SM or the alternative scenarios, considering the scattering rates predicted by the two models, and the test statistics is computed for each considered κ_V^2 value and detector scenario assuming a unity scale factor for the fake lepton background and a total luminosity of 3 ab^{-1}. Figure 6.24 shows two examples of the test statistics obtained and used to evaluate the seperation of $\kappa_V^2 = 0.5, 0.8$ with respect to the SM for the Phase-II scenario. The separation is estimated as the Gaussian quantile of the CL_s value calculated as follows:

$$\text{CL}_\text{S} = \frac{\text{CL}_\text{S+B}}{\text{CL}_\text{B}} = \frac{P(q > \tilde{q}(\text{H}_0)|\text{H}_a)}{P(q > \tilde{q}(\text{H}_0)|\text{H}_0)} \tag{6.11}$$

Independently from the direct measurement of κ_V that is obtained studying Higgs boson properties, the same sign WW bosons scattering would allow to exclude values of κ_V^2 lower than about 0.75 at 95% CL with the Phase-II detector. In contrast, worse limits by about 20% are obtained with the Phase-I Aged detector, as shown in Figure 6.25.

6.7.5 Anomalous couplings results

As described in the previous Sections, BSM physics effects altering the VBS properties are described with the EFT formalism adding a series of high dimensional operators to the SM Lagrangian. These operators are the low energy effective description of interactions mediated

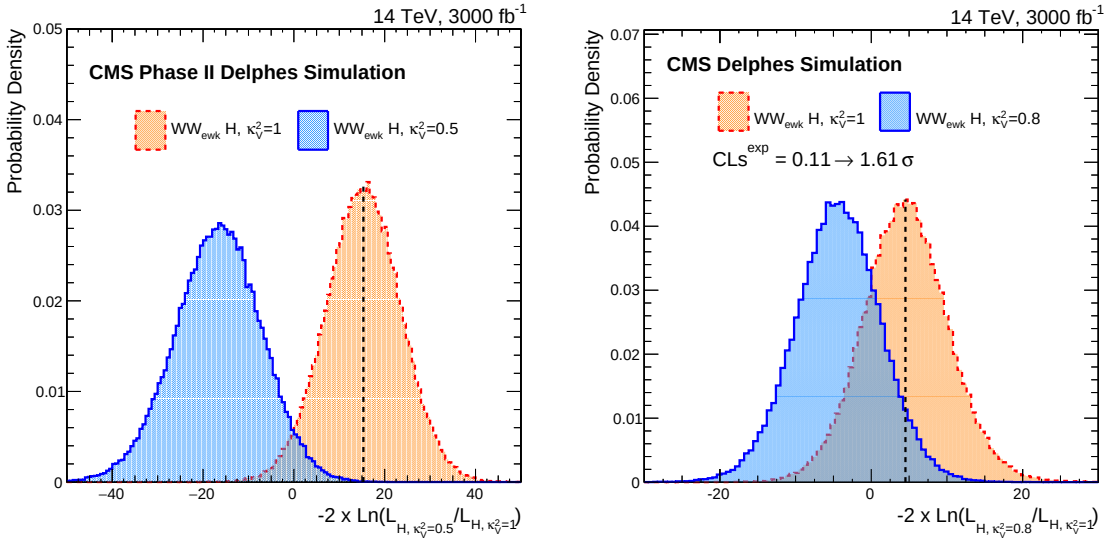

Figure 6.24: Distributions of the test statistic evaluated for different partial unitarized cases $\kappa_V^2 = 0.5$ (left) and $\kappa_V^2 = 0.8$ (right) compared to the SM one after 3 ab^{-1} and for a unity scale factor of the fake lepton rate in the Phase-II detector scenario.

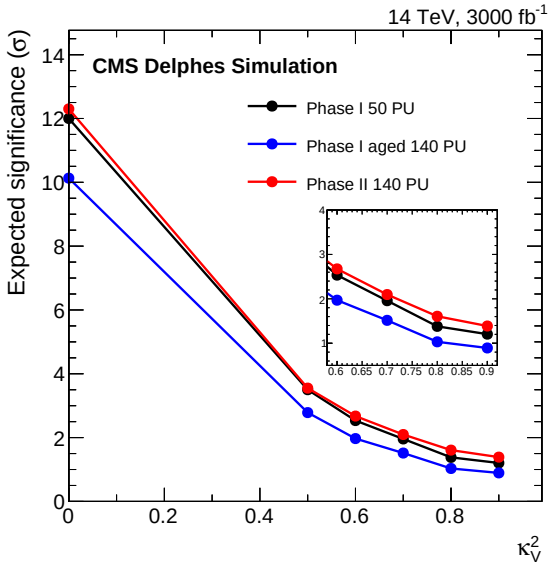

Figure 6.25: The expected separation between the SM and partial unitarized scenarios as a function of κ_V^2 for the three detector configurations. Projections are obtained considering 3 ab^{-1} of integrated luminosity and a unity scale factor for the fake lepton background.

by massive fields inside the spectrum of an unknown BSM theory, for which no direct observation is possible. While dimension-six operators are well constrained by other measurements ($W^{\pm}W^{\mp}$, $V\gamma$, WV), the VBS process is also modified by dimension-eight ones through anomalous quartic gauge couplings.

In the scattering of same sign W-bosons, nine possible CP-conserving dimension-eight operators could modify the scattering process [61] involving only the Higgs doublet and the $SU(2)_L$ field strengths ($W^{\mu\nu}$). The S-operators correspond to the effect of on-shell production of scalar particles (S_0,S_1), while the T-operators (T_0,T_1,T_2) correspond to transversely polarized vec-

tor bosons. Such operators would alter the large interference between the components of VBS processes and potentially lead to strong enhancements of the cross section along high energy tails of the distributions.
To assess the potential of the VBS analysis to constrain EFT operators, a mono-dimensional template likelihood fit to the di-lepton invariant mass ($m_{\ell\ell}$) is adopted. Events are simulated with non-zero values for the Wilson coefficients for these operators with MadGraph[1], over a range comparable to the sensitivity region that can be probed at the HL-LHC. These coefficients are scanned independently, while in the S_0-S_1 case also a simultaneous scan is performed since this combination provides a direct measurement of new physics that only modifies the SM quartic gauge couplings.
After the analysis selections, the dependance of the signal rate as a function of the Wilson coefficient of a certain operator in each bin of $m_{\ell\ell}$ is fitted with a second order polynomial, as shown in Figure 6.26 for the S_0 operator (left) and the simultaneous S_0-S_1 scan (right).
In addition, normalizing the event rate to an integrated luminosity of 3 ab^{-1} and considering the CMS Phase-II detector, Figure 6.27 shows the expected $m_{\ell\ell}$ distributions for the SM signal and the modifications induced by two values of the Wilson coefficients associated to the S_0 and the T_0 operators.

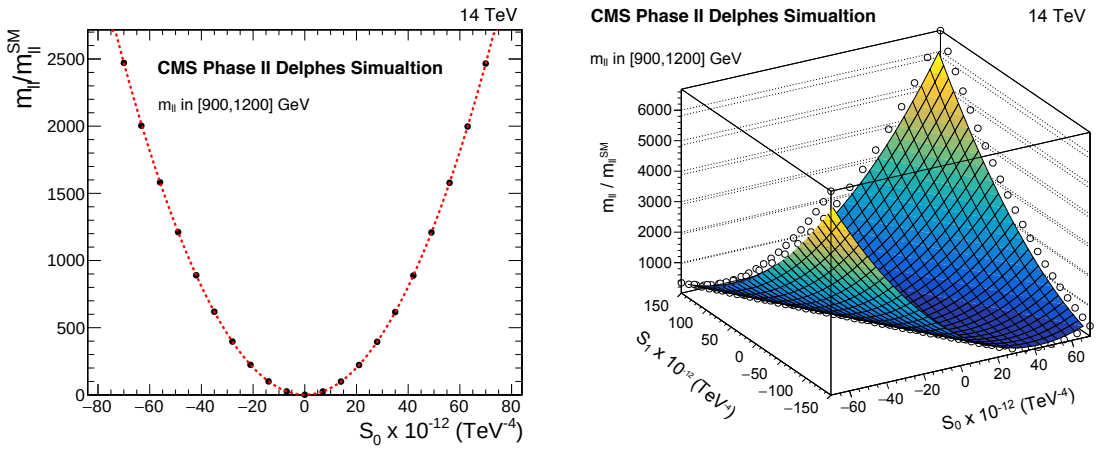

Figure 6.26: (Left) Effect of the Wilson coefficient scan for the S_0 operator on the predicted rate for the same sign WW boson scattering in the $m_{\ell\ell}$ range between 900-1200 GeV. (Right) The same effect is displayed for the simultaneous scan of the Wilson coefficients for S_0 and S_1.

The expected limits at 95% CL on the nine operators are listed in Table 6.8, where the Phase-I and Phase-II detectors show similar performances, while the Phase-I Aged scenario shows a worsening of about 10%. The effect of possible aQGC on the WZ scattering process in the WW same sign signal region is expected to be negligible. In addition to the limits on individual operators, Figure 6.28 (left) shows the two-dimensional 95% CL contour on the S_0 and S_1 coefficients. To investigate the importance of both experimental and theoretical systematics on the limit extraction, the dependance of the exclusion sensitivity is studied as a function of the integrated luminosity. Figure 6.28 (right) shows the trend of the exclusion bound on S_1 versus the luminosity, proving how the analysis is expected to become systematically limited after about 2.5 ab^{-1}.

[1] Wilson coefficients are the coupling constant to higher-order operators introduced in the EFT parametrization.

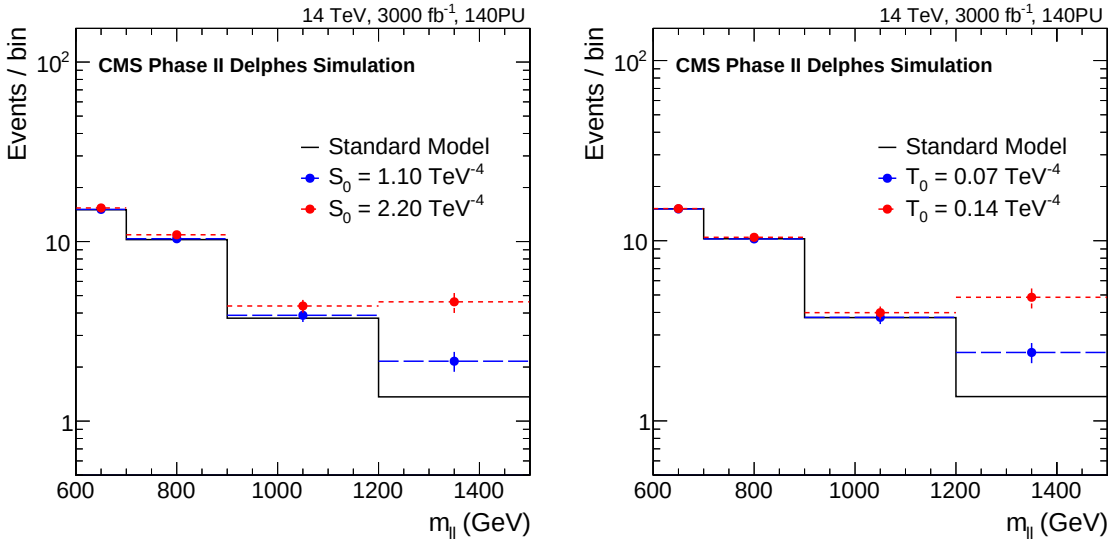

Figure 6.27: Distributions of the di-lepton invariant mass, when $m_{\ell\ell} > 600$ GeV, for the SM signal and its modifications for two values of the coefficient associated with the S_0 (left) and T_0 (right) operators.

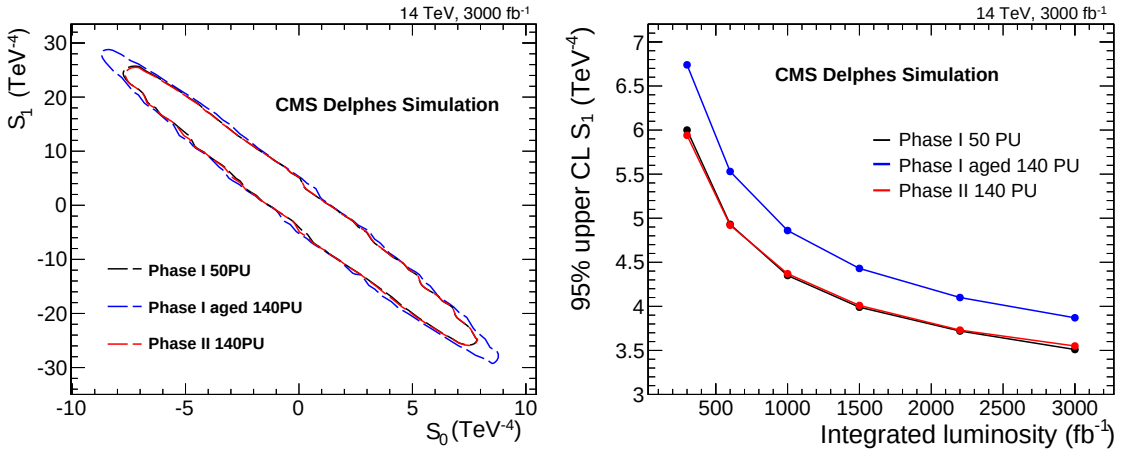

Figure 6.28: (Left) Expected 95% CL contours on the coefficients associated to the S_0 and S_1 operators for 3 ab^{-1} of data, a unity scale factor of the fake lepton rate and for the three detector scenarios. (Right) Limits on the coefficient associated to the S_1 operator for the different detector scenarios as a function of the collected luminosity.

6.8 The WZ boson scattering

This channel is identified by searching for three identified and isolated charged leptons in the final state, with $p_T > 20$ GeV, where two of them must have the same flavor, opposite charge and an invariant mass compatible with the Z-boson one within ± 10 GeV.
The undetected neutrino generates missing transverse energy. The longitudinal component of the neutrino momentum is obtained by requiring it to be produced, together with the remaining charged lepton, from the decay of a on-shell W-boson, following the same procedure illustrated in Section 4.2.1.
The background due to $Z/\gamma^* \to \ell\ell$ produced in association with jets and semi-leptonic $t\bar{t}$ events, when one jet is wrongly identified as a charged lepton, is not found to be significant,

Operator (TeV^{-4})	Phase-I	Phase-II	Phase-I Aged
S_0/Λ^4	1.06	1.07	1.17
S_1/Λ^4	3.51	3.55	3.87
T_0/Λ^4	0.067	0.077	0.083
T_1/Λ^4	0.036	0.033	0.036
T_2/Λ^4	0.119	0.111	0.119
M_0/Λ^4	0.78	0.75	0.82
M_1/Λ^4	1.10	1.06	1.14
M_6/Λ^4	1.56	1.49	1.63
M_7/Λ^4	1.37	1.32	1.45

Table 6.8: Expected 95% CL limits on the coefficients for BSM higher-order operators in the EFT Lagrangian, after 3 ab^{-1} of data and for a unity scale factor of the fake lepton rate, as a function of the three detector scenarios. The units are TeV^{-4}.

especially once the full analysis selections are applied. In contrast, the irreducible WZ boson production in association with jets coming from strong vertices $O(\alpha_{ew}^4\alpha_s^2)$ is much more important than the corresponding one in the same-sign WW scattering, representing the leading background of the analysis.
The presence of additional loosely identified leptons is vetoed to remove contaminations from ZZ+jets electroweak ($O(\alpha_{ew}^6)$) and QCD ($O(\alpha_{ew}^4\alpha_s^2)$) productions and $\not{E}_T$ should be larger than 30 GeV. Any same flavor (SF) and opposite sign (OS) lepton pair should have an invariant mass $m_{\ell\ell}$ larger than 20 GeV.
The two highest p_T jets, with $p_T > 30$ GeV, not in geometrical overlap with the already identified leptons, are chosen as VBS tag-jets. Their pseudorapidity separation ($\Delta\eta_{jj}$) must be larger than 4 units, while the invariant mass of the tag jet system is required to be larger than 600 GeV.
Eventually, to maximize the analysis performance, the selected events are further classified into four categories according to the flavor content of the final state: $\mu\mu\mu$, $\mu\mu e$, $ee\mu$ and eee. The selection strategy is summarized in Table 6.9.
Applying the selections previously described, the distributions of the distance in pseudorapidity between the two leptons with same charge ($\Delta\eta_{\ell\ell}^{\text{SC}} = |\eta_{\ell_1}^{\pm} - \eta_{\ell_2}^{\pm}|$), $p_T^{\ell_1}$ and $\Delta\varphi_{jj}$ are reported in Figure 6.29 for the Phase-II detector and an integrated luminosity of 3 ab^{-1}. The main backgrounds are represented by QCD production of ZZ and WZ boson pairs, as well as the electroweak production of ZZ pairs. The signal is displayed including (full blue line) or excluding (dotted red line) the Higgs boson contribution stacked on top of the backgrounds, while the full red line shows the difference between the SM and the Higgs-less case.
The experimental sources of systematic uncertainties are evaluated similarly to the same sign WW case described in Section 6.7.1, while theoretical systematics have been already listed in Table 6.5.

6.8.1 The WZ scattering results

The inclusive electroweak WZ scattering cross section is determined by a two-dimensional binned likelihood fit of (p_T^{jj},$\Delta\eta_{\ell\ell}^{\text{SC}}$). The expected uncertainty on the signal strength, obtained after a simultaneous fit of the four event categories, is shown in Figure 6.30 (left) as a function of the integrated luminosity for the three different detector configurations. An uncertainty of

Selection	Value	Comments
`Tight` Leptons		
Electron p_T and η	$p_T > 20$ GeV, $\lvert\eta\rvert < 2.5$	avoid `ECAL` gap.
Muon p_T and η	$p_T > 20$ GeV, $\lvert\eta\rvert < 2.4$	
Isolation as Eq. 6.4	$I_{rel} < 0.25$ (0.65)	`Phase-I` 50PU (`Phase-II` 140PU)
`Loose` Leptons		
Lepton p_T, Isolation	$p_T > 20$ GeV, $I_{rel} < 0.35$ (0.75)	no loose leptons
Jet selection		
Jet p_T and η	$p_T > 30$ GeV, $\lvert\eta\rvert < 4.7$	`CHS` + jet area median correction
Event selections		
Num. `Tight` leptons	3	two with `OS` and `SF`
Num. `Loose` leptons	0	
Missing energy $\not{E}_T$	$\not{E}_T > 30$ GeV	reduce Z+jets and ZZ
Di-lepton invariant mass $m_{\ell\ell}$	$m_{\ell\ell} > 20$ GeV	for each `OS` and `SF` pair
Di-lepton invariant mass $m_{\ell\ell}$	$m_{\ell\ell} \neq m_Z \pm 10$ GeV	for only one `OS` and `SF` pair
Num. jets $p_T > 30$ GeV	2	
`VBS` jet selections		
Jet separation $\Delta\eta_{jj}$	$\Delta\eta_{jj} > 4$	
Di-jet invariant mass m_{jj}	$m_{jj} > 600$ GeV	
Event categories		
Lepton flavor	$\mu\mu\mu$, $\mu\mu e$, $ee\mu$, eee	

Table 6.9: List of the selections and the event categories used in this analysis to enhance the sensitivity to the fully leptonic WZ boson scattering signal against reducible and irreducible backgrounds.

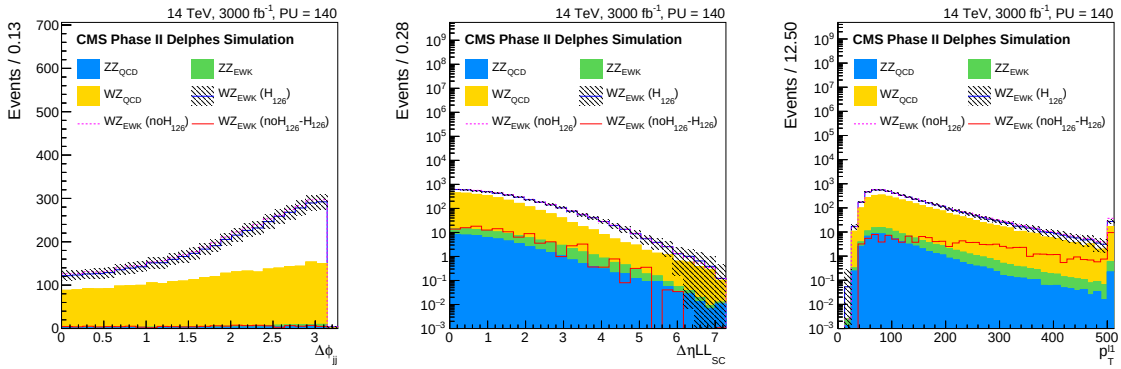

Figure 6.29: From left to right: distributions of the difference in azimuthal angle between the tag jets ($\Delta\varphi_{jj}$), the difference in pseudorapidity between the same charge leptons ($\Delta\eta_{\ell\ell}^{SC}$) and the leading lepton transverse momentum ($p_T^{\ell_1}$) for the `Phase-II` detector and the WZ selection. Signal and the backgrounds are normalized to the integrated luminosity of 3 ab^{-1}.

about 10% is achieved after 3 ab^{-1} with `Phase-I` and `Phase-II` detectors, while the `Phase-I Aged` scenario would provide worse result by about 20%.

In addition, the expected sensitivity to the longitudinal component of the WZ boson scattering has been also studied, as in the same sign WW case. Different observables are tested and the two-dimensional distribution of ($\Delta\varphi_{jj}$,$p_T^{\ell_1}$) is the one which provides the best discrimination between the longitudinal and the transverse modes. The expected significance for measuring the longitudinal cross section, after combining all the tri-lepton categories, is shown in Figure 6.30 (right) as a function of the collected luminosity.

The best result is achieved with the Phase-II detector, 1.4 σ significance with 3 ab^{-1}, thanks to an improved lepton reconstruction and identification efficiency with respect to the Phase-I scenatio, while worse performances of about 30% are reached with the Phase-I Aged detector.

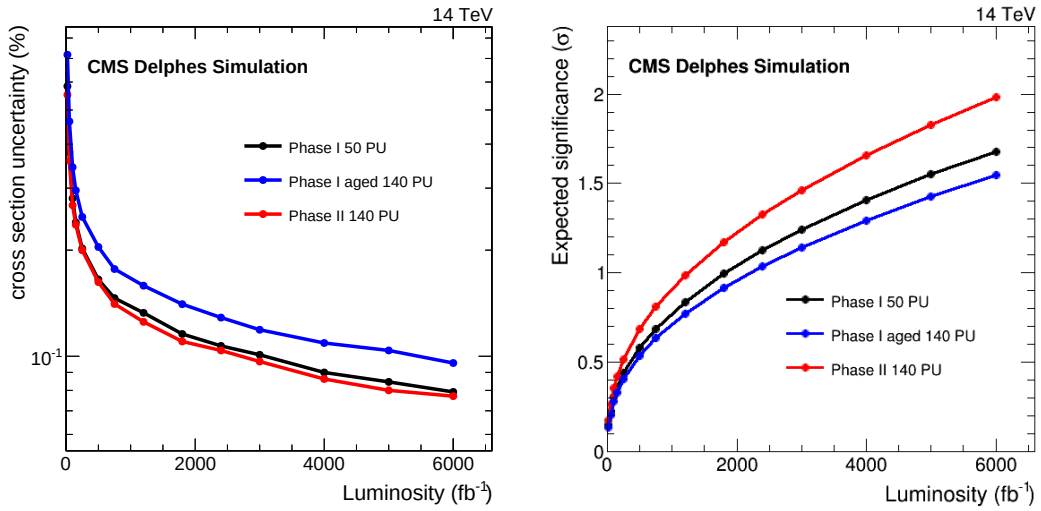

Figure 6.30: (Left) The expected uncertainty for the various detector scenarios on the electroweak WZ cross section as a function of the collected luminosity. (Right) The expected discovery significance for the longitudinal WZ scattering component for the various detector scenarios as a function of the integrated luminosity.

6.9 Combined sensitivity to the longitudinal scattering

Results obtained from the WW and WZ analyses are combined to determine the overall performance for the observation of the longitudinal scattering, where the Higgs boson alone fulfills the unitarization of the VBS cross section. In the combination, common systematic sources are considered fully correlated between the two channels such as the uncertainty of the luminosity of the data sample, lepton efficiencies, jet/lepton energy scale and resolution, as well as theoretical uncertainties due to QCD scale and the PDF determination.
Figure 6.31 shows the evolution of the combination as a function of the integrated luminosity (right), for a unity mis-tag lepton scale factor, as well as the dependance on the jet-to-lepton mis-identification scale factor with 3 ab^{-1} of data (left). The Phase-II detector scenario provides better performances in all the investigated conditions, reaching about 2.75 σ significance after 3 ab^{-1} considering the central prediction for the mis-identified lepton background.

6.10 Conclusions

A feasibility study for the measurement of vector boson scattering properties has been presented in this Chapter. Two promising fully leptonic final states are considered: the same sign WW and WZ bosons scattering topologies. For both of them, the precision in the measurement of the electroweak cross section is studied, as well as the expected discovery significance of its longitudinal component, which represents one of the most challenging measurements for the HL-LHC physics programme. In addition, the exclusion power to residual non-unitarisation

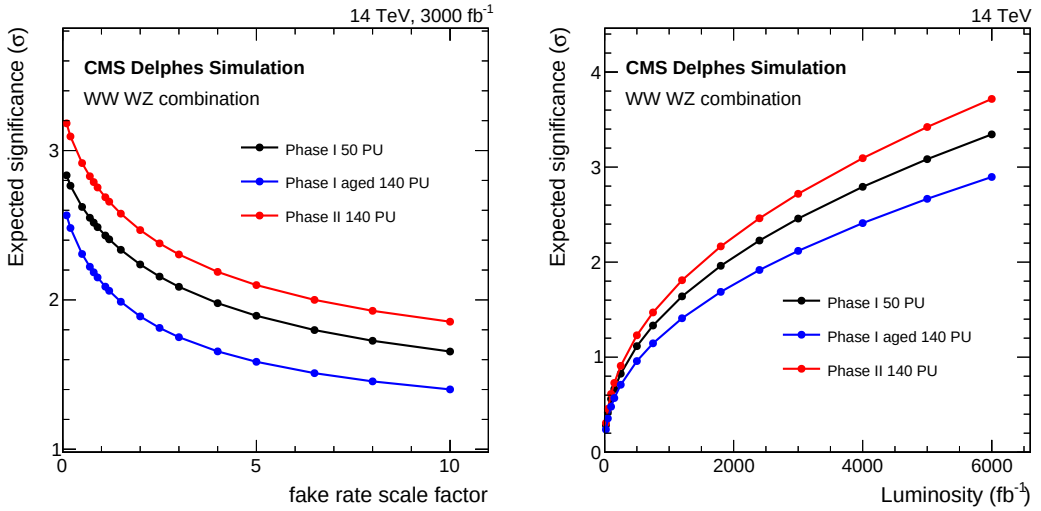

Figure 6.31: (Left) The expected discovery significance for the longitudinal vector bosons scattering for the various detector scenarios as a function of the fake rate scale factor after 3 ab^{-1} of data, for the combination of samesign WW and WZ analyses. (Right) The evolution of the discovery sensitivity, for a unity scale factor of the fake rate, as a function of the collected luminosity.

effects at high energy has been presented for the same sign WW case. Finally, BSM sensitivity is probed in the framework of effective field theories by extracting expected limits on the anomalous quartic gauge couplings through the $W^{\pm}W^{\pm}$ scattering.
Results are presented for three detector scenarios, characterized by alternative possible conditions of CMS during the HL-LHC data taking. The current detector design, aged by the strong irradiation, is compared to a model with tracker system, muon system and forward calorimetry upgraded to stand the harsh environment foreseen during the HL-LHC. Eventually, also the non-aged current CMS detector is tested, to provide a benchmark of the physics performances of the present detector. Signal and background events have been processed with Delphes, which provides a parametric simulation of the CMS detector that has been validated to the full CMS simulation when possible. An attempt is made to include also reducible backgrounds, as the ones due to jets mis-identified as leptons. Given the large uncertainty on the detailed detector simulation, results are expressed as a function of a data-to-simulation scale factor for the mis-tag lepton rate.

Conclusions

The most relevant achievements of my three year Ph.D. work in CMS have been presented in this thesis. The field of interests of my research activity has been twofold.
Looking at LHC Run-I data taking at $\sqrt{s} = 8$ TeV, my efforts have been mainly dedicated to searches for heavy particles decaying into di-boson pairs exploring boosted topologies, where the fragmentation products of hadronically decaying vector bosons ($V \to qq'$) overlap in the detector, preventing their identification as resolved jets. These hadronic boosted signatures dominate the experimental sensitivity when looking at particles with invariant masses larger than 600 GeV.
This led me to study specific algorithms to identify boosted vector bosons decaying hadronically (V-jets), which have been also applied in other searches for new physics characterized high p_T jets originating from W/Z-boson, Higgs boson or top quark decays.
I focused on exploring the WW/WZ semi-leptonic topology using jet substructure techniques to tag the hadronically decaying W/Z bosons, interpreting the result in different BSM contexts. In particular, with no significant excess observed above the predicted background, upper limits have been produced for massive gravitons predicted by Bulk and Randall-Sundrum extra dimension models ($800 < m_{\mathrm{G}} < 2500$ GeV), as well as for a SM-like Higgs boson at high mass ($600 < m_{\mathrm{H}} < 1000$ GeV).
In the former case, a model independent result of the analysis has also been performed, to provide a generic recipe that can be used for a re-interpretation over different phenomenological models. In the latter one, the analysis has been categorised in exclusive jet bins to disentangle the gluon fusion production from the vector boson fusion (VBF) mode and the result was also interpreted in the context of a BSM heavy Higgs scenario, called electroweak singlet model, which predicts the existence of two Higgs scalars with the mass of the lighter one being around 125 GeV. A deviation with a local significance of about $2.5\,\sigma$ has been observed in the VBF category for $700 < m_{\mathrm{WW}} < 800$ GeV.
In addition, I continued to work on vector boson jet tagging in view of the LHC Run-II, with a particular emphasis on jet substructure correlations and performances in a high pileup environment. I have contributed also to the commissioning of a powerful pileup rejection method, called PileUp Per Particle Identification (PUPPI), for high p_T jets and missing energy, looking at the first Run-II data.
Then, I performed feasibility studies for the Vector Boson Scattering (VBS) in view of the CMS

detector upgrade, foreseen for the high luminosity LHC (HL-LHC) data taking. If no evidence for new physics will be observed with Run-II collisions, this rare process would become crucial to search for BSM physics via precision electroweak measurements.
Two different scattering topologies have been considered represented by the scattering of same sign W-boson pairs and the fully leptonic WZ boson one. The future CMS sensitivity to the longitudinal electroweak scattering, to partial unitarized scenarios and to new physics in the EWSB sector, parametrized through anomalies in the vector boson self couplings, has been investigated. In addition, the difference in the performance between the proposed upgraded CMS detector and the present one, which will undergo a strong degradation due to radiation damage, has been studied showing how the upgraded technology allows better performances in all the considered benchmarks.

Although no clear hints of BSM physics have been observed during Run-I, a number of puzzles, such as the origin of dark matter, the matter-antimatter asymmetry, the hierarchy problem and the electroweak vacuum stability, suggest that the SM cannot be considered the ultimate theory of nature. The LHC Run-II represents a unique opportunity for particle physics, where the machine will give access to an energy range never explored before, since the beam collisions will happen at a centre of mass energy of 13 TeV, allowing searches for new physics over a broad phenomenology, as well as to continue the precise measurements of SM processes.

Bibliography

[1] ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group Collaboration, "Precision electroweak measurements on the Z resonance", *Phys.Rept.* **427** (2006) 257–454, arXiv:hep-ex/0509008. doi:10.1016/j.physrep.2005.12.006.

[2] CMS Collaboration, "Measurement of inclusive W and Z boson production cross sections in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Rev. Lett.* **112** (2014) 191802, arXiv:1402.0923. doi:10.1103/PhysRevLett.112.191802.

[3] CMS Collaboration, "Measurements of jet multiplicity and differential production cross sections of Z+jets events in proton-proton collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev.* **D91** (2015), no. 5, 052008, arXiv:1408.3104. doi:10.1103/PhysRevD.91.052008.

[4] CMS Collaboration, "Differential cross section measurements for the production of a W boson in association with jets in proton–proton collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett.* **B741** (2015) 12–37, arXiv:1406.7533. doi:10.1016/j.physletb.2014.12.003.

[5] CMS Collaboration, "Measurement of W^+W^- and ZZ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Lett.* **B721** (2013) 190–211, arXiv:1301.4698. doi:10.1016/j.physletb.2013.03.027.

[6] A. Lesov, "The Weak Force: From Fermi to Feynman", *ArXiv e-prints* (October, 2009) arXiv:0911.0058.

[7] L.B. Okun, "Leptons and Quarks". North-Holland, Moscow, 1980.

[8] Michael E. Peskin and Daniel V. Schroeder, "An introduction to quantum field theory". Addison-Wesley Publishing Company, Second printing, January 1996.

[9] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266. doi:10.1103/PhysRevLett.19.1264.

[10] S. L. Glashow, "Partial-symmetries of weak interactions", *Nuclear Physics* **22** (1961), no. 4, 579 – 588. doi:http://dx.doi.org/10.1016/0029-5582(61)90469-2.

[11] CMS Collaboration, "Measurement of the $pp \to ZZ$ production cross section and constraints on anomalous triple gauge couplings in four-lepton final states at $\sqrt{s}$ =8 TeV", *Phys. Lett.* **B740** (2015) 250–272, arXiv:1406.0113. doi:10.1016/j.physletb.2014.11.059.

[12] CMS Collaboration, "Search for $WW\gamma$ and $WZ\gamma$ production and constraints on anomalous quartic gauge couplings in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Rev.* **D90** (2014), no. 3, 032008, arXiv:1404.4619. doi:10.1103/PhysRevD.90.032008.

[13] F. Mandl and G. Shawn, "Quantum field theory". John Wiley & Sons, Reprinted, April 1986.

[14] Particle Data Group Collaboration, "Review of Particle Physics", *Chin.Phys.* **C38** (2014) 090001. doi:10.1088/1674-1137/38/9/090001.

[15] James D. Bjorken and Sidney D. Drell, "Relativistic Quantum Fields". McGraw-Hill Book Company, 1965.

[16] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323. doi:10.1103/PhysRevLett.13.321.

[17] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509. doi:10.1103/PhysRevLett.13.508.

[18] P. Higgs, "Broken symmetries, massless particles and gauge fields", *Physics Letters* **12** (1964), no. 2, 132 – 133. doi:http://dx.doi.org/10.1016/0031-9163(64)91136-9.

[19] T. W. B. Kibble, "Symmetry Breaking in Non-Abelian Gauge Theories", *Phys. Rev.* **155** (Mar, 1967) 1554–1561. doi:10.1103/PhysRev.155.1554.

[20] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries", *Phys. Rev.* **127** (Aug, 1962) 965–970. doi:10.1103/PhysRev.127.965.

[21] Belle Collaboration, "Measurement of CKM elements and the unitarity triangle", *eConf* **C030626** (2003) FRAT04, arXiv:hep-ex/0308060.

[22] A. Falkowski, S. Pokorski, and J. Roberts, "Modelling strong interactions and longitudinally polarized vector boson scattering", *JHEP* **0712** (2007) 063, arXiv:0705.4653. doi:10.1088/1126-6708/2007/12/063.

[23] J. Elias-Miro, J. R. Espinosa, G. F. Giudice et al., "Stabilization of the Electroweak Vacuum by a Scalar Threshold Effect", *JHEP* **1206** (2012) 031, arXiv:1203.0237. doi:10.1007/JHEP06(2012)031.

[24] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", arXiv:1101.0593. doi:10.5170/CERN-2011-002.

[25] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30, arXiv:1207.7235. doi:10.1016/j.physletb.2012.08.021.

[26] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1, arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020.

[27] CMS Collaboration, "Observation of the diphoton decay of the 125 GeV Higgs boson and measurement of its properties", *Eur. Phys. J. C* **74** (2014) 3076, arXiv:1407.0558. doi:10.1140/epjc/s10052-014-3076-z.

[28] CMS Collaboration, "Measurement of the properties of a Higgs boson in the four-lepton final state", *Phys. Rev. D* **89** (2014) 092007, arXiv:1312.5353. doi:10.1103/PhysRevD.89.092007.

[29] CMS Collaboration, "Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states", *J. High Energy Phys.* **01** (2014) 096, arXiv:1312.1129. doi:10.1007/JHEP01(2014)096.

[30] CMS Collaboration, "Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks", *Phys. Rev. D* **89** (2014) 012003, arXiv:1310.3687. doi:10.1103/PhysRevD.89.012003.

[31] CMS Collaboration, "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons", *J. High Energy Phys.* **05** (2014) 104, arXiv:1401.5041. doi:10.1007/JHEP05(2014)104.

[32] CMS Collaboration, "Study of the Mass and Spin-Parity of the Higgs Boson Candidate via Its Decays to Z Boson Pairs", *Phys. Rev. Lett.* **110** (2013) 081803, arXiv:1212.6639. doi:10.1103/PhysRevLett.110.081803.

[33] CMS Collaboration, "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV", (2014). arXiv:1411.3441. Submitted for publication in *Phys. Rev. D*.

[34] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV", *Eur. Phys. J.* **C75** (2015), no. 5, 212, arXiv:1412.8662. doi:10.1140/epjc/s10052-015-3351-7.

[35] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", arXiv:1307.1347. doi:10.5170/CERN-2013-004.

[36] M. Veltman, "Limit on mass differences in the Weinberg model", *Nuclear Physics B* **123** (1977), no. 1, 89 – 99. doi:http://dx.doi.org/10.1016/0550-3213(77)90342-X.

[37] B. Patt and F. Wilczek, "Higgs-field portal into hidden sectors", arXiv:hep-ph/0605188.

[38] V. Barger, P. Langacker, M. McCaskey et al., "CERN LHC phenomenology of an extended standard model with a real scalar singlet", *Phys. Rev. D* **77** (Feb, 2008) 035005. doi:10.1103/PhysRevD.77.035005.

[39] M. Bowen, Y. Cui, and J. D. Wells, "Narrow trans-TeV Higgs bosons and H —¿ hh decays: Two LHC search paths for a hidden sector Higgs boson", *JHEP* **03** (2007) 036, arXiv:hep-ph/0701035. doi:10.1088/1126-6708/2007/03/036.

[40] S. Bock, R. Lafaye, T. Plehn et al., "Measuring hidden Higgs and strongly-interacting Higgs scenarios", *Physics Letters B* **694** (2010), no. 1, 44 – 53. doi:http://dx.doi.org/10.1016/j.physletb.2010.09.032.

[41] T. Robens and T. Stefaniak, "Status of the Higgs Singlet Extension of the Standard Model after LHC Run 1", *Eur.Phys.J.* **C75** (2015), no. 3, 104, arXiv:1501.02234. doi:10.1140/epjc/s10052-015-3323-y.

[42] G. Branco, P. Ferreira, L. Lavoura et al., "Theory and phenomenology of two-Higgs-doublet models", *Physics Reports* **516** (2012), no. 1–2, 1 – 102. Theory and phenomenology of two-Higgs-doublet models. doi:http://dx.doi.org/10.1016/j.physrep.2012.02.002.

[43] N. Craig and S. Thomas, "Exclusive signals of an extended Higgs sector", *Journal of High Energy Physics* **2012** (2012), no. 11,. doi:10.1007/JHEP11(2012)083.

[44] L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension", *Phys.Rev.Lett.* **83** (1999) 3370–3373, arXiv:hep-ph/9905221. doi:10.1103/PhysRevLett.83.3370.

[45] L. Randall and R. Sundrum, "An Alternative to compactification", *Phys.Rev.Lett.* **83** (1999) 4690–4693, arXiv:hep-th/9906064. doi:10.1103/PhysRevLett.83.4690.

[46] A. Oliveira, "Gravity particles from Warped Extra Dimensions, a review. Part I - KK Graviton", arXiv:1404.0102.

[47] ATLAS Collaboration, "Evidence for Electroweak Production of $W^{\pm}W^{\pm}jj$ in pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector", *Phys.Rev.Lett.* **113** (2014), no. 14, 141803, arXiv:1405.6241. doi:10.1103/PhysRevLett.113.141803.

[48] CMS Collaboration, "Study of vector boson scattering and search for new physics in events with two same-sign leptons and two jets", *Phys.Rev.Lett.* **114** (2015), no. 5, 051801, arXiv:1410.6315. doi:10.1103/PhysRevLett.114.051801.

[49] B. W. Lee, C. Quigg, and H. B. Thacker, "Strength of Weak Interactions at Very High Energies and the Higgs Boson Mass", *Phys. Rev. Lett.* **38** (Apr, 1977) 883–885. doi:10.1103/PhysRevLett.38.883.

[50] D. Espriu and B. Yencho, "Longitudinal WW scattering in light of the "Higgs boson" discovery", *Phys.Rev.* **D87** (2013), no. 5, 055017, arXiv:1212.4158. doi:10.1103/PhysRevD.87.055017.

[51] J. Chang, K. Cheung, C.-T. Lu et al., "WW scattering in the era of post-Higgs-boson discovery", *Phys.Rev.* **D87** (2013) 093005, arXiv:1303.6335. doi:10.1103/PhysRevD.87.093005.

[52] K. Cheung, C.-W. Chiang, and T.-C. Yuan, "Partially Strong WW Scattering", *Phys.Rev.* **D78** (2008) 051701, arXiv:0803.2661. doi:10.1103/PhysRevD.78.051701.

[53] ATLAS Collaboration, "Physics at a High-Luminosity LHC with ATLAS", arXiv:1307.7292.

[54] OPAL Collaboration, "Measurement of triple gauge boson couplings from W^+W^- production at LEP energies up to 189-GeV", *Eur. Phys. J.* **C19** (2001) 1–14, arXiv:hep-ex/0009022. doi:10.1007/s100520100597.

[55] DELPHI Collaboration, "Study of triple-gauge-boson couplings ZZZ, ZZγ and Z$\gamma\gamma$ LEP", *Eur. Phys. J.* **C51** (2007) 525–542, arXiv:0706.2741. doi:10.1140/epjc/s10052-007-0345-0.

[56] CDF Collaboration, "Measurement of the WZ cross section and triple gauge couplings in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev.* **D86** (2012) 031104, arXiv:1202.6629. doi:10.1103/PhysRevD.86.031104.

[57] CDF Collaboration, "Measurement of the W^+W^- production cross section and search for anomalous $WW\gamma$ and WWZ couplings in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. Lett.* **104** (2010) 201801, arXiv:0912.4500. doi:10.1103/PhysRevLett.104.201801.

[58] CMS Collaboration, "Measurement of the W^+W^- cross section in pp collisions at $\sqrt{s}$ = 8 TeV and limits on anomalous gauge couplings", arXiv:1507.03268.

[59] CMS Collaboration, "Measurement of the sum of WW and WZ production with W+dijet events in pp collisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J.* **C73** (2013), no. 2, 2283, arXiv:1210.7544. doi:10.1140/epjc/s10052-013-2283-3.

[60] CMS Collaboration, "Measurement of the $Z\gamma$ production cross section in pp collisions at 8 TeV and search for anomalous triple gauge boson couplings", *JHEP* **04** (2015) 164, arXiv:1502.05664. doi:10.1007/JHEP04(2015)164.

[61] O. Eboli, M. Gonzalez-Garcia, and J. Mizukoshi, "$pp \to jje^{\pm}\mu^{\pm}\nu\nu$ and $jje^{\pm}\mu^{\mp} + \nu\nu$ at O(α(em)^6) and O(α(em)^4 α(s)^2) for the study of the quartic electroweak gauge boson vertex at CERN LHC", *Phys.Rev.* **D74** (2006) 073005, arXiv:hep-ph/0606118. doi:10.1103/PhysRevD.74.073005.

[62] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001. doi:10.1088/1748-0221/3/08/S08001.

[63] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[64] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003. doi:10.1088/1748-0221/3/08/S08003.

[65] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002. doi:10.1088/1748-0221/3/08/S08002.

[66] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005. doi:10.1088/1748-0221/3/08/S08005.

[67] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl.Phys.* **B126** (1977) 298. doi:10.1016/0550-3213(77)90384-4.

[68] S. Alekhin, S. Alioli, R. D. Ball et al., "The PDF4LHC Working Group Interim Report", arXiv:1101.0536.

[69] CMS Collaboration, "The CMS magnet project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997. https://cds.cern.ch/record/331056.

[70] CMS Collaboration, V. Karimäki, M. Mannelli, P. Siegrist et al., "The CMS tracker system project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997. https://cds.cern.ch/record/368412.

[71] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, arXiv:1405.6569. doi:10.1088/1748-0221/9/10/P10009.

[72] CMS Collaboration, S. Chatrchyan et al., "The CMS electromagnetic calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997. http://cds.cern.ch/record/349375.

[73] CMS Collaboration, S. Chatrchyan et al., "The CMS hadron calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997. https://cds.cern.ch/record/357153.

[74] CMS Collaboration, "The CMS muon project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997. https://cds.cern.ch/record/343814.

[75] CMS Collaboration, "CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems", technical report. http://cds.cern.ch/record/706847.

[76] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, arXiv:1206.4071. doi:10.1088/1748-0221/7/10/P10002.

[77] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl.Instrum.Meth.* **A262** (1987) 444–450. doi:10.1016/0168-9002(87)90887-4.

[78] P. Adzic, R. Alemany-Fernandez, C. Almeida et al., "Energy resolution of the barrel of the CMS electromagnetic calorimeter", *JINST* **2** (2007) P04004. doi:10.1088/1748-0221/2/04/P04004.

[79] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", arXiv:1502.02701.

[80] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET",. http://cds.cern.ch/record/1194487.

[81] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* **0804** (2008) 063, arXiv:0802.1189. doi:10.1088/1126-6708/2008/04/063.

[82] CMS Collaboration, "Pileup Jet Identification", Technical Report CMS-PAS-JME-13-005, CERN, Geneva, (2013). http://cds.cern.ch/record/1581583".

[83] CMS Collaboration, "Pileup Removal Algorithms", Technical Report CMS-PAS-JME-14-001, CERN, Geneva, (2014). https://cds.cern.ch/record/1751454.

[84] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, arXiv:1107.4277. doi:10.1088/1748-0221/6/11/P11002.

[85] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys.Lett.* **B659** (2008) 119–126, arXiv:0707.1378. doi:10.1016/j.physletb.2007.09.077.

[86] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 02, P02006, arXiv:1411.0511. doi:10.1088/1748-0221/10/02/P02006.

[87] J. M. Butterworth, A. R. Davison, M. Rubin et al., "Jet substructure as a new Higgs search channel at the LHC", *Phys. Rev. Lett.* **100** (2008) 242001, arXiv:0802.2470. doi:10.1103/PhysRevLett.100.242001.

[88] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Techniques for improved heavy particle searches with jet substructure", *Phys. Rev.* **D80** (2009) 051501, arXiv:0903.5081. doi:10.1103/PhysRevD.80.051501.

[89] CMS Collaboration, "Studies of jet mass in dijet and W/Z + jet events", *JHEP* **1305** (2013) 090, arXiv:1303.4811. doi:10.1007/JHEP05(2013)090.

[90] CMS Collaboration, "Jet Substructure Algorithms", Technical Report CMS-PAS-JME-10-013, CERN, Geneva, (2011). https://cds.cern.ch/record/1333700.

[91] ATLAS Collaboration, "Jet mass and substructure of inclusive jets in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS experiment", *JHEP* **1205** (2012) 128, arXiv:1203.4606. doi:10.1007/JHEP05(2012)128.

[92] ATLAS Collaboration, "Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector", *JHEP* **1309** (2013) 076, arXiv:1306.4945. doi:10.1007/JHEP09(2013)076.

[93] CMS Collaboration, "Identification techniques for highly boosted W bosons that decay into hadrons", *JHEP* **1412** (2014) 017, arXiv:1410.4227. doi:10.1007/JHEP12(2014)017.

[94] CMS Collaboration, "V Tagging Observables and Correlations", Technical Report CMS-PAS-JME-14-002, CERN, Geneva, (2014). https://cds.cern.ch/record/1754913".

[95] S. Bentvelsen and I. Meyer, "The Cambridge jet algorithm: Features and applications", *Eur.Phys.J.* **C4** (1998) 623–629, arXiv:hep-ph/9803322. doi:10.1007/s100520050232.

[96] D. Krohn, J. Thaler, and L.-T. Wang, "Jet Trimming", *JHEP* **1002** (2010) 084, arXiv:0912.1342. doi:10.1007/JHEP02(2010)084.

[97] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches", *Phys.Rev.* **D81** (2010) 094023, arXiv:0912.0033. doi:10.1103/PhysRevD.81.094023.

[98] A. J. Larkoski, S. Marzani, G. Soyez et al., "Soft Drop", *JHEP* **1405** (2014) 146, arXiv:1402.2657. doi:10.1007/JHEP05(2014)146.

[99] M. Dasgupta, A. Fregoso, S. Marzani et al., "Towards an understanding of jet substructure", *JHEP* **1309** (2013) 029, arXiv:1307.0007. doi:10.1007/JHEP09(2013)029.

[100] A. J. Larkoski and J. Thaler, "Unsafe but Calculable: Ratios of Angularities in Perturbative QCD", *JHEP* **1309** (2013) 137, arXiv:1307.1699. doi:10.1007/JHEP09(2013)137.

[101] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", *JHEP* **1103** (2011) 015, arXiv:1011.2268. doi:10.1007/JHEP03(2011)015.

[102] J. Thaler and K. Van Tilburg, "Maximizing Boosted Top Identification by Minimizing N-subjettiness", *JHEP* **1202** (2012) 093, arXiv:1108.2701. doi:10.1007/JHEP02(2012)093.

[103] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys.Rev.* **D48** (1993) 3160–3166, arXiv:hep-ph/9305266. doi:10.1103/PhysRevD.48.3160.

[104] S. D. Ellis, A. Hornig, T. S. Roy et al., "Qjets: A Non-Deterministic Approach to Tree-Based Jet Substructure", *Phys.Rev.Lett.* **108** (2012) 182003, arXiv:1201.1914. doi:10.1103/PhysRevLett.108.182003.

[105] A. J. Larkoski, G. P. Salam, and J. Thaler, "Energy Correlation Functions for Jet Substructure", *JHEP* **1306** (2013) 108, arXiv:1305.0007. doi:10.1007/JHEP06(2013)108.

[106] D. Krohn, M. D. Schwartz, T. Lin et al., "Jet Charge at the LHC", *Phys.Rev.Lett.* **110** (2013), no. 21, 212001, arXiv:1209.2421. doi:10.1103/PhysRevLett.110.212001.

[107] CMS Collaboration, "Performance of quark/gluon discrimination in 8 TeV pp data", Technical Report CMS-PAS-JME-13-002, CERN, Geneva, (2013). https://cds.cern.ch/record/1599732".

[108] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", Technical Report CMS-PAS-LUM-13-001, CERN, Geneva, (2013). https://cds.cern.ch/record/1598864.

[109] S. Bolognesi, Y. Gao, A. V. Gritsan et al., "On the spin and parity of a single-produced resonance at the LHC", *Phys.Rev.* **D86** (2012) 095031, arXiv:1208.4018. doi:10.1103/PhysRevD.86.095031.

[110] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **0711** (2007) 070, arXiv:0709.2092. doi:10.1088/1126-6708/2007/11/070.

[111] S. Alioli, P. Nason, C. Oleari et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **1006** (2010) 043, arXiv:1002.2581. doi:10.1007/JHEP06(2010)043.

[112] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **0605** (2006) 026, arXiv:hep-ph/0603175. doi:10.1088/1126-6708/2006/05/026.

[113] CMS Collaboration, "Study of the underlying event at forward rapidity in pp collisions at $\sqrt{s} = 0.9$, 2.76, and 7 TeV", *JHEP* **1304** (2013) 072, arXiv:1302.2394. doi:10.1007/JHEP04(2013)072.

[114] J. Alwall, R. Frederix, S. Frixione et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **1407** (2014) 079, arXiv:1405.0301. doi:10.1007/JHEP07(2014)079.

[115] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput.Phys.Commun.* **178** (2008) 852–867, arXiv:0710.3820. doi:10.1016/j.cpc.2008.01.036.

[116] R. Corke and T. Sjostrand, "Multiparton Interactions with an x-dependent Proton Size", *JHEP* **1105** (2011) 009, arXiv:1101.5953. doi:10.1007/JHEP05(2011)009.

[117] J. Pumplin, D. Stump, J. Huston et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **0207** (2002) 012, arXiv:hep-ph/0201195. doi:10.1088/1126-6708/2002/07/012.

[118] A. D. Martin, R. Roberts, W. Stirling et al., "MRST2001: Partons and α_s from precise deep inelastic scattering and Tevatron jet data", *Eur.Phys.J.* **C23** (2002) 73–87, arXiv:hep-ph/0110215. doi:10.1007/s100520100842.

[119] H.-L. Lai, M. Guzzi, J. Huston et al., "New parton distributions for collider physics", *Phys.Rev.* **D82** (2010) 074024, arXiv:1007.2241. doi:10.1103/PhysRevD.82.074024.

[120] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **0206** (2002) 029, arXiv:hep-ph/0204244. doi:10.1088/1126-6708/2002/06/029.

[121] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl.Instrum.Meth.* **A506** (2003) 250–303. doi:10.1016/S0168-9002(03)01368-8.

[122] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur.Phys.J.* **C72** (2012) 1896, arXiv:1111.6097. doi:10.1140/epjc/s10052-012-1896-2.

[123] CMS Collaboration, "Search for massive resonances in dijet systems containing jets tagged as W or Z boson decays in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **1408** (2014) 173, arXiv:1405.1994. doi:10.1007/JHEP08(2014)173.

[124] CMS Collaboration, "Search for leptonic decays of W ' bosons in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **1208** (2012) 023, arXiv:1204.4764. doi:10.1007/JHEP08(2012)023.

[125] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, arXiv:1211.4462. doi:10.1088/1748-0221/8/04/P04013.

[126] CMS Collaboration, "Search for anomalous $t\bar{t}$ production in the highly-boosted all-hadronic final state", *JHEP* **1209** (2012) 029, arXiv:1204.2488. doi:10.1007/JHEP09(2012)029, 10.1007/JHEP03(2014)132.

[127] CMS Collaboration, "Search for heavy resonances in the W/Z-tagged dijet mass spectrum in pp collisions at 7 TeV", *Phys.Lett.* **B723** (2013) 280–301, arXiv:1212.1910. doi:10.1016/j.physletb.2013.05.040.

[128] A. Hoecker, P. Speckmayer, J. Stelzer et al., "TMVA: Toolkit for Multivariate Data Analysis", *PoS* **ACAT** (2007) 040, arXiv:physics/0703039.

[129] ATLAS Collaboration Collaboration, "Charged particle multiplicities in p p interactions at $\sqrt{s} = 0.9$ and 7 TeV in a diffractive limited phase-space measured with the ATLAS detector at the LHC and new PYTHIA6 tune", Technical Report ATLAS-CONF-2010-031, CERN, Geneva, (Jul, 2010). http://cds.cern.ch/record/1277665.

[130] "ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11", Technical Report ATL-PHYS-PUB-2011-009, CERN, Geneva, (Jul, 2011). https://cds.cern.ch/record/1363300.

[131] CMS Collaboration, "Measurement of the W-boson helicity in top-quark decays from $t\bar{t}$ production in lepton+jets events in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **1310** (2013) 167, arXiv:1308.3879. doi:10.1007/JHEP10(2013)167.

[132] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7$ TeV", *Phys.Lett.* **B722** (2013) 5–27, arXiv:1210.6718. doi:10.1016/j.physletb.2013.03.024.

[133] M. Cacciari, G. P. Salam, and G. Soyez, "On the use of charged-track information to subtract neutral pileup", arXiv:1404.7353.

[134] G. Soyez, G. P. Salam, J. Kim et al., "Pileup subtraction for jet shapes", *Phys.Rev.Lett.* **110** (2013), no. 16, 162001, arXiv:1211.2811. doi:10.1103/PhysRevLett.110.162001.

[135] P. Berta, M. Spousta, D. W. Miller et al., "Particle-level pileup subtraction for jets and jet shapes", *JHEP* **1406** (2014) 092, arXiv:1403.3108. doi:10.1007/JHEP06(2014)092.

[136] D. Bertolini, P. Harris, M. Low et al., "Pileup Per Particle Identification", *JHEP* **10** (2014) 59, arXiv:1407.6013. doi:10.1007/JHEP10(2014)059.

[137] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, arXiv:1410.8849. doi:10.1007/JHEP04(2015)040.

[138] CMS Collaboration, "Search for massive resonances decaying into pairs of boosted bosons in semi-leptonic final states at $\sqrt{s} = 8$ TeV", *JHEP* **1408** (2014) 174, arXiv:1405.3447. doi:10.1007/JHEP08(2014)174.

[139] A. Belyaev, N. D. Christensen, and A. Pukhov, "CalcHEP 3.4 for collider physics within and beyond the Standard Model", *Comput.Phys.Commun.* **184** (2013) 1729–1769, arXiv:1207.6082. doi:10.1016/j.cpc.2013.01.014.

[140] CMS Collaboration, "Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV", *JHEP* **1101** (2011) 080, arXiv:1012.2466. doi:10.1007/JHEP01(2011)080.

[141] CMS Collaboration, "Trigger strategies for Higgs searches in 2011", CMS Analysis Note CMS-AN-2011/065, (2011).

[142] J. Beringer, J. F. Arguin, R. M. Barnett et al., "Review of Particle Physics*", *Phys. Rev. D* **86** (Jul, 2012) 010001. doi:10.1103/PhysRevD.86.010001.

[143] M. Gouzevitch, A. Oliveira, J. Rojo et al., "Scale-invariant resonance tagging in multijet events and new physics in Higgs pair production", *JHEP* **1307** (2013) 148, arXiv:1303.6636. doi:10.1007/JHEP07(2013)148.

[144] LHCb Collaboration, "Observation of J/ψ pair production in pp collisions at $\sqrt{s} = 7\ TeV$", *Phys. Lett.* **B707** (2012) 52–59, arXiv:1109.0963. doi:10.1016/j.physletb.2011.12.015.

[145] CMS Collaboration, "Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events", Technical Report CMS-PAS-BTV-13-001, CERN, Geneva, (2013). "https://cds.cern.ch/record/1581306".

[146] M. L. Tiku, "Power Function of the F-Test Under Non-Normal Situations", *Journal of the American Statistical Association* **66** (1971), no. 336, 913–916. doi:10.1080/01621459.1971.10482371.

[147] T. C. collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *Journal of Instrumentation* **7** (2012), no. 10, P10002. http://stacks.iop.org/1748-0221/7/i=10/a=P10002.

[148] F. Garwood, "Fiducial Limits for the Poisson Distribution", *Biometrika* **28** (1936) 437. doi:10.1093/biomet/28.3-4.437.

[149] A. L. Read, "Presentation of search results: the CLs technique", *Journal of Physics G: Nuclear and Particle Physics* **28** (2002), no. 10, 2693. http://stacks.iop.org/0954-3899/28/i=10/a=313.

[150] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth.* **A434** (1999) 435–443, arXiv:hep-ex/9902006. doi:10.1016/S0168-9002(99)00498-2.

[151] CMS Collaboration, "Search for narrow resonances using the dijet mass spectrum in pp collisions at $\sqrt{s}$=8 TeV", *Phys.Rev.* **D87** (2013), no. 11, 114015, arXiv:1302.4794. doi:10.1103/PhysRevD.87.114015.

[152] G. Cowan, K. Cranmer, E. Gross et al., "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J.* **C71** (2011) 1554, arXiv:1007.1727. [Erratum: Eur. Phys. J.C73,2501(2013)]. doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z.

[153] G. Passarino, "{WW} scattering and perturbative unitarity", *Nuclear Physics B* **343** (1990), no. 1, 31 – 59. doi:http://dx.doi.org/10.1016/0550-3213(90)90593-3.

[154] M. S. Chanowitz and M. K. Gaillard, "The TeV physics of strongly interacting W's and Z's", *Nuclear Physics B* **261** (1985) 379 – 431. doi:http://dx.doi.org/10.1016/0550-3213(85)90580-2.

[155] A. Ballestrero, G. Bevilacqua, D. B. Franzosi et al., "How well can the LHC distinguish between the SM light Higgs scenario, a composite Higgs and the Higgsless case using VV scattering channels?", *JHEP* **11** (2009) 126, arXiv:0909.3838. doi:10.1088/1126-6708/2009/11/126.

[156] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett.* **B716** (2012) 1–29, arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020.

[157] CMS Collaboration, "Search for a standard-model-like Higgs boson with a mass in the range 145 to 1000 GeV at the LHC", *Eur.Phys.J.* **C73** (2013) 2469, arXiv:1304.0213. doi:10.1140/epjc/s10052-013-2469-8.

[158] CMS Collaboration, "Search for a Standard Model-like Higgs boson decaying into WW to $\ell\nu q\overline{q}$ in exclusive jet bins in pp collisions at $\sqrt{s} = 8$ TeV", Technical Report CMS-PAS-HIG-14-008, CERN, Geneva, (2015). https://cds.cern.ch/record/2001192.

[159] CMS Collaboration, "Search for a Higgs Boson in the Mass Range from 145 to 1000 GeV Decaying to a Pair of W or Z Bosons", arXiv:1504.00936.

[160] A. Ballestrero, A. Belhouari, G. Bevilacqua et al., "PHANTOM: A Monte Carlo event generator for six parton final states at high energy colliders", *Comput.Phys.Commun.* **180** (2009) 401–417, arXiv:0801.3359. doi:10.1016/j.cpc.2008.10.005.

[161] N. Kauer and G. Passarino, "Inadequacy of zero-width approximation for a light Higgs boson signal", *JHEP* **08** (2012) 116, arXiv:1206.4803. doi:10.1007/JHEP08(2012)116.

[162] G. Passarino, C. Sturm, and S. Uccirati, "Higgs Pseudo-Observables, Second Riemann Sheet and All That", *Nucl. Phys.* **B834** (2010) 77–115, arXiv:1001.3360. doi:10.1016/j.nuclphysb.2010.03.013.

[163] S. Goria, G. Passarino, and D. Rosco, "The Higgs Boson Lineshape", *Nucl. Phys.* **B864** (2012) 530–579, arXiv:1112.5517. doi:10.1016/j.nuclphysb.2012.07.006.

[164] J. M. Campbell, R. K. Ellis, and C. Williams, "Gluon-Gluon Contributions to W+ W- Production and Higgs Interference Effects", *JHEP* **10** (2011) 005, arXiv:1107.5569. doi:10.1007/JHEP10(2011)005.

[165] CMS Collaboration Collaboration, "Search for a Standard Model-like Higgs boson decaying into WW to $\ell\nu q\overline{q}$ in pp collisions at $\sqrt{s} = 8$ TeV", Technical Report CMS-PAS-HIG-13-008, CERN, Geneva, (2013). http://cds.cern.ch/record/1546778.

[166] CMS Collaboration, "Technical Proposal for the Phase-II Upgrade of the CMS Detector",. http://cds.cern.ch/record/2020886.

[167] DELPHES 3 Collaboration, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment", *JHEP* **02** (2014) 057, arXiv:1307.6346. doi:10.1007/JHEP02(2014)057.

[168] C. Collaboration, "Technical proposal for the upgrade of the CMS detector through 2020", Technical Report CERN-LHCC-2011-006. LHCC-P-004, CERN, Geneva, (Jun, 2011). https://cds.cern.ch/record/1355706.

[169] CMS Collaboration, "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in *pp* Collisions at $\sqrt{s} = 7$ TeV", *JINST* **8** (2013) P09009, arXiv:1306.2016. doi:10.1088/1748-0221/8/09/P09009.

[170] M. Cacciari, G. P. Salam, and G. Soyez, "SoftKiller, a particle-level pileup removal method", *Eur. Phys. J.* **C75** (2015), no. 2, 59, arXiv:1407.0408. doi:10.1140/epjc/s10052-015-3267-2.

[171] CMS Collaboration, "Underlying Event Tunes and Double Parton Scattering", Technical Report CMS-PAS-GEN-14-001, CERN, Geneva, (2014). https://cds.cern.ch/record/1697700.

[172] C. Englert, E. Re, and M. Spannowsky, "Triplet Higgs boson collider phenomenology after the LHC", *Phys. Rev.* **D87** (2013), no. 9, 095014, arXiv:1302.6505. doi:10.1103/PhysRevD.87.095014.

[173] C. Englert, E. Re, and M. Spannowsky, "Pinning down Higgs triplets at the LHC", *Phys. Rev.* **D88** (2013) 035024, arXiv:1306.6228. doi:10.1103/PhysRevD.88.035024.

[174] N. Kauer, T. Plehn, D. L. Rainwater et al., "$H \to W^+W^-$ as the discovery mode for a light Higgs boson", *Phys. Lett.* **B503** (2001) 113–120, arXiv:hep-ph/0012351. doi:10.1016/S0370-2693(01)00211-8.

[175] K. Doroba, J. Kalinowski, J. Kuczmarski et al., "The W_LW_L Scattering at the LHC: Improving the Selection Criteria", *Phys. Rev.* **D86** (2012) 036011, arXiv:1201.2768. doi:10.1103/PhysRevD.86.036011.

[176] B. Jager, C. Oleari, and D. Zeppenfeld, "Next-to-leading order QCD corrections to W+ W+ jj and W- W- jj production via weak-boson fusion", *Phys. Rev.* **D80** (2009) 034022, arXiv:0907.0580. doi:10.1103/PhysRevD.80.034022.

[177] G. Bozzi, B. Jager, C. Oleari et al., "Next-to-leading order QCD corrections to W+ Z and W- Z production via vector-boson fusion", *Phys. Rev.* **D75** (2007) 073004, arXiv:hep-ph/0701105. doi:10.1103/PhysRevD.75.073004.

[178] CMS Collaboration, "Vector Boson Scattering and Quartic Gauge Coupling Studies in WZ Production at 14 TeV", Technical Report CMS-PAS-FTR-13-006, CERN, Geneva, (2013). https://cds.cern.ch/record/1606835.

[179] C. Degrande, N. Greiner, W. Kilian et al., "Effective Field Theory: A Modern Approach to Anomalous Couplings", *Annals Phys.* **335** (2013) 21–32, arXiv:1205.4231. doi:10.1016/j.aop.2013.04.016.

[180] CMS Collaboration, "Search for the standard model Higgs boson produced in association with top quarks in multilepton final states", Technical Report CMS-PAS-HIG-13-020, CERN, Geneva, (2013). http://cds.cern.ch/record/1604480.

[181] CMS Collaboration, "Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states", *JHEP* **01** (2014) 096, arXiv:1312.1129. doi:10.1007/JHEP01(2014)096.

[182] J. S. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra", in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*. 2011. arXiv:1103.0354.

[183] CMS Collaboration, "Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett.* **B710** (2012) 26–48, arXiv:1202.1488. doi:10.1016/j.physletb.2012.02.064.

[184] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group Collaboration, "Procedure for the LHC Higgs boson search combination in Summer 2011", Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, (Aug, 2011). http://cds.cern.ch/record/1379837.