UNIVERSITÀ DI PADOVA
DIPARTIMENTO DI FISICA E ASTRONOMIA GALILEO GALILEI
CORSO DI DOTTORATO DI RICERCA IN FISICA
XXXIII CICLO

TESI DI DOTTORATO

A LEPTON PAIR AND JETS FINAL STATE IN CMS AND UNITARITY INTERPRETATIONS

COORDINATORE
CH.MO PROF. FRANCO SIMONETTO

RELATORI
CH.MO PROF. ROBERTO ROSSIN
CH.MA DOTT. SSA PATRIZIA AZZI

CANDIDATO
MATTEO PRESILLA

ANNO ACCADEMICO 2019/2020

DECLARATION

I, *Matteo Presilla*, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.
I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

DETAILS OF COLLABORATION AND PUBLICATIONS:
In this thesis are reported the results of research carried out within the CMS Collaboration, and documented in the internal analysis reports:

- AN-18-112 (FTR-18-006) "Search for heavy composite Majorana neutrino at the HL/HE-LHC" (main contributors: P.Azzi, M. Narain, M. Presilla, W. Zhang) - published in CERN Yellow Rep.Monogr. 7 (2019).
- AN-18-111 (EXO-20-011) "Search for heavy composite Majorana neutrino at the LHC with Run 2 data" (main contributors: P.Azzi, V. Mariani, M. Presilla).
- AN-20-076 "Search for electroweak and anomalous electroweak production of VZ boson pair plus two jets in proton-proton collisions at 13 TeV" (main contributors: P. Azzi, A. Hakimi, M. Presilla, J.B. Sauvan).

Furthermore, the phenomenological research, independent of CMS group activities, has been mainly carried out with Dr. S. Biondini. Dr. R. Leonardi, Dr. O. Panella, (and others) and is documented in:

- "Perturbative unitarity bounds for effective composite models", S. Biondini, R.Leonardi, O.Panella, M. Presilla - PLB 795 (2019) 644-649.
- "Joining LHC Phenomenology of Heavy Composite Neutrinosand Baryogenesis", S. Biondini, M. Presilla - in submission.

During the graduate program, I have also worked on the following publications, whose content is not directly reported in this manuscript:

- "Like-sign dileptons with mirror type composite neutrinos at the HL-LHC", M.Presilla, R.Leonardi, O.Panella - CERN Yellow Rep.Monogr. 7 (2019).
- "Beyond the Standard Model Physics at the HL-LHC and HE-LHC", X. Cid Vidal et al. - CERN Yellow Rep.Monogr. 7 (2019).
- "Like-sign dileptons with mirror type composite neutrinos at the HL-LHC", M.Presilla, R.Leonardi, O.Panella - CERN Yellow Rep.Monogr. 7 (2019).
- "Non-Commutativity effects in the Dirac equation in crossed electric and magnetic fields", D. Nath, M.Presilla, O.Panella, P.Roy - EPL.

Finally, some other research activity non mentioned in this thesis has been performed within the MUonE Collaboration.

ABSTRACT

This thesis addresses the study of a signature with a pair of same flavor leptons and jets with the CMS detector in the context of two different scenarios: precision physics within the Standard Model, and the effective theory of composite models. On one hand, the dilepton plus jets signature allows the selection of the Vector Boson Scattering process of a ZV in the semi-leptonic final state, where V is the vector boson decaying hadronically (V=W, Z) and Z decays in charged leptons (e, μ), in proton-proton collisions at 13 TeV. In addition, the dilepton plus jets final state can also be exploited in LHC and in future collider (HL/HE-LHC) data to search for the presence of new exotic states (Majorana neutrinos) arising from a compositeness framework. Constraints from unitarity are considered for the first time and are shown to play a fundamental role in the sensitivity of new physics searches.

CONTENTS

INTRODUCTION

New physics beyond the Standard Model (SM) is manifestly needed to address at least two open problems in fundamental physics: the existence of dark matter and the baryon asymmetry of the Universe. It is nowadays clear that the SM alone cannot provide a particle candidate or a mechanism to successfully explain either of these compelling issues. Compositeness of ordinary fermions is one possible scenario Beyond the Standard Model (BSM) that may lead to a solution of the hierarchy problem or the explanation of the proliferation of ordinary fermions. Elementary particles are thought to be bound states of some as–yet–unobserved fundamental constituents generically referred to as *preons*, and the existence of new excited states interacting with SM fermions via effective couplings is foreseen. The heavy composite Majorana neutrino N_ℓ would be a particular case of such excited states, that shows a connection with baryogenesis via thermal leptogenesis. Indeed, composite neutrinos can be produced in the thermal bath of the early universe due to their interactions with the SM particles. As is customary in an Effective Field Theory (EFT) approach, the effects of the high-energy physics scale, here Λ, are captured in higher-dimensional operators that describe processes within a lower energy domain, where the fundamental building blocks of the theory cannot show up. On general grounds, this *bottom-up* approach can parametrize the observable effects of a large class of new physics scenarios. However, it suffers from a limited validity range, i.e. up to energy/momentum scales smaller than the large energy scale that sets the operator expansion. Since collider experiments are involving more and more energetic particle collisions, the use and the applicability of effective operators can be questioned, and the translation of experimental data into a theoretical framework has to be done with particular care. Very general principles consisting of analyticity and perturbative unitarity, directly descending from NR Quantum Mechanics, come in aid as a fundamental axiom for collider physics.

Interestingly, the particle that nowadays we consider responsible for the electroweak symmetry breaking, the Higgs boson, rescues the SM from perturbative unitarity violation in scattering between vectors bosons. The Vector Boson Scattering (VBS) process represents also an ideal test bench for the spontaneous symmetry breaking mechanism at hadron colliders, giving primary access to the non-Abelian gauge structure of the electroweak interactions through quartic gauge couplings. The possibility to access the full Run 2 data collected by the Compact Muon Solenoid (CMS) detector, offers, therefore, a chance to explore the rarest and undiscovered channels, like the semileptonic VBS process, that are shown to have high sensitivity to new physics as well. As proven by the discoveries that led to the consolidation of the SM, electroweak precision physics can play a fundamental role in establishing the existence of new physics and guiding its theoretical interpretations. We will try to emphasize the relevance of a solid unitarity interpretation, and how intertwined it is with both the precision and the direct search strategies.

The outline of the thesis is as follows. In the first chapter, we provide an introduction to the theoretical framework used throughout the thesis. In Chapter 2, we present the study of partial wave unitarity bounds in the parameter space of dimension-5 and dimension-6 effective operators that arise in a compositeness scenario. After deducing the unitarity bound for the production process of a composite neutrino, we implement such a bound and compare it with the recent experimental exclusion curves. In the next chapter, we introduce the experimental setup used to collect the data, analyzed in the remaining chapters: the LHC and the CMS detector. In Chapter 4, we perform an experimental search for the heavy composite Majorana neutrino in the two same-flavor leptons (electrons or muons) at least one-large radius jet. The analysis uses proton-proton collisions at $\sqrt{s} = 13$ TeV collected by the CMS experiment at the CERN LHC from 2016 to 2018. In Chapter 5, we present another CMS Run 2 data analysis for the first search for electroweak production of VZ boson pairs and two jets. Then, we step into the future of the LHC in Chapter 6, extending the search in Chapter 4 to the High-Luminosity LHC (HL-LHC) and the High-Energy LHC (HE-LHC), and match the results from future colliders with the phase space preferred by baryogenesis. To conclude, we present the outcomes of this work at the end of the present document, underlying how the experimental and theoretical work are deeply connected and represent a key for the interpretation approach of experimental data in the future.

Part I

PHENOMENOLOGICAL STUDIES

1 THE STANDARD MODEL AND ITS EXTENSIONS

1 WHAT IS STANDARD MODEL AND WHAT IS NOT

The interactions and the fundamental constituents of the Universe are elegantly described by a mathematical consistent renormalizable theory called the Standard Model (SM) represented with the following Lagrangian density [1, 2]:

$$
\begin{gathered}
\mathcal{L}_{SM} = -\tfrac{1}{2}\partial_\nu g^a_\mu \partial_\nu g^a_\mu - g_s f^{abc}\partial_\mu g^a_\nu g^b_\mu g^c_\nu - \tfrac{1}{4} g_s^2 f^{abc} f^{ade} g^b_\mu g^c_\nu g^d_\mu g^e_\nu + \tfrac{1}{2} i g_s^2 (\bar{q}^\sigma_i \gamma^\mu q^\sigma_j) g^a_\mu + \\
\bar{G}^a \partial^2 G^a + g_s f^{abc} \partial_\mu \bar{G}^a G^b g^c_\mu - \partial_\nu W^+_\mu \partial_\nu W^-_\mu - M^2 W^+_\mu W^-_\mu - \tfrac{1}{2}\partial_\nu Z^0_\mu \partial_\nu Z^0_\mu - \\
\tfrac{1}{2c_w^2} M^2 Z^0_\mu Z^0_\mu - \tfrac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu - \tfrac{1}{2}\partial_\mu H \partial_\mu H - \tfrac{1}{2} m_h^2 H^2 - \partial_\mu \phi^+ \partial_\mu \phi^- - M^2 \phi^+ \phi^- - \\
\tfrac{1}{2}\partial_\mu \phi^0 \partial_\mu \phi^0 - \tfrac{1}{2c_w^2} M \phi^0 \phi^0 - \beta_h [\tfrac{2M^2}{g^2} + \tfrac{2M}{g} H + \tfrac{1}{2}(H^2 + \phi^0\phi^0 + 2\phi^+\phi^-)] + \tfrac{2M^4}{g^2}\alpha_h - \\
igc_w[\partial_\nu Z^0_\mu (W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - Z^0_\nu (W^+_\mu \partial_\nu W^-_\mu - W^-_\mu \partial_\nu W^+_\mu) + Z^0_\mu (W^+_\nu \partial_\nu W^-_\mu - \\
W^-_\nu \partial_\nu W^+_\mu)] - igs_w[\partial_\nu A_\mu (W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - A_\nu (W^+_\mu \partial_\nu W^-_\mu - W^-_\mu \partial_\nu W^+_\mu) + \\
A_\mu (W^+_\nu \partial_\nu W^-_\mu - W^-_\nu \partial_\nu W^+_\mu)] - \tfrac{1}{2} g^2 W^+_\mu W^-_\mu W^+_\nu W^-_\nu + \tfrac{1}{2} g^2 W^+_\mu W^-_\nu W^+_\mu W^-_\nu + \\
g^2 c_w^2 (Z^0_\mu W^+_\mu Z^0_\nu W^-_\nu - Z^0_\mu Z^0_\mu W^+_\nu W^-_\nu) + g^2 s_w^2 (A_\mu W^+_\mu A_\nu W^-_\nu - A_\mu A_\mu W^+_\nu W^-_\nu) + \\
g^2 s_w c_w [A_\mu Z^0_\nu (W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - 2 A_\mu Z^0_\mu W^+_\nu W^-_\nu] - g\alpha [H^3 + H\phi^0\phi^0 + \\
2H\phi^+\phi^-] - \tfrac{1}{8} g^2 \alpha_h [H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + \\
2(\phi^0)^2 H^2] - gMW^+_\mu W^-_\mu H - \tfrac{1}{2} g \tfrac{M}{c_w^2} Z^0_\mu Z^0_\mu H - \tfrac{1}{2} ig [W^+_\mu (\phi^0 \partial_\mu \phi^- - \phi^- \partial_\mu \phi^0) - \\
W^-_\mu (\phi^0 \partial_\mu \phi^+ - \phi^+ \partial_\mu \phi^0)] + \tfrac{1}{2} g [W^+_\mu (H \partial_\mu \phi^- - \phi^- \partial_\mu H) - W^-_\mu (H \partial_\mu \phi^+ - \\
\phi^+ \partial_\mu H)] + \tfrac{1}{2} g \tfrac{1}{c_w} (Z^0_\mu (H \partial_\mu \phi^0 - \phi^0 \partial_\mu H) - ig \tfrac{s_w^2}{c_w} M Z^0_\mu (W^+_\mu \phi^- - W^-_\mu \phi^+) + \\
igs_w M A_\mu (W^+_\mu \phi^- - W^-_\mu \phi^+) - ig \tfrac{1-2c_w^2}{2c_w} Z^0_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) + \\
igs_w A_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) - \tfrac{1}{4} g^2 W^+_\mu W^-_\mu [H^2 + (\phi^0)^2 + 2\phi^+\phi^-] - \\
\tfrac{1}{4} g^2 \tfrac{1}{c_w^2} Z^0_\mu Z^0_\mu [H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2 \phi^+\phi^-] - \tfrac{1}{2} g^2 \tfrac{s_w^2}{c_w} Z^0_\mu \phi^0 (W^+_\mu \phi^- + W^-_\mu \phi^+) - \\
\tfrac{1}{2} ig^2 \tfrac{s_w^2}{c_w} Z^0_\mu H (W^+_\mu \phi^- - W^-_\mu \phi^+) + \tfrac{1}{2} g^2 s_w A_\mu \phi^0 (W^+_\mu \phi^- + W^-_\mu \phi^+) + \\
\tfrac{1}{2} ig^2 s_w A_\mu H (W^+_\mu \phi^- - W^-_\mu \phi^+) - g^2 \tfrac{s_w}{c_w} (2c_w^2 - 1) Z^0_\mu A_\mu \phi^+\phi^- - g^1 s_w^2 A_\mu A_\mu \phi^+\phi^- - \\
\bar{e}^\lambda (\gamma\partial + m_e^\lambda) e^\lambda - \bar{\nu}^\lambda \gamma\partial \nu^\lambda - \bar{u}^\lambda_j (\gamma\partial + m_u^\lambda) u^\lambda_j - \bar{d}^\lambda_j (\gamma\partial + m_d^\lambda) d^\lambda_j + \\
igs_w A_\mu [-(\bar{e}^\lambda \gamma^\mu e^\lambda) + \tfrac{2}{3} (\bar{u}^\lambda_j \gamma^\mu u^\lambda_j) - \tfrac{1}{3} (\bar{d}^\lambda_j \gamma^\mu d^\lambda_j)] + \tfrac{ig}{4c_w} Z^0_\mu [(\bar{\nu}^\lambda \gamma^\mu (1+\gamma^5) \nu^\lambda) + \\
(\bar{e}^\lambda \gamma^\mu (4s_w^2 - 1 - \gamma^5) e^\lambda) + (\bar{u}^\lambda_j \gamma^\mu (\tfrac{4}{3} s_w^2 - 1 - \gamma^5) u^\lambda_j) + (\bar{d}^\lambda_j \gamma^\mu (1 - \tfrac{8}{3} s_w^2 - \gamma^5) d^\lambda_j)] + \\
\tfrac{ig}{2\sqrt{2}} W^+_\mu [(\bar{\nu}^\lambda \gamma^\mu (1+\gamma^5) e^\lambda) + (\bar{u}^\lambda_j \gamma^\mu (1+\gamma^5) C_{\lambda\kappa} d^\kappa_j)] + \tfrac{ig}{2\sqrt{2}} W^-_\mu [(\bar{e}^\lambda \gamma^\mu (1+\gamma^5) \nu^\lambda) + \\
(\bar{d}^\kappa_j C^\dagger_{\lambda\kappa} \gamma^\mu (1+\gamma^5) u^\lambda_j)] + \tfrac{ig}{2\sqrt{2}} \tfrac{m_e^\lambda}{M} [-\phi^+ (\bar{\nu}^\lambda (1-\gamma^5) e^\lambda) + \phi^- (\bar{e}^\lambda (1+\gamma^5) \nu^\lambda)] - \\
\tfrac{g}{2} \tfrac{m_e^\lambda}{M} [H (\bar{e}^\lambda e^\lambda) + i\phi^0 (\bar{e}^\lambda \gamma^5 e^\lambda)] + \tfrac{ig}{2M\sqrt{2}} \phi^+ - m_d^\kappa (\bar{u}^\lambda_j C_{\lambda\kappa} (1-\gamma^5) d^\kappa_j) + m_u^\lambda (\bar{u}^\lambda_j C_{\lambda\kappa} (1 + \\
\gamma^5) d^\kappa_j) + \tfrac{ig}{2M\sqrt{2}} \phi^- (m_d^\lambda (\bar{d}^\lambda_j C^\dagger_{\lambda\kappa} (1+\gamma^5) u^\kappa_j) - m_u^\kappa (\bar{d}^\lambda_j C^\dagger_{\lambda\kappa} (1-\gamma^5) u^\kappa_j)) - \\
\tfrac{g}{2} \tfrac{m_u^\lambda}{M} H (\bar{u}^\lambda_j u^\lambda_j) - \tfrac{g}{2} \tfrac{m_d^\lambda}{M} H (\bar{d}^\lambda_j d^\lambda_j) + \tfrac{ig}{2} \tfrac{m_u^\lambda}{M} \phi^0 (\bar{u}^\lambda_j \gamma^5 u^\lambda_j) - \tfrac{ig}{2} \tfrac{m_d^\lambda}{M} \phi^0 (\bar{d}^\lambda_j \gamma^5 d^\lambda_j) + \bar{X}^+ (\partial^2 - \\
M^2) X^+ + \bar{X}^- (\partial^2 - M^2) X^- + \bar{X}^0 (\partial^2 - \tfrac{M^2}{c_w^2}) X^0 + \bar{Y} \partial^2 Y + igc_w W^+_\mu (\partial_\mu \bar{X}^0 X^- -
\end{gathered}
$$

$$\partial_\mu \bar{X}^+ X^0) + igs_w W_\mu^+ (\partial_\mu \bar{Y} X^- - \partial_\mu \bar{X}^+ Y) + igc_w W_\mu^- (\partial_\mu \bar{X}^- X^0 - \partial_\mu \bar{X}^0 X^+) + igs_w W_\mu^- (\partial_\mu \bar{X}^- Y - \partial_\mu \bar{Y} X^+) + igc_w Z_\mu^0 (\partial_\mu \bar{X}^+ X^+ - \partial_\mu \bar{X}^- X^-) + igs_w A_\mu (\partial_\mu \bar{X}^+ X^+ - \partial_\mu \bar{X}^- X^-) - \frac{1}{2} g M [\bar{X}^+ X^+ H + \bar{X}^- X^- H + \frac{1}{c_w^2} \bar{X}^0 X^0 H] + \frac{1-2c_w^2}{2c_w} igM [\bar{X}^+ X^0 \phi^+ - \bar{X}^- X^0 \phi^-] + \frac{1}{2c_w} igM [\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-] + igMs_w [\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-] + \frac{1}{2} igM [\bar{X}^+ X^+ \phi^0 - \bar{X}^- X^- \phi^0].$$

The SM [3,4] describes all known particles and the interactions (except for gravity) within the framework of a quantum field theory. More strictly, the SM is built on a local gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$ (here C, L, and Y stand for the color charge, the left-handed chirality, and the weak hypercharge), where the carriers of the forces are the generators of the symmetry group. The fundamental blocks of matter are the fermion fields, particles with spin 1/2, organized in three generations, each one consisting of two leptons and two flavors of quarks (Table 1.1). While the quark fields couple with all the force carriers, the leptons are only involved in electroweak interactions.

	Leptons		Quarks	
1st Generation	$\begin{pmatrix} \nu_{eL} \\ e_L^- \end{pmatrix}$,	e_R^-	$\begin{pmatrix} u_L \\ d_L \end{pmatrix}$	$, u_R, d_R$
2nd Generation	$\begin{pmatrix} \nu_{\mu L} \\ \mu_L^- \end{pmatrix}$,	μ_R^-	$\begin{pmatrix} c_L \\ s_L \end{pmatrix}$	$, c_R, s_R$
3rd Generation	$\begin{pmatrix} \nu_{\tau L} \\ \tau_L^- \end{pmatrix}$,	τ_R^-	$\begin{pmatrix} t_L \\ b_L \end{pmatrix}$	$, t_R, b_R$

Table 1.1: Right-handed fermion singlets and lef-handed doublets for each generation in the SM. All the fermions have an anti-fermion counterpart not reported here.

At the *zeroeth level*, all the SM fields are massless, and the quarks and leptons fields are combined in multiplets. Since the electroweak interactions are chiral, the left-handed (LH) and the right-handed (RH) components are assigned to different representations of the SU(2) group: the LH fields transform as SU(2) doublets, and the RH fields transform as singlets. At the *first level*, the $SU(2)_L \times U(1)_Y$ gauge group is spontaneously broken down to $U(1)_{EM}$ at a scale $v = 246$ GeV. This spontaneous symmetry breaking is simply and effectively described by the famous *Englert-Brout-Higgs mechanism* [5,6]. The introduction of the so called *Higgs field*, a scalar field with non-vanishing vacuum expectation value, allows the definition of non-zero masses for W and Z gauge bosons, as well as for fermions. Nearly fifty years after the theorization of this mechanism, the ATLAS and CMS Collaborations claimed the observation of a new particle with a mass of approximately 125 GeV [7–9], consistent with the SM predictions. The consequent measurement of its properties with the unprecedented amount of data available in proton-proton collisions at the Large Hadron Collider (LHC) [10] confirmed the success of the SM framework.

Despite its incredible accuracy (especially in the electroweak sector), new physics beyond the Standard Model (BSM) is invoked to address several theoretical and experimental open issues. Among these, we can mention:

Neutrino masses and nature. Described as massless fermions within the SM, neutrinos became the foremost actor of recent new physics measurements. The observation of the flavor oscillation in solar and atmospheric neutrino experiments [11–13] attested that neutrinos have small masses, which cannot be generated by the Higgs mechanism. The picture is further complicated by the ambiguity on the nature of neutrinos with definite masses: are they Majorana (lepton number violating) or Dirac (lepton number conserving) fermions?

Dark matter and dark energy evidence. Cosmological observations, such as the cosmic microwave background and the structure and movements of galaxies, show that the baryonic matter described by the SM constitutes only roughly the 5% of the Universe. The remaining fraction is expected to be formed for about 25% by non-visible matter, the Dark Matter, and a large energy component, called Dark Energy.

Inclusion of the gravitational force. The gravitational force is currently not related to microscopic interactions of the SM. Whether this is a good approximation at the electroweak scale $(m_{weak} \approx 100\,\mathrm{GeV})$ where gravity is so weak to be negligible, at the so-called Planck scale $\left(M_p \approx 10^{19}\,\mathrm{GeV}\right)$ the effects of quantum gravity are expected to become important.

Matter-antimatter asymmetry in the Universe. The baryon-antibaryon asymmetry observed in the Universe (discussed in Chapter 6.3) is not explained by the Standard Model.

The hierarchy problem. The Higgs field, introduced for the spontaneous symmetry breaking mechanism in the SM, is quadratically sensitive to high scales. The radiative corrections (one-loop) to the Higgs mass tend to push that mass up to the next physics scale after the electroweak. In the worst scenario in which the next physics scale lies at the Plank mass, the bare potential must be incredibly fine-tuned to cancel off the quantum corrections and get the correct physical mass parameter [14].

The forces unification problem. The SM gauge group $SU(3)_c\times\ SU(2)_L \times U(1)_Y$ consists of three different subgroups, each having its coupling constant. These couplings run with the scale, and it would be natural if they converge toward the same point at some scale, usually referred to as *the unification scale*. However, precision measurement has shown that the three coupling constants do not meet in a single point [15, 16].

The masses, and the flavor problem. Why does the SM include three generations of quarks and leptons, the last two being a heavier version of the first? Why do the masses of the fermions span over many orders of magnitude? The SM cannot directly explain this.

2 BEYOND THE STANDARD MODEL

The difficulties underlined in the previous section suggest that the SM could be part of a more fundamental underlying theory, and motivated the vast experimental program focused on high precision studies of the SM predictions and the physics landscape beyond it. Among the several extensions of the SM proposed in the past, we will cite the most famous in collider phenomenology without the desire to explain them exhaustively, addressing a complete discussion to specific references [10,14,17].

Grand Unification Theories. Grand unification theories (GUT) [4] postulate that the SM gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$, consisting of three different subgroups each having its coupling constant, originates from a larger symmetry group with only one coupling constant. One of the most popular GUT scenarios, the Left-Right symmetric model (LRM), is based on an $SO(10)$ group that is supposed to break to:

$$SO(10) \to SU(3)_c \times SU(2)_L \times SU(2)_R \times U(1)_{B-L},$$

where the $U(1)_{B-L}$ is the global gauge symmetry associated with the conservation of baryon minus lepton number. The LRM predicts the existence of new gauge bosons, Z' and $W'^{\pm}$, that couple to right-handed fermions (including the right-handed neutrinos, not present in the SM).

Supersymmetry. Supersymmetry (SUSY) refers to possible relations between the spectrum and interactions of fermions (half-integer spin particles) and bosons (integer spin particles). SUSY predicts a supersymmetric partner for every SM particle, with the same mass but with the spin differing by $1/2$. If SUSY exists in nature, it must be a broken symmetry at some scale since no evidence of these particles has been found so far. Supersymmetry offers an elegant solution to the hierarchy problem: the fermion loop that makes the Higgs boson mass sensitive to high-scales is canceled by the SUSY partner's contribution. Moreover, in the so-called *R-parity conserving* SUSY models, the lightest supersymmetric particle is stable and it is considered a good Dark Matter candidate.

Extra-dimensions. The four dimensions that we experience could be embedded in a higher-dimensional space. Given the lack of evidence for these extra dimensions, they should be either compactified [18] or characterized by such a strong curvature to make it hard to escape into them [19]. In these models, gravity can propagate in the extra dimensions, and therefore the SM particles experience only a small fraction of the total gravitational force. That will solve the hierarchy problem since the scale of gravity lies around the TeV region. Moreover, the new particles that arise in some scenarios, the so-called Kaluza-Klein excitations [20], are often presented as a valid Dark Matter candidate.

Effective theories, including also *composite models*, conclude this short summary, and are intentionally left out from the list since a specific introduction to their formalism is presented in the next section.

3 EFFECTIVE THEORIES

Throughout the following part of this Chapter, we delineate the basics of effective field theories. In particular, we are going to highlight the fundamental concepts behind these BSM scenarios and provide the formalism necessary to understand the results given in the next Chapters. First, we discuss the roots to effective theory (Sec. 1.3.1), and we cite the example of two most famous EFT scenarios: the Standard Model EFT (SMEFT) and the Higgs EFT (HEFT) (Sec. 1.3.2). We then shift to the effective theory of composite models (Sec. 1.3.3), which represents the benchmark framework of this thesis.

3.1 THE EFFECTIVE FIELD THEORY APPROACH

What is an effective theory? The answer contains at least these two words: *appropriate*, and *important* (citing the famous lecture by H. Georgi [21]). Nature manifests in many scales, and we can choose to divide its parameter space into different regions, in each of which there is an *appropriate* description of the *important* physics. Such an *appropriate* description of the *important* physics is an effective theory. We can think of mechanics: we learn Newtonian mechanics as a separate discipline, and not as the limit of relativistic mechanics for small velocities. In the parameter space that we experience every day all velocities are much smaller than the speed of light, and relativity can be ignored. However, there is nothing wrong in treating mechanics in a fully relativistic way, it just makes it easier to make calculations by not including relativity. In other words, if there are very large or very small parameters compared to the physical quantities we are interested in, the most appropriate picture for performing calculations consists of setting to zero small parameters, and large parameters to infinity, considering all finite effects as perturbations.
In a quantum field theory, the effective approach traces its roots on algebraic calculations in pion-pion scattering with the so-called Chiral Lagrangian [22]. We can recall that there exists a limit of QCD in which the π, the K, and the η mesons are massless Goldstones of a Spontaneous Symmetry Breaking of $SU(3) \times SU(3)$, we call "chiral symmetry".

More generally, to study a particular system it is necessary to single out the most relevant degrees of freedom as the building blocks to attain a simple description of the problem at hand. It is crucial to make the appropriate choice of these variables capable to capture the most important effects at a given scale. The degrees of freedom that become relevant at any higher energy scale are not taken into account and do not appear explicitly in the formulation of the theory. That is the case, for instance, of a system containing a heavy particle: it cannot be created at energies smaller than its mass, M, and therefore a theory and its corresponding Lagrangian valid at such small energies does not contain this degree of freedom. The rigorous description of this property is provided by the Appelquist-Carazzone theorem [23], who showed that heavy degrees of freedom decouple at energy scales much lower than their mass. In this context, decoupling means that any effect of the heavy degrees of freedom is, up to logarithmic contributions, suppressed by inverse powers of the heavy scale M. In the strict heavy mass limit, $M \to \infty$, the heavy state does not provide any correction [24].

However, the EFT approach encodes more than a convenience in calculations. We can invert the process, and suppose we only can experience the low-energy theory. This approach commonly goes under the name *bottom up*. That might be the case of the SM. As the hierarchy issue seems to suggest, some of the interactions lie at a scale so large that we cannot see them directly now, and perhaps not ever. Nonetheless, we can use an EFT to parametrize our ignorance of what is going on at higher energies. All we can do is to impose the SM known invariance $SU(3)_C \times SU(2)_L \times U(1)_Y$, and possibly the conservation laws measured in nature with good precision (like baryon number conservation, etc.) to write a general Lagrangian density valid up to the electroweak scale in Taylor expansion:

$$\mathcal{L}_{eff} = \mathcal{L}_0 + \frac{1}{\Lambda}\mathcal{L}_1 + \frac{1}{\Lambda^2}\mathcal{L}_2 + ..., \tag{1.1}$$

where $\mathcal{L}_0$ stands for the dimension four SM Lagrangian density, $\mathcal{L}_1$ is a dimension 5 term, etc.. A milestone reference for this expansion, including the detailed discussion on the type of operators allowed by the chosen symmetry, is [25].

This "old-fashioned effective theory" was rapidly reprized and expanded in the last decades, flowing into two main well-defined scenarios. In the next section, we will briefly cover these two effective field theory variants called *SMEFT* and *HEFT*. The following discussion is based on [26–29].

3.2 TWO EFT PROTOTYPES FOR THE SM: THE SMEFT VS. THE HEFT

With the striking success of the SM phenomenology, the model-independent approach is increasingly becoming a central theme for the LHC general-purpose experiments. This *bottom-up* searching strategy, will have wider applicability and higher possibilities for discovery as the LHC Run 3, and especially the High-Luminosity LHC, will be able to deliver increased statistics and enhanced detector capabilities. Two main directions of thoughts have emerged within this sector, and their main difference is reflected in the description of the electroweak symmetry breaking mechanism: the linear realization of the Higgs sector in the SMEFT, and non-linear realization in the HEFT, which works in the broken phase using the physical Higgs boson and an independent set of Goldstone bosons.

The SMEFT. The $SU(2)_L \times SU(2)_R$ symmetry in the Higgs sector is realized linearly, with the Higgs field being a $SU(2)_L\times$ $SU(2)_R$ bi-doublet. The perturbative expansion is characterized by the inverse powers of the new physics scale Λ, similar to Eq. 1.1. Assuming Baryon and Lepton number conservation, the SMEFT Lagrangian is written as the sum:

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \sum_i c_i^{(6)} \mathcal{O}_i^{(6)} + \sum_i c_i^{(8)} \mathcal{O}_i^{(8)} + \ldots, \tag{1.2}$$

where $\mathscr{L}_{\rm SM}$ stands for the SM Lagrangian, $c_i^{(n)}$ are dimensionless coefficients (known as *Wilson coefficients*) and $\mathcal{O}^{(n)}$ are operators of canonical dimension $n > 4$ that encode the suppression Λ^{4-n}. Dots represent higher order operators in the expansion in inverse powers of the cut-off. The SMEFT is renormalizable order by order in the expansion in Λ : independently from the number of loops of a given diagram, the quantum corrections calculated from the Lagrangian truncated at some order of Λ and calculated to the same order in Λ can be renormalized by the couplings present in the effective Lagrangian.

Considering observables whose typical energy is much smaller than the cut-off, the higher the canonical dimension of an operator the smaller is its contribution to those observables. Once the typical energy involved is closer to the cut-off, the ordering of the operators in canonical dimensions is not so meaningful anymore and operators of higher dimensions may dominate. In this case, the effective description breaks down and cannot be considered reliable anymore. An entire Chapter (2) of this thesis is devoted to this issue, which represents a *leitmotif* of this work.

The HEFT. In this case, the $SU(2)_L \times SU(2)_R$ symmetry is realized non-linearly, on the three Goldstone bosons absorbed by the gauge fields. The physical Higgs field may be either a singlet of the custodial symmetry $SU(2)_C$ or an additional field never forming the Higgs doublet with the three Goldstone bosons (GBs). This Lagrangian is the most general description of the EW and Higgs couplings, satisfying the gauge symmetry of the SM: in specific limits, it may reduce to the SM Lagrangian, or may coincide with the SMEFT one, or may match the description of BSM specific scenarios, like the composite Higgs models.

The HEFT can be considered as the merging between the chiral perturbation theory applied to the longitudinal components of the gauge bosons and the Higgs and the SMEFT once dealing with the other SM particles. Following the conventional notation of the EW Chiral Lagrangian, the SM GBs π^a are described employing a dimensionless unitary matrix transforming as a bi-doublet of the global $SU(2)_L \times SU(2)_R$ symmetry

$$\mathbf{U}(x) = e^{i\sigma_a \pi^a(x)/v}, \quad \mathbf{U}(x) \to L\mathbf{U}(x)R^\dagger \tag{1.3}$$

with $v = 246$GeV being the EW scale defined through the W mass, and $L(R)$ denote the $SU(2)_{L(R)}$ transformation. It is customary to introduce two objects, the vector and the scalar chiral fields, that transform in the adjoint of $SU(2)_L$:

$$\mathbf{V}_\mu(x) \equiv \left(\mathbf{D}_\mu \mathbf{U}(x)\right)\mathbf{U}(x)^\dagger, \quad \mathbf{T}(x) \equiv \mathbf{U}(x)\sigma_3\mathbf{U}(x)^\dagger \tag{1.4}$$

where the covariant derivative reads

$$\mathbf{D}_\mu \mathbf{U}(x) \equiv \partial_\mu \mathbf{U}(x) + igW_\mu(x)\mathbf{U}(x) - \frac{ig'}{2}B_\mu(x)\mathbf{U}(x)\sigma_3 \tag{1.5}$$

While $\mathbf{V}_\mu(x)$ is $SU(2)_C$ conserving, $\mathbf{T}(x)$ is not and therefore plays the role of a custodial symmetry breaking spurion.

In the rest of the Chapter, we focus on another type of effective approach that considers some unknown structure for ordinary fermions, the so-called compositeness framework.

3.3 CONTACT INTERACTIONS AND COMPOSITENESS

Compositeness is a BSM scenario widely discussed in the phenomenological and experimental community [30–36]. Strongly inspired by the hadronic spectroscopy, the proliferation of quarks and leptons has naturally led to the speculation that they can be composite structures, bound states of more fundamental constituents. This scenario is strictly connected with the EFT formalism discussed in the previous sections, but foresees new actors in the game - the *preons*- that show up at some unknown energy scale, that we call the compositeness scale Λ. In the very first realization of the composite model, the existence of sub-constituents of the ordinary matter (whose dynamic is not directly questioned) manifests as a higher-order correction to the SM Lagrangian with four-fermion operators of dimension-6 (Eq. 1.1). The most general flavor-diagonal color-singlet and chirality invariant Lagrangian is written with the following convention [37]:

$$\mathcal{L} = \mathcal{L}_{LL} + \mathcal{L}_{RR} + \mathcal{L}_{LR} + \mathcal{L}_{RL}, \tag{1.6}$$

with:

$$\begin{aligned}
\mathcal{L}_{LL} &= \frac{g_\star^2}{2\Lambda^2} \sum_{i,j} \eta_{LL}^{ij} \left(\bar{\psi}_L^i \gamma_\mu \psi_L^i\right) \left(\bar{\psi}_L^j \gamma^\mu \psi_L^j\right), \\
\mathcal{L}_{RR} &= \frac{g_\star^2}{2\Lambda^2} \sum_{i,j} \eta_{RR}^{ij} \left(\bar{\psi}_R^i \gamma_\mu \psi_R^i\right) \left(\bar{\psi}_R^j \gamma^\mu \psi_R^j\right), \\
\mathcal{L}_{LR} &= \frac{g_\star^2}{2\Lambda^2} \sum_{i,j} \eta_{LR}^{ij} \left(\bar{\psi}_L^i \gamma_\mu \psi_L^i\right) \left(\bar{\psi}_R^j \gamma^\mu \psi_R^j\right), \\
\mathcal{L}_{RL} &= \frac{g_\star^2}{2\Lambda^2} \sum_{i,j} \eta_{RL}^{ij} \left(\bar{\psi}_R^i \gamma_\mu \psi_R^i\right) \left(\bar{\psi}_L^j \gamma^\mu \psi_L^j\right).
\end{aligned} \tag{1.7}$$

Here ψ are the SM fermion spinors with the indices i, j labelling their species, while color and other indices are suppressed. The coupling $g_\star^2$ is conventionally fixed to $g_\star^2/4\pi = 1$, as stemming the strong dynamics that is foreseen by the model, while the helicity parameters $\eta_{\alpha\beta}$ are taken to have unit magnitude. Different chiral structures have been investigated by the ATLAS and the CMS Collaborations with the recent LHC Run 2 data [38, 39], pushing the limit range on the compositeness scale in the range between 20 TeV and 40 TeV.

On a more general ground, composite models foresee two additional features:

i. the existence of *excited states* of the known fermions, often called excited leptons or quarks ($l^\star$, $q^\star$), whose masses are expected to be very high, but still less than the binding scale, $m^\star \leq \Lambda$. They are expected to share the main quantum numbers with their lighter partners [40]: an excited lepton, for instance, shares a leptonic quantum number with one of the existing leptons (and similarly for the excited quarks);

ii. the presence of a dimension-5 gauge-type interactions for the associated production with a SM fermion. Given the supposed mass difference between the excited fermion and its lighter SM partner, the excited leptons and quarks

$(l^\star, q^\star)$ interacts with light fermions via a *magnetic-type gauge coupling*[1] generally expressed as:

$$
\begin{aligned}
\mathcal{L}_{gauge} =& \frac{\lambda_\gamma^{(\psi^*)} e}{2m_{\psi^*}} \bar{\psi}^* \sigma^{\mu\nu} \left(\eta_L \frac{1-\gamma_5}{2} + \eta_R \frac{1+\gamma_5}{2} \right) \psi F_{\mu\nu} \\
&+ \frac{\lambda_Z^{(\psi^*)} e}{2m_{\psi^*}} \bar{\psi}^* \sigma^{\mu\nu} \left(\eta_L \frac{1-\gamma_5}{2} + \eta_R \frac{1+\gamma_5}{2} \right) \psi Z_{\mu\nu} \\
&+ \frac{\lambda_W^{(l^*)} g}{2m_{l^*}} \bar{l}^* \sigma^{\mu\nu} \frac{1-\gamma_5}{2} \nu W_{\mu\nu} \\
&+ \frac{\lambda_W^{(\nu^*)} g}{2m_{\nu^*}} \bar{\nu}^* \sigma^{\mu\nu} \left(\eta_L \frac{1-\gamma_5}{2} + \eta_R \frac{1+\gamma_5}{2} \right) l W^\dagger_{\mu\nu} + \text{h.c.},
\end{aligned}
\tag{1.8}
$$

where the various λ couplings can be expressed in terms of SM couplings in specific scenarios, $g = e/\sin\theta_W, \psi = \nu$ or $l, F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the photon field strength, $Z_{\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu$, etc.. The normalization of the coupling is chosen such that $\max\left(|\eta_L|, |\eta_R|\right) = 1$. Therefore, for chirality conservation $\eta_L \eta_R = 0$.

Although in the past many studies addressed compositeness model building, no correct or compelling model can be identified at present. In the literature we can identify mainly three distinct scenarios [41]:

1. *Sequential type Model.* Extending directly the dynamical structure of SM, one can organize excited leptons in left-handed doublets and right-handed singlets of the $SU(2)_L$

$$
L_L^\star = \begin{pmatrix} \nu^\star \\ e^\star \end{pmatrix}_L, \qquad [\nu_R^\star], \qquad e_R^\star, \tag{1.9}
$$

where square parentheses denote that the are two cases possible: with and without the excited right-handed neutrino.
Consequently, the magnetic-type couplings can written as:

$$
\mathcal{L}_{gauge} = \frac{1}{2\Lambda} \bar{e}_R \phi^\dagger \sigma^{\mu\nu} \left(g f \frac{\tau^a}{2} W^a_{\mu\nu} + g' f' \frac{Y}{2} B_{\mu\nu} \right) L_L^* + h.c, \tag{1.10}
$$

and

$$
\mathcal{L}_{gauge} = \frac{1}{2\Lambda} \bar{\ell}_L \sigma^{\mu\nu} \left(g f \frac{\tau^a}{2} W^a_{\mu\nu} + g' f' \frac{Y}{2} B_{\mu\nu} \right) \phi \nu_R^* + h.c., \tag{1.11}
$$

if the heavy neutrino is considered. In the previous equations we have introduced the f, f' dimensionless parameters, usually taken as 1. In both the variants, these gauge couplings lead to a v/Λ suppressed coupling, where $v \sim 246$ GeV is the vacuum expectation value of $\sqrt{2}\phi$.

1 A γ_μ transition mediated by SM bosons between standard fermions and excited fermions would result in a non-conserved electromagnetic current due to their masses difference $j^\mu_{e.m.} \sim \bar{\psi}_{e^\star} \gamma^\mu \psi_e$.

2. *Mirror type Model.* Here multiplets of the excited leptons are assumed to be grouped as:

$$[\nu_L^\star], \qquad e_L^\star, \qquad L_R^\star = \begin{pmatrix} \nu^\star \\ e^\star \end{pmatrix}_R, \tag{1.12}$$

and neglecting the presence of the left-handed neutrino singlet we automatically get a $\nu^\star$ Majorana particle, and thus a lepton number violation scenario.
The gauge invariant interaction reads:

$$\mathcal{L}_{gauge} = \frac{1}{2\Lambda}\overline{L}_R^\star \sigma^{\mu\nu}\left(gf\frac{\tau^a}{2}W_{\mu\nu}^a + g'f'\frac{Y}{2}B_{\mu\nu}\right)L_L + h.c., \tag{1.13}$$

where now $L_R^\star$ is the spinor according to (1.12), whilst L_L is the SM weak doublet.

3. *Homo-doublet Model.* In this third scenario no excited singlets are contemplated:

$$L_L^\star = \begin{pmatrix} \nu^\star \\ e^\star \end{pmatrix}_L, \qquad L_R^\star = \begin{pmatrix} \nu^\star \\ e^\star \end{pmatrix}_R, \tag{1.14}$$

and the reference Lagrangian density for the gauge couplings is analogous to Eq. 1.13. Since left-handed and right-handed are almost considered with the same mass, this model leads intrinsically to a lepton number conservation scenario, as it is known that two left-handed and two right-handed Majorana fields give Dirac particles.
However, by introducing an L-R symmetry breaking with mixing between $\nu_L^\star - \nu_R^\star$ one can afford lepton number violation.

Throughout this thesis, we will take as a reference for our investigations the heavy composite Majorana neutrino scenario already considered by the CMS Collaboration in Ref. [42].

2 UNITARITY

We mentioned in the previous chapter that the theory of strong interactions, or Quantum Chromodynamics (QCD), is presently explained in terms of quarks and gluons (generators of the $SU(3)_C$ symmetry group). The level of accuracy in the predictions within its perturbation theory range, i.e. in the regime of very high-energy collisions and large transferred momentum, is remarkably high. However, we can think that the bulk of interactions in hadron-colliders involves low-momentum partons, for which the perturbative expansion is not applicable. Very general principles consisting of analyticity, unitarity, and crossing symmetry, directly descending from non-relativistic Quantum Mechanics, come in aid as fundamental axioms to be used in collider physics.

In this chapter, we focus on one of them, shaping the concept of perturbative unitarity in hadron collider processes, starting from its very first derivation in Quantum Field Theory.

We begin by recalling the so-called *optical theorem* in Sec. 2.1, and we then focus on some remarkable consequences in hadron collider searches. We then present several possible healing mechanisms for the unitarity violation in particle physics phenomenology. In Section 2.2, we examine in detail a possible approach to the problem in the case of exotic composite fermions searches at the LHC, that we will see has a relevant impact on the experimental searches, discussed in detail in Chapter 4.

1 GENERAL PRINCIPLES AND THE OPTICAL THEOREM

Theoretically, unitarity does not mean much but *probability of quantum transitions adding up to one*. Despite its simplicity, the rigorous application of this basic principle in the amplitudes of high-energy processes becomes a powerful phenomenological tool for SM and BSM physics. In the end, citing A. Martin and F. Cheung (Lectures at Brandeis Summer School) *"it seems to us to be a necessary task to explore a bit-by-bit the rigorous consequences of analyticity, unitarity, and crossing. Who knows if someday one will not be able to reassemble the pieces of the puzzle"* [43].

The very basic quantity of interest for this discussion is the transition probability from a given initial state $|i\rangle$ to a certain final state $|f\rangle$ after some non-specified collision. This quantity is well known to be described by the S-matrix:

$$S_{fi} = \langle f|S|i\rangle. \tag{2.1}$$

Given that there is some probability that nothing occurs we can write it, in matricial form, as

$$S = \mathbb{1} + i\mathcal{T}, \tag{2.2}$$

and the newly defined $\mathcal{T}$ matrix encompasses all the possible non-trivial transitions information.
Unitarity of the S matrix $S^\dagger S = \mathbb{1}$ then directly implies

$$i\left(\mathcal{T}^\dagger - \mathcal{T}\right) = \mathcal{T}^\dagger \mathcal{T}. \tag{2.3}$$

If we explicit the four-momentum conservation, we can write

$$\langle f|\mathcal{T}|i\rangle = (2\pi)^4\delta^4\left(p_i - p_f\right)\mathcal{M}(i\to f), \tag{2.4}$$

and easily write the LH-side of Eq. 2.3 as:

$$\begin{aligned}\left\langle f\left|i\left(\mathcal{T}^\dagger - \mathcal{T}\right)\right|i\right\rangle &= i\langle i|\mathcal{T}|f\rangle^\star - i\langle f|\mathcal{T}|i\rangle\\ &= i(2\pi)^4\delta^4\left(p_i - p_f\right)\left(\mathcal{M}^\star(f\to i) - \mathcal{M}(i\to f)\right),\end{aligned} \tag{2.5}$$

and the RH one as:

$$\begin{aligned}&\left\langle f\left|\mathcal{T}^\dagger\mathcal{T}\right|i\right\rangle = \textstyle\sum_X \int d\Pi_X \left\langle f\left|\mathcal{T}^\dagger\right|X\right\rangle \langle X|\mathcal{T}|i\rangle\\ &= \textstyle\sum_X (2\pi)^4\delta^4\left(p_f - p_X\right)(2\pi)^4\delta^4\left(p_i - p_X\right)\int d\Pi_X \mathcal{M}(i\to X)\mathcal{M}^\star(f\to X),\end{aligned} \tag{2.6}$$

where we have used the compliteness relation of the Hilbert space

$$\mathbb{1} = \sum_X \int d\Pi_X |X\rangle\langle X| \tag{2.7}$$

summing over single- and multi-particle state $|X\rangle$, and with $d\Pi_X = \prod_{j\in X}\dfrac{d^3p_j}{(2\pi)^3}\dfrac{1}{2E_j}$ being the Lorentz-invariant phase space.
Plugging the two sides, we finally get the most general form of the *optical theorem*:

$$\mathcal{M}(i\to f) - \mathcal{M}^\star(f\to i) = i\sum_X \int d\Pi_X (2\pi)^4\delta^4\left(p_i - p_X\right)\mathcal{M}(i\to X)\mathcal{M}^\star(f\to X), \tag{2.8}$$

that is valid order-by-order in perturbation theory.
To get a more readable form, we can consider the case in which $|i\rangle = |f\rangle = |A\rangle$, and write the cross section as

$$\sigma(A\to X) = \frac{1}{4E_{\mathrm{CM}}\,|\vec{p}_i|}\int d\Pi_X (2\pi)^4\delta^4\left(p_A - p_X\right)|\mathcal{M}(A\to X)|^2. \tag{2.9}$$

The next step is to use this definition in Eq. 2.8 to get the more common relation between the imaginary part of the forward scattering amplitude and the total cross section:

$$\mathrm{Im}\,\mathcal{M}(A\to A) = 2E_{\mathrm{CM}}\,|\vec{p}_i|\sum_X \sigma(A\to X). \tag{2.10}$$

Experimentally, this equation means that the total cross section can be measured either by counting all the final states produced, or via the forward elastic scattering amplitude. Moreover, another important implication of this theorem in high-energy physics is that the amplitudes cannot be arbitrarily large [44]. Indeed, writing

Eq. 2.10 in a more compact notation we have that $\mathrm{Im}\,\mathcal{M} \leq |\mathcal{M}|^2 \implies |\mathcal{M}| < 1$. Bounds for the amplitudes can be made explicit in many ways. An example worth to be mentioned is the so-called Froissart bound, that puts a strong upper constraint on the high-energy processes total cross-section of the order of $\ln^2 E_{CM}$.
The focus of this chapter will instead be another general bound descending from the optical theorem: the perturbative unitarity bound.

1.1 PARTIAL WAVE UNITARITY BOUNDS

Let us consider an elastic scattering process with 2 particles in the initial state and the same two particles in the final state. The total cross section in the center-of-mass reference frame for these processes can be expanded in terms of partial waves:

$$\sigma_{\text{tot}}(AB \to AB) = \frac{1}{32\pi E_{\text{CM}}^2} \int d\cos\theta |\mathcal{M}(\theta)|^2, \tag{2.11}$$

where the amplitude

$$\mathcal{M}(\theta) = 16\pi \sum_{j=0}^{\infty} a_j (2j+1) P_j(\cos\theta) \tag{2.12}$$

is written in terms of the Legendre polynomials $P_j(\cos\theta)$, satisfying $P_j(1) = 1$ and

$$\int_{-1}^{1} P_j(\cos\theta) P_k(\cos\theta) d\cos\theta = \frac{2}{2j+1}\delta_{jk}. \tag{2.13}$$

If we integrate Eq. 2.11 on $\cos\theta$ we get

$$\sigma_{\text{tot}} = \frac{16\pi}{E_{\text{CM}}^2} \sum_{j=0}^{\infty} (2j+1) \left|a_j\right|^2. \tag{2.14}$$

Now, for the optical theorem the imaginary part of the forward scattering amplitude, at $\theta = 0$, is related to the total cross section:

$$\begin{aligned} \mathrm{Im}\,\mathcal{M}(\theta = 0) &= 2E_{\text{CM}} \left|\vec{p}_i\right| \sum_X \sigma_{\text{tot}}(AB \to X) \\ &\geq 2E_{\text{CM}} \left|\vec{p}_i\right| \sigma_{\text{tot}}(AB \to AB) \end{aligned} \tag{2.15}$$

and therefore

$$\sum_{j=0}^{\infty} (2j+1)\,\mathrm{Im}\,(a_j) \geq \frac{2\left|\vec{p}_i\right|}{E_{\text{CM}}} \sum_{j=0}^{\infty} (2j+1) \left|a_j\right|^2. \tag{2.16}$$

We can see its meaning through an extremely simple process, although the same arguments apply to more general cases but with a more involved mathematical derivation.
Let us focus on the high-energy limit $|\vec{p}_i| = \frac{1}{2}E_{\text{CM}}$ and drop out the sum on j by looking at the scattering between angular momentum eigenstates:

$$\mathrm{Im}\,(a_j) \geq \left|a_j\right|^2. \tag{2.17}$$

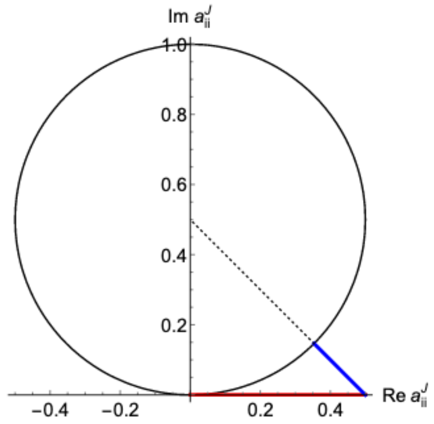


Figure 2.1: Unitarity constraint in the Argand plane. A Born value of $a_j = \frac{1}{2}$ and $\mathrm{Im}\, a_j = 0$ (red line) requires a correction (blue line), which amounts to at least the $\sqrt{2}-1$ 40% of the the tree-level value in order to come back inside the unitarity circle [45].

This means that $|a_j|$ must lie inside the circle in the Argand plane defined by

$$(\mathrm{Re}\, a_j)^2 + \left(\mathrm{Im}\, a_j - \frac{1}{2}\right)^2 \leq \frac{1}{4} \tag{2.18}$$

that implies

$$0 \leq \mathrm{Im}\, a_j \leq 1 \quad \text{and} \quad |\mathrm{Re}\, a_j| \leq \frac{1}{2}. \tag{2.19}$$

The assumption that the tree-level amplitude is real suggests the following perturbativity criterium

$$\left|\mathrm{Re}\,(a_j)^{\mathrm{Born}}\right| \leq \frac{1}{2}. \tag{2.20}$$

In fact, a Born value of Re $a_j = \frac{1}{2}$ and $\mathrm{Im}\, a_j = 0$ needs at least a correction of 40% in order to restore unitarity (Fig. 2.1).
As we approach the upper bound of the inequality in Eq. 2.16, higher-order corrections are becoming more and more important for our calculation, and we are losing the predictability of our theory. As we will see in the next section, although we have obtained the partial wave unitarity bound in its simpler formulation, i.e. for elastic scattering and tree-level amplitudes, its validity is granted also for the inelastic scattering processes [44].

Much earlier than the discovery of the Higgs boson, perturbative unitarity's claim to fame is the Vector Boson Scattering (VBS) process, which is also extensively faced in this thesis with an experimental search within the CMS Collaboration at the LHC. The constraint has been used to infer an upper bound on the Higgs boson mass or on the scale where the SM description of weak interactions would need to be completed in the ultraviolet (UV) in terms of some new strongly coupled dynamics [46,47]. Indeed, within the SM, the sum of all these contributions results in $a(s,\theta) = \sum_i a_i(s,\theta) \propto s/M_W^2$. As a consequence when $s \gg M_W$, the cross section

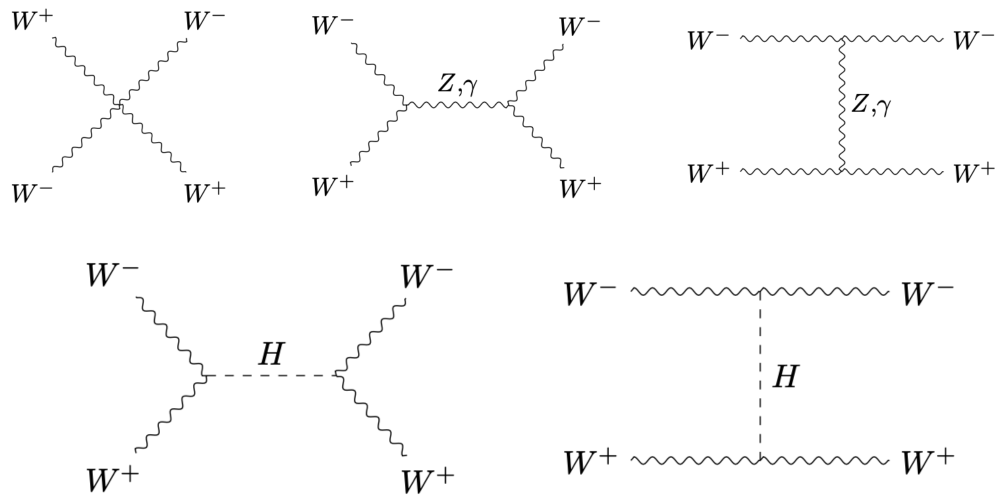


Figure 2.2: Contributions from the triple (TGC) and quartic gauge couplings (QGC) for the scattering (top line), and contributions from the Higgs s-channel and t-channel exchange (bottom line).

grows proportional to s and the unitarity condition is not satisfied (Figure 2.2). Lee, Quigg, and Thacker [48] found that if the Higgs boson mass exceeds the value $\left(8\pi\frac{\sqrt{2}}{3}G_F\right)^{\frac{1}{2}} \approx 1$ TeV, partial-wave unitarity is not respected by the tree diagrams for the two-body scattering of gauge bosons, and the weak interactions must become strong at high energies.

2 PERTURBATIVE UNITARITY BOUNDS FOR EFFECTIVE COMPOSITE MODELS

It is well known that partial wave unitarity condition is a powerful tool to estimate the perturbative validity of effective field theories (EFTs). It has been used in the past to provide useful insights both in strong and electroweak interactions [46] as well as in quantum gravity [49], in composite Higgs models [50], and in searches of scalar di-boson resonances [45, 51], searches for dark matter effective interactions [52] and on generic dimension-6 operators [53]. In this section, we present the partial wave unitarity bound in the parameter space of dimension-5 and dimension-6 effective operators that arise in the compositeness scenario presented in the first Chapter. These are routinely used in experimental searches at the LHC to constraint contact and gauge interactions between ordinary Standard Model fermions and excited (composite) states of mass M. After deducing the unitarity bound for the production process of a composite neutrino, we implement such a bound and compare it with the recent experimental exclusion curves for LHC Run 2, the High-Luminosity and High-Energy configurations of the LHC. Our results also apply to the searches where a generic single excited state is produced via contact interactions. We find that the unitarity bound, so far overlooked, is quite compelling, and significant portions of the parameter space (M, Λ) become excluded in addition to the standard request $M \leq \Lambda$.

We anticipate that one possible BSM alternative, widely discussed in literature and routinely pursued in high-energy experiments, is a composite-fermions scenario, which offers a possible solution to the hierarchy pattern of fermion masses [54–58]. In this context [30, 32–34, 59, 60], SM quarks "q" and leptons "ℓ" are assumed to be bound states of some as yet not observed fundamental constituents generically referred as *preons*. If quarks and leptons have an internal substructure, they are expected to be accompanied by heavy excited states ℓ^*, q^* of masses M that should manifest themselves at an unknown energy scale, the compositeness scale Λ.

As is customary in an EFT approach, the effects of the high-energy physics scale, here Λ, are captured in higher-dimensional operators that describe processes within a lower energy domain, where the fundamental building blocks of the theory cannot show up. Hence, the heavy excited states may interact with the SM ordinary fermions via dimension-5 gauge interactions of the $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ SM gauge group of the magnetic-moment type (so that the electromagnetic current conservation is not spoiled by e.g. $\ell^*\ell\gamma$ processes [32]). Besides, the exchange of preons and/or binding quanta of the unknown interactions between ordinary fermions (f) and/or the excited states (f^*) results in effective contact interactions (CI) that couple the SM fermions and heavy excited states [10, 31, 33, 60]. In the latter case, the dominant effect is expected to be given by the dimension-6 four-fermion interactions scaling with the inverse square of the compositeness scale Λ:

$$\mathcal{L}^6 = \frac{g_*^2}{\Lambda^2}\frac{1}{2} j^\mu j_\mu \,, \tag{2.21a}$$

$$\begin{aligned} j_\mu = {} & \eta_L \bar{f}_L \gamma_\mu f_L + \eta\prime_L \bar{f}^*_L \gamma_\mu f^*_L + \eta\prime\prime_L \bar{f}^*_L \gamma_\mu f_L + h.c. \\ & + (L \to R)\,, \end{aligned} \tag{2.21b}$$

where $g_*^2 = 4\pi$ and the η's factors are usually set equal to unity. In this work the right-handed currents will be neglected for simplicity (this is also the setting adopted by the experimental collaborations).

As far as gauge interactions (GI) are concerned, let us consider the first lepton family and assume that the excited neutrino and the excited electron are grouped into left-handed singlets and a right-handed SU(2) doublet: $e^*_L, \nu^*_L\,, L^*_R = (\nu^*_R\,, e^*_R)^T$, so that a magnetic type coupling between the left-handed SM doublet and the right-handed excited doublet via the $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ gauge fields can be written down [32, 41]:

$$\mathcal{L}^5 = \frac{1}{2\Lambda}\, \bar{L}^*_R \sigma^{\mu\nu} \left(g f \frac{\boldsymbol{\tau}}{2} \cdot \boldsymbol{W}_{\mu\nu} + g' f' Y B_{\mu\nu} \right) L_L + h.c.\,. \tag{2.22}$$

Here, $L^T = (\nu_{\ell L}, \ell_L)$ is the ordinary lepton doublet, g and g' are the $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ gauge couplings and $\boldsymbol{W}_{\mu\nu}$, $B_{\mu\nu}$ are the field strength tensor of the corresponding gauge fields respectively; $\boldsymbol{\tau}$ are the Pauli matrices and Y is the hypercharge, f and f' are dimensionless couplings and are expected (and assumed) to be of order unity.

Excited states interacting with the SM sector through the model Lagrangians (2.21a)-(2.21b) and (2.22) have been extensively searched for at high-energy collider facilities. The current strongest bounds are due to the recent LHC experiments. Charged leptons (e^*, μ^*) have been searched for in the channel $pp \to \ell\ell^* \to \ell\ell\gamma$ [61–

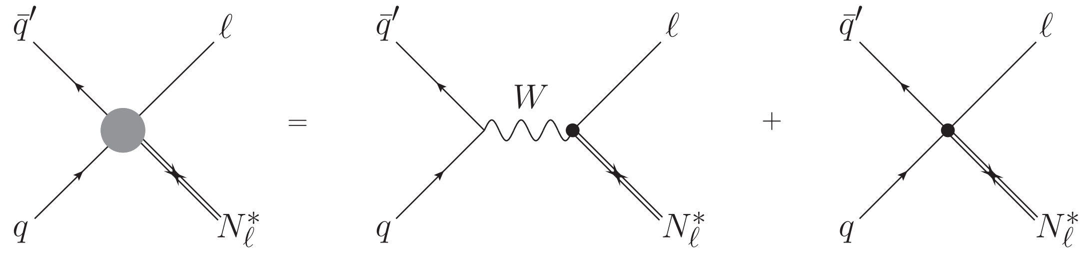


Figure 2.3: Feynman diagrams depicting the mechanisms responsible for the process $q\bar{q}' \to N^*\ell$, where ℓ stands for both $\ell^\pm$. The dark grey blob (diagram on the left) describes the production of an on-shell heavy Majorana neutrino N in proton-proton collisions at LHC. The production is possible both with gauge interactions (first diagram on the right-hand side) and with four-fermion contact interactions (second diagram on the right-hand side).

67], i.e. produced via CI and then decay via GI, and in the channel $pp \to \ell\ell^* \to \ell\ell q\bar{q}'$ [68] where both production and decay proceed through CI. Neutral excited leptons have been also discussed in the literature and the corresponding phenomenology at LHC has been discussed in detail in the case of a heavy composite Majorana neutrino N^* [69]. A dedicated experimental analysis has been carried out by the CMS collaboration [42] on LHC data collected at $\sqrt{s} = 13$ TeV and looking for the process

$$pp \to \ell N_\ell^* \to \ell\ell q\bar{q}', \tag{2.23}$$

with dilepton (dielectrons or dimuons) plus diquark final states. The existence of N_ℓ^* is excluded for masses up to 4.60 (4.70) TeV at 95% confidence level, assuming $M = \Lambda$. Moreover, the composite Majorana neutrinos of this model can be responsible for baryogenesis via leptogenesis [70,71]. The phenomenology of other excited states has also been discussed in a series of recent papers [69, 72–78].

We emphasize that in all phenomenological studies referenced above,as well as all experimental analyses that have searched for excited states at colliders, it is customary to impose they constrain $M \leq \Lambda$ on the parameter space of the model. To the best of our knowledge unitarity has never been taken into account and/or discussed in connection with the effective interactions of the so-called excited states. The main goal of this work is to report instead that the unitarity bounds, as extracted from Eq. (2.21a)-(2.21b) and (2.22), are quite compelling and should be included in future studies of such effective composite models because they constrain rather strongly the parameter space. While we present an explicit calculation of the unitarity bound for heavy composite neutrino searches, we expect that similar bounds (i.e. equally compelling) would apply for excited electrons (e^*), muons (μ^*), and quarks (q^*). Indeed, the effective operators that describe the latter excited states have the very same structure as those referred to as the composite neutrinos.

2.1 UNITARITY IN SINGLE-EXCITED-FERMION PRODUCTION

For the derivation of the unitarity bound, we adopt a standard method that makes use of the optical theorem and the expansion of the scattering amplitude in partial waves. To specify the CI and GI Lagrangians for a definite situation, we consider the production of the excited Majorana neutrino at the LHC. However, we shall highlight when the results apply to other composite fermion states in the following.

The central object that we shall derive from the operators in the effective Lagrangians (2.27) and (2.28) is the interacting part of the S matrix, indicated with T in this letter. It enters the partial wave decomposition of the scattering amplitude as follows

$$\mathcal{M}_{i\to f}(\theta) = 8\pi \sum_j (2j+1) T^j_{i\to f} d^j_{\lambda_f\,\lambda_i}(\theta)\,, \tag{2.24}$$

where j is the eigenvalue of the total angular momentum J of the incoming (outgoing) pair, $d^j_{\lambda_f\,\lambda_i}(\theta)$ is the Wigener d-function and λ_i (λ_f) is the total helicity of the initial (final) state pair. Without loss of generality, we consider azimuthally symmetric processes and fix $\phi = 0$ accordingly. From the optical theorem and the decomposition in Eq. (2.24), one can find the perturbative unitarity condition of an inelastic process for each j to be

$$\sum_{f\neq i} \beta_i \beta_f |T^j_{i\to f}|^2 \leq 1\,, \tag{2.25}$$

where β_i (β_f) is obtained from the two-body phase space and reads for two generic particles with masses m_1 and m_2

$$\beta = \frac{\sqrt{[\hat{s} - (m_1 - m_2)^2]\,[\hat{s} - (m_1 + m_2)^2]}}{\hat{s}}\,. \tag{2.26}$$

It corresponds to the particle velocity when $m_1 = m_2$. It is important to notice that the unitarity bound is imposed on the subprocess involving the proton valence quarks as initial state, namely $q\bar{q}' \to \ell N^*_\ell$ as shown in Figure 2.3. Then, for the process of interest, the relevant interaction(s) are as follows:

$$\mathcal{L}_{\rm CI} = \frac{g_*^2\,\eta}{\Lambda^2} \bar{q}'\gamma^\mu P_L q\, \bar{N}\gamma_\mu P_L \ell + h.c.\,, \tag{2.27}$$

$$\mathcal{L}_{\rm GI} = \frac{g\,f}{\sqrt{2}\Lambda} \bar{N}\sigma^{\mu\nu}(\partial_\mu W^+_\nu) P_L \ell + h.c.\,, \tag{2.28}$$

with the LH projector defined as $P_L = (1-\gamma_5)/2$. Accordingly, in Eq. (2.26), $\hat{s}$ denotes the center-of-mass energy in each collision and it is obtained from the nominal collider energy and the parton momentum fractions as $\hat{s} = x_1 x_2\, s$. As far as the kinematics are concerned, $\beta_i = 1$ can be used since the valence quark masses are negligible with respect to the center-of-mass energy. Instead, one finds $\beta_f = 1 - M^2/\hat{s}$ for the final state, where the composite neutrino mass has to be kept.

The core of the method relies on the derivation of the amplitude for the process of interest induced by the contact and gauge-mediated ones effective Lagrangians (2.27) and (2.28). Then, one matches the result so-obtained for $\mathcal{M}_{i\to f}$ with the r.h.s of Eq. (2.24) and extracts the corresponding $T^j_{i\to f}$ for each definite eigenvalue of the total angular momentum (j). The latter is inserted into Eq. (2.25) to derive the unitarity condition that the model parameters (Λ, M, g_*, g) and the center-of-mass energy have to obey. To this end, the amplitude $\mathcal{M}_{i\to f}$ is decomposed in terms of definite helicity states and, therefore, we have to express the initial and final state particles spinors accordingly [79]. The helicity of each particle in the initial or final state is $\lambda = \pm 1/2$, being all the involved particles fermions (also the composite neutrinos are spin-1/2 fermion). We shall simply use $\pm$ to label the initial and final state helicity combinations, $(+,+)$, $(+,-)$, $(-,+)$ and $(-,-)$. Since in the center-of-mass frame the incoming and outgoing particles travel in opposite directions, the helicities in the Wigner d-functions are defined as $\lambda_i = \lambda_q - \lambda_{\bar{q}'}$ and $\lambda_f = \lambda_{N^*} - \lambda_\ell$. One can adopt two different bases for expressing the spinors and the gamma matrices, the Dirac and chiral bases (see e.g. appendix in ref. [52]). We used both the options to derive $T^j_{i\to f}$ and checked that our findings are indeed invariant upon the choice of the basis.

We give the result for the CI Lagrangian in Eq. (2.27) first. The non-vanishing helicity amplitudes read

$$T^{j=1}_{(-,+)\to(-,+)} = -\frac{\hat{s}\, g_*^2}{12\pi\Lambda^2}\left(1-\frac{M^2}{\hat{s}}\right)^{\frac{1}{2}}, \tag{2.29}$$

$$T^{j=1}_{(-,+)\to(+,+)} = \frac{\sqrt{\hat{s}}\, M\, g_*^2}{12\sqrt{2}\pi\Lambda^2}\left(1-\frac{M^2}{\hat{s}}\right)^{\frac{1}{2}}. \tag{2.30}$$

Only the amplitude with $j = 1$ is non-zero, due to the initial helicity state. The same occurs with the vector and axial-vector operators studied in [52] for dark matter pair production at colliders. We notice that a finite composite neutrino mass allows for the helicity flip in the final state originating the term in Eq. (2.30). We obtain the same result if we work with right-handed particles in the CI operator, however the helicities in Eqs. (2.29) and (2.30) flip as $+ \leftrightarrow -$. Using Eq. (2.24) and summing over the non-vanishing final helicity states, we obtain

$$\frac{g_*^4\, \hat{s}\,(2\hat{s}+M^2)}{288\pi^2\Lambda^4}\left(1-\frac{M^2}{\hat{s}}\right)^2 \leq 1\,. \tag{2.31}$$

As far as the GI process is concerned, we proceed the same way. A dimension-5 operator is involved and, in this case, the W boson mediates the scattering between the initial and final states. We keep the W boson mass in our expression, even if it is much smaller than the typical $\hat{s}$ values of the pp collisions. The SM electroweak current enters besides the one from the composite model and the helicity amplitudes are found to be

$$T^{j=1}_{(-,+)\to(-,+)} = -\frac{ig^2}{24\pi\Lambda}\frac{\hat{s}^{3/2}}{\hat{s}-m_W^2}\left(1-\frac{M^2}{s}\right)^{\frac{1}{2}}, \tag{2.32}$$

$$T^{j=1}_{(-,+)\to(+,+)} = \frac{ig^2}{24\sqrt{2}\pi\Lambda}\frac{\hat{s}\, M}{\hat{s}-m_W^2}\left(1-\frac{M^2}{\hat{s}}\right)^{\frac{1}{2}}, \tag{2.33}$$

and the corresponding result for the unitarity bound is

$$\frac{g^4}{1152\,\pi^2\Lambda^2}\frac{\hat{s}^2\,(2\hat{s}+M^2)}{(\hat{s}-m_W^2)^2}\left(1-\frac{M^2}{\hat{s}}\right)^2 \leq 1\,. \tag{2.34}$$

A comment is in order. The unitarity bound in Eq. (2.31) is valid for the more generic production process $q\bar{q}' \to f^* f$, i.e. excited charged or neutral leptons and excited quarks accompanied by a SM fermion. This statement traces back to the particle-blind choice adopted in the CIs framework, where the η's are set to unity in all the cases. More care has to be taken about wider applicability of the result for GIs in Eq. (2.34). Here, different factors can enter according to the gauge couplings and gauge bosons that describe the processes involving excited charged leptons and quarks instead of composite neutrinos.

2.2 IMPLEMENTING THE BOUND

The production of heavy composite Majorana neutrinos has been studied by the CMS Collaboration by studying the final state with two leptons and at least one large-radius jet, with data from pp collisions at $\sqrt{s} = 13$ TeV and with an integrated luminosity of 2.3 fb^{-1} [42]. Good agreement between the data and the SM expectations was observed in the search, but the whole dataset of Run 2 of the LHC still needs to be analyzed. Therefore, the issue of the unitarity condition on the accessible parameter space (M, Λ) urges to be assessed. As usual in BSM searches, the absence of a signal excess over the SM background is translated into an experimental bound on the parameter space (M, Λ). Moreover, the sensitivity of this search was investigated for two future collider scenarios: the High-Luminosity LHC (HL-LHC), with a center-of-mass energy of 14 TeV and an integrated luminosity of 3 ab^{-1}, and the High-Energy LHC (HE-LHC), with a center-of-mass energy of 27 TeV and an integrated luminosity of 15 ab^{-1} [80]. The projection studies, included in the recent Yellow Report CERN publication [81,82], have shown the potential of such facilities in reaching much higher neutrino masses.

In this section, the perturbative unitarity bounds are applied to these searches in the dilepton and a large-radius jet channel with the CMS detector for the three different collider scenarios. Figure 2.4 presents the approach developed.

As already clear from the rather different coupling values entering the Lagrangians (2.27) and (2.28), namely $g/\sqrt{2} \approx 0.4$ versus $g_*^2 = 4\pi$, the production mechanism of a heavy composite neutrino and other excited states is dominated by the contact interaction mechanism [76]. In particular, it was shown that cross-sections in contact-mediated production are usually more than two orders of magnitude larger than the gauge mediated ones for all values of the Λ and M relevant in the analyses. This means that it is a reasonable approximation to consider only the bounds given in Eq. (2.31) to constrain the unitarity violation of the signal samples.

To estimate the effect of the unitarity condition on LHC searches, we need to implement the bounds in the case of hadron collisions. Then, the square of the center-of-mass energy of the colliding partons system, $\hat{s} = x_1 x_2 s$ does not have a definite value. To this aim, we have estimated $\hat{s}$ in each event generated in the Monte Carlo (MC) samples, and we have input the result into Eq. (2.31) to obtain level

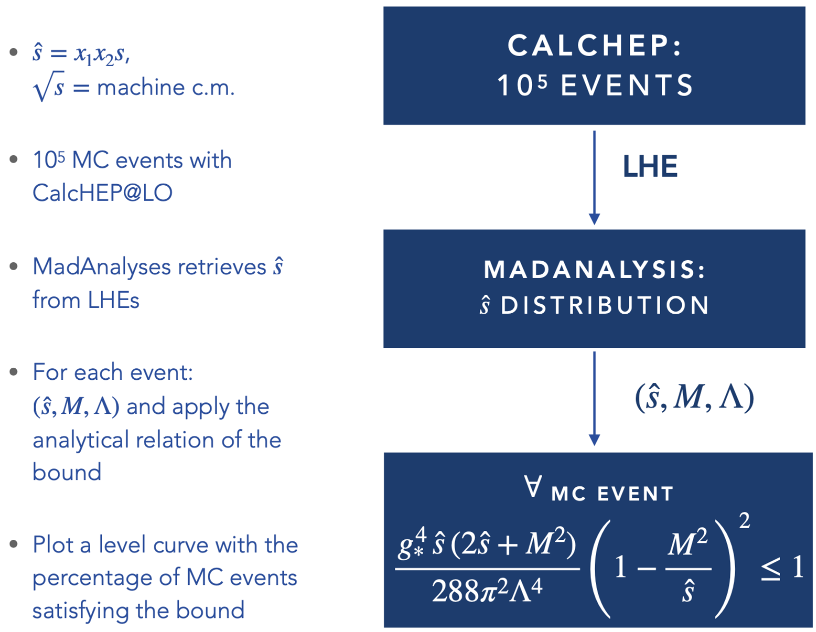


Figure 2.4: Scheme of implementation of the unitarity bound in Eq. 2.31 for the production of heavy composite Majorana neutrino at the LHC.

curves on the parameter space for which the unitarity bound is satisfied to some extent. Indeed, the constraint in Eq. (2.31) should not be interpreted too strictly. A violation of such bounds would signal the breakdown of the EFT expansion and call for higher-order operators in $\sqrt{\hat{s}}/\Lambda$ to help in restoring the unitarity of the process. Therefore, we implement a theoretical uncertainty by allowing up to 50% of the events to violate the bound, that corresponds to assigning a relative correction to the cross-section $\delta\sigma/\sigma_{\rm LO} \lesssim 0.5$ from higher-order terms.

The MC samples for the signal are generated at Leading Order (LO) with CalcHEP (v3.6) [83] for $\sqrt{s} = 13, 14$ and 27 TeV proton-proton collisions, using the NNPDF3.0 LO parton distribution functions with the four-flavor scheme [84], spanning over the (Λ, M) region covered by the experimental searches [80]. The information on the parton momenta is then retrieved from the Les Houches Event (LHE) files of each signal process through MadAnalysis [85]. We have explicitly checked that, in the mass range explored, the $\hat{s}$-distributions in our MC simulations are peaked at values around M^2 almost irrespective of the nominal collider energy $\sqrt{s} = 13, 14, 27$ TeV. This is somehow expected on general grounds since in the generated signal events the available energy, $\sqrt{\hat{s}}$, is mostly used to produce a heavy excited state (of mass M). Of course, the larger the collider energy the more prominent the distribution tails at high $\hat{s}$ values. These expectations are corroborated by analytical expressions for the $\hat{s}$-distributions that can be retrieved from ref. [69], involving only the product of the parton luminosity functions and the hard production process cross-section.

The results are presented in Figs. 2.5, 2.6 and 2.7 for the Run 2, HL-LHC and HE-LHC scenario respectively. The regions below the solid (violet) lines in Figs. 2.5-2.7 delimit the parameter space for which respectively 100%, 95% and 50% of the events satisfy the unitarity bound, i.e. they define the regions where the model should not

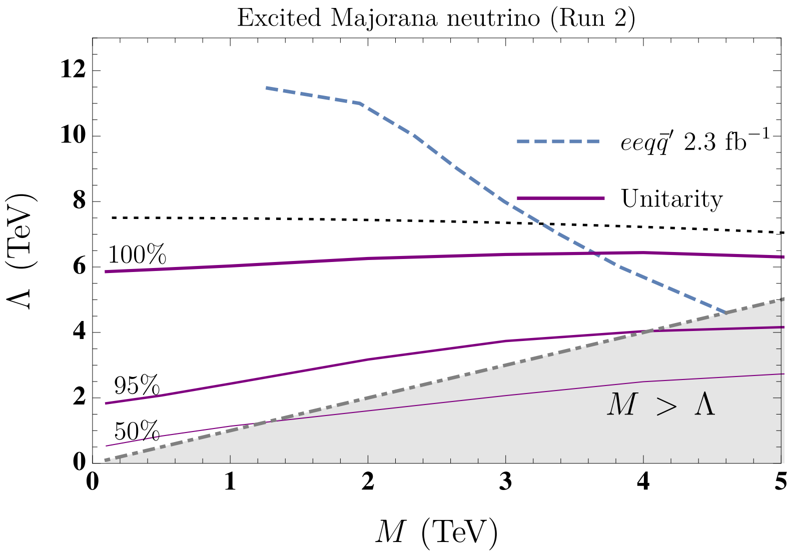


Figure 2.5: The unitarity bound in the (M, Λ) plane compared with the Run 2 exclusion at 95% CL from [42], dashed line (blue), for the $eeq\bar{q}'$ final state signature. The solid (violet) lines with decreasing thickness represent the unitarity bound respectively for 100%, 95% and 50% event fraction satisfying Eq. 2.31. The dot-dashed (gray) line stands for the $M = \Lambda$ condition. Here and in the following figures both Λ and M start at 100 GeV, and the dotted (black) curve corresponds to the theoretical unitarity bound (Eq. 2.31 with $\hat{s} = s$).

be trusted because unitarity is violated for such (M, Λ) values. It is important to underline that the impact of the unitarity bound is strongly dependent on this fraction of events (f) that satisfy the condition in Eq. (2.31). This conforms with the results in [52], at least in the (Λ, M) region considered here. Moreover, we observe that there is a value of the compositeness scale Λ, which depends on the parton collision energy $\sqrt{\hat{s}}$ and the excited fermion mass M, above which the unitarity bound saturates. Indeed, one can estimate an upper bound for such a value from Eq. (11) by setting the collision energy $\sqrt{\hat{s}} = \sqrt{s}$; it is represented with the dotted (black) line in Fig. 2.5-2.8 for the corresponding nominal energies $\sqrt{s} = 13$, 14, 27 TeV. An approximated (maximal) value of $\Lambda \approx \sqrt{s/3}$, which saturates the unitary bound, is obtained when $s \gg M^2$.

2.3 DISCUSSION AND RESULTS

Let us elaborate on our findings and explain their impact on the experimental analyses carried out at the LHC. First of all, the experimental outcomes are summarized with exclusion regions in the (M, Λ) plane, which are in turn set with the 95% C.L. observed (Run 2) [42] and expected limit (HL/HE-LHC) [81,82], namely the dashed blue lines in Figure 2.5, 2.6 and 2.7 respectively. Above these lines, the model is still viable, whereas below it is excluded. The experimental collaborations quote routinely the largest excluded excited-state mass by intersecting the 95% C.L. exclusion curves with the $M \leq \Lambda$ constraint (dot-dashed gray line and gray shaded region in Figure 2.5, 2.6 and 2.7). This is the widely discussed condition imposed on the model validity and it originates from asking the heavy excited states to be at most as heavy as the new physics scale Λ. Despite being a reasonable constraint,

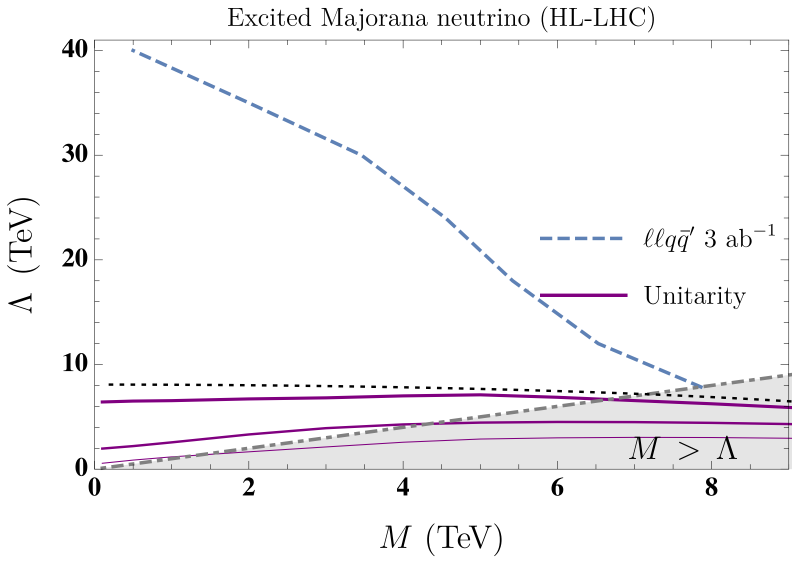


Figure 2.6: The unitarity bound in the plane (M, Λ) for the three event fractions as in Fig. 2.5 compared with the exclusion from the High Luminosity projections study in [81] for LHC at $\sqrt{s} = 14$ TeV at 3 ab^{-1} of integrated luminosity.

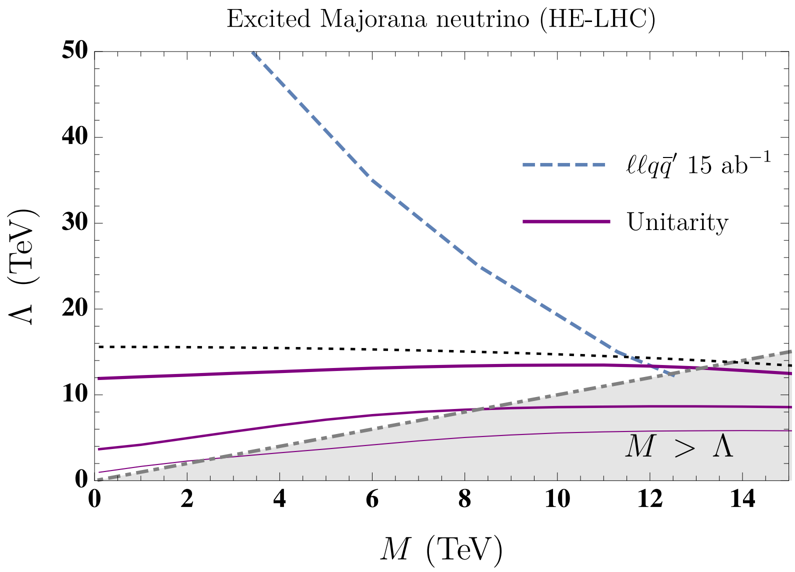


Figure 2.7: The unitarity bound in the plane (M, Λ) for the three event fractions as in Fig. 2.5 compared with the exclusion curve from the HE-LHC projection studies in [81] for $\sqrt{s} = 27$ TeV at 15 ab^{-1} of integrated luminosity.

it does not take into account the typical energy scale that enters the production process, i.e. $\sqrt{\hat{s}}$.

Comparing the unitarity bounds, as represented by the solid violet lines in Figure 2.5, 2.6 and 2.7, with the usual prescription of $M \leq \Lambda$ we can frame the interplay between the two. For small values of the heavy neutrino masses (M), the unitarity bound is more restrictive than $M \leq \Lambda$ and it shrinks the available parameter space quite considerably (at least for $f = 100\%$ and 95%). On the other hand, for higher values of the masses, the two become compelling. The relative importance depends on both the collider nominal energy and event fraction f. If one applies

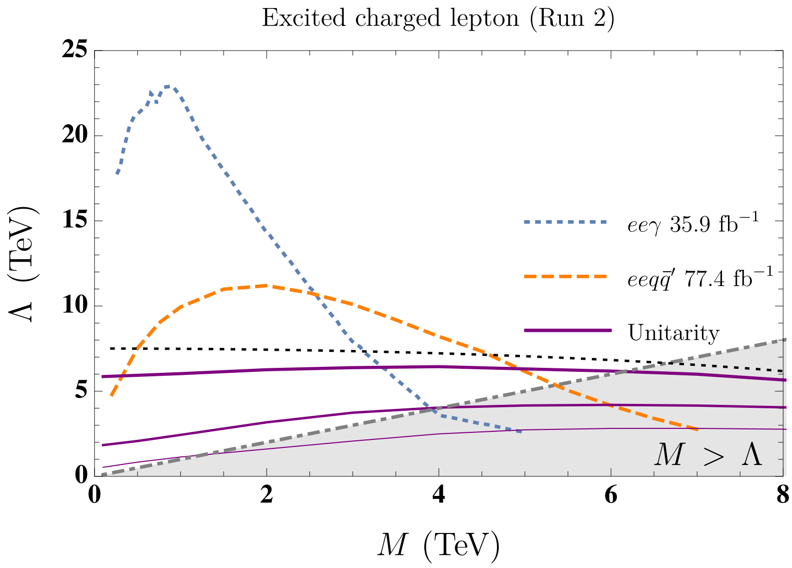


Figure 2.8: The unitarity bound in the plane (M, Λ) for the three event fractions as in Fig. 2.5 compared with the exclusion from the Run 2 for charged leptons searches with two different final states [67, 68].

	LHC Run 2 (N^*) 2.3 fb^{-1}, $\sqrt{s} = 13$ TeV	LHC Run 2 (e^*) 35.9 fb^{-1}, $\sqrt{s} = 13$ TeV	LHC Run 2 (e^*) 77.4 fb^{-1}, $\sqrt{s} = 13$ TeV
$M = \Lambda$	$M \leq 4.6$ TeV [42]	$M \leq 4.0$ TeV [67]	$M \leq 5.5$ TeV [68]
Unitarity 100%	$M \leq 3.6$ TeV ($\Lambda = 6.4$ TeV)	$M \leq 3.3$ TeV ($\Lambda = 6.5$ TeV)	$M \leq 4.9$ TeV ($\Lambda = 6.4$ TeV)
Unitarity 50%	–	$M \leq 4.9$ TeV ($\Lambda = 2.8$ TeV)	$M \leq 7.0$ TeV ($\Lambda = 2.9$ TeV)

Table 2.1: In the first line we quote the bounds reported in the CMS analysis of like sign dilpetons and diquark for excited neutrinos [42] and the bounds from CMS for two analyses for excited charged leptons [67,68]. In the second (third) line, we quote instead the strongest mass bound obtained from Figs. 2.5 and 2.8 when the perturbative unitarity bound with $f = 100\%$ (50%) crosses the 95% C.L. exclusion curve from the experimental studies.

the unitarity bound to the experimental results by following the same prescription as outlined before for $M \leq \Lambda$, then the maximal neutrino mass values are those collected in Table 2.1, second row. For example, for LHC Run 2, we find $M \leq 3.6$ TeV for $\Lambda = 6.4$ TeV instead of $M \leq 4.6$ TeV ($\Lambda = 4.6$ TeV), when the unitarity bound is required to be satisfied by the 100% of events. As anticipated, the new constraint set by the unitarity bound offers an alternative theoretical input for ongoing and future experimental analyses on this effective composite model. We provide the unitarity bound down to $M = 100$ GeV. Smaller values clash with the original model setting that assumes some new physics above the electroweak scale triggering fermion excitations [32, 33].

While in this study we have concentrated on the impact of unitarity bounds on the heavy composite neutrino production at the LHC, HL-LHC, and HE-LHC, we expect that similar bounds will affect the searches for charged excited leptons.

Leaving more detailed studies for future work, we apply the unitarity bound in Eq. (2.31) for the recent experimental results reported in [67, 68], where charged excited leptons are produced via CI in the process $pp \to \ell^*\ell$, with $\ell^* = e^*, \mu^*$. Being the effective Lagrangian for the production mechanism the very same as for excited Majorana neutrinos (if one insists on $\eta = 1$ for the different states), we can use the unitarity bound as extracted for $\sqrt{s} = 13$ TeV and apply it for the process involving charged excited leptons. The comparison with the experimental exclusion limit at 95% C.L. is shown in Figure 2.8. As anticipated, the comparison with the observed and expected limits produces different mass reaches. For the case of LHC Run 2 searches recently performed in CMS, we quote the corresponding mass values in Table 2.1 for the searches of excited charged leptons. As for the analyses on the excited charged states, we can exploit the data as provided in the region $M > \Lambda$ and inspect the interplay with the unitarity bound for $f = 100\%, 95\%, 50\%$. On the other hand, we cannot provide as many mass values for the excited neutrino searches, due to the lack of experimental data in the same region $M > \Lambda$. When $f = 100\%$ is considered, the largest excited lepton mass reads 3.3 TeV (4.9 TeV) instead of 4.0 TeV (5.5 TeV) for the $ee\gamma$ ($eeq\bar{q}'$) final state.

In conclusion, we studied the perturbative unitarity bound extracted from the effective gauge and contact Lagrangians for a composite-fermion model. On general grounds, an effective theory is valid up to energy/momentum scales smaller than the large energy scale that sets the operator expansion. Since collider experiments are involving more and more energetic particle collisions, the use and the applicability of effective operators can be questioned. To address this issue and to be on the safe side, one can impose the unitarity condition both on the EFT parameters (M, Λ, g^*, g) and the energy involved in a given process. To the best of our knowledge, such a constraint was not derived for the model Lagrangians in Eq. (2.27) and (2.28), and we have obtained the corresponding unitarity bounds, namely Eqs. (2.31) and (2.34). Thus, the applicability of the effective operators describing the production of composite neutrinos (and other excited states) has to be restricted accordingly. We have considered an estimation of the theoretical error on the unitarity bound, which is derived at leading order in the EFT expansion, by allowing up to 50% of the events to evade the constraint. This originates the violet lines in Figures 2.5-2.8.

Based on the results, further investigations may be devoted to a better understanding of the theoretical error. Indeed the strong dependence of the unitarity bound on the fraction of events f, especially so in the low-mass region, calls perhaps for an estimate of possible higher-order terms in the effective theory expansion (operators of dimension-7 for contact interactions). In doing so, one could pinpoint a particular choice of f more rigorously.

Nonetheless, we will see in Chapter 4 how the findings here discussed will have a significant impact on ongoing and future experimental searches for excited states coupling to the SM fermions with the interactions given in (2.27) and (2.28). At the very least, the unitarity bounds play the role of a collider-driven theoretical tool for interpreting the experimental results of the considered composite models, more rigorous than the simple relation $M \leq \Lambda$.

Part II

EXPERIMENTAL SEARCHES LHC RUN 2 DATA WITH THE CMS DETECTOR

3 EXPERIMENTAL SETUP: THE CMS DETECTOR

In this section, we provide an introduction of the Large Hadron Collider (LHC) and its facilities, with an emphasis on the Compact Muon Solenoid (CMS) experiment. The data taken from the second Run started in 2016 and concluded in 2018, a collection of approximately 140 fb^{-1} of integrated luminosity of proton-proton collisions with a center-of-mass energy of 13 TeV, are then used in two distinct searches in Chapter 4.

1 THE LARGE HADRON COLLIDER

The LHC is the world's largest and most powerful particle accelerator, built overlapping the France and Swiss border, near Geneva. It consists of a ring-shaped tunnel with a circumference of 27 km, placed to an average depth of 100 m. This facility can accelerate proton beams up to energies of 7 TeV per particle or heavy nuclei up to energies of 5.74 TeV per nucleus. The tunnel contains two adjacent pipes crossing each other at four points. The particles bunches travel in the beam pipes in opposite directions and collide in the intersection points. The particles are bend on the curvilinear trajectory using 1232 magnetic dipoles, while 392 magnetic quadrupoles maintain the beams focused for the maximization of the interaction probability of the particles in the intersection points. These superconducting magnets are composed of a niobium and titanium alloy and are maintained at the operating temperature of 1.9 K with liquid helium. Before entering the LHC, the particles pass through a chain of accelerators that gradually bring the beam to higher energies (Figure 3.2, on the right). Each technology of particle acceleration has maximum and minimum

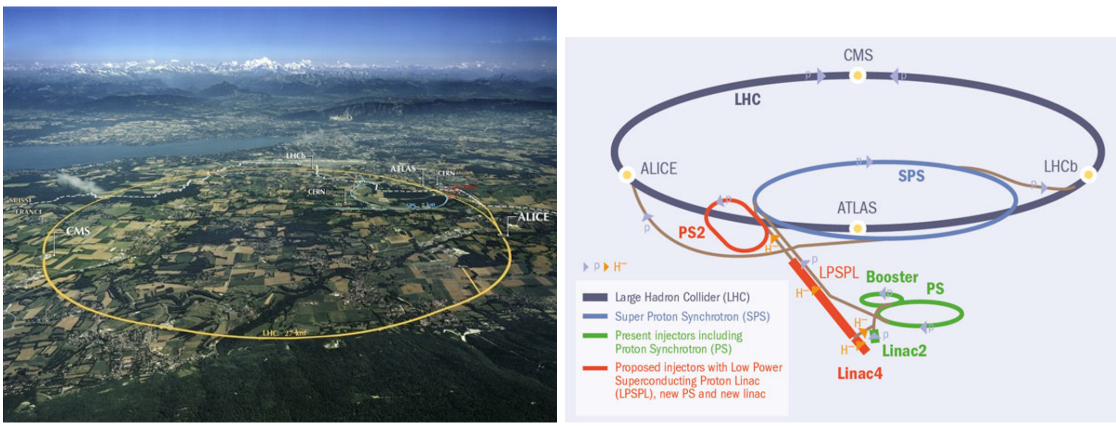


Figure 3.1: On the left, the aerial view of the CERN LHC and the scheme [86]. On the right, the scheme of the accelerator chain at CERN.

limits of operating energy. The particles at low energy are produced by two linear accelerators, called LINAC, one for the protons and one for the heavy nuclei. Then the particles go through the ring accelerators, first the PS booster, then the Proton Synchrotron, then the Super Proton Synchrotron, and finally, they enter the LHC. Along the LHC ring, there are six principal experiments. The two largest are CMS and ATLAS [87], the former being described in detail in the next section. Among the others, we can mention the LHCb experiment [88], mainly implemented to study the physics of the B mesons, and ALICE [89], suited for the study of heavy ions collisions.

2 THE CMS EXPERIMENT

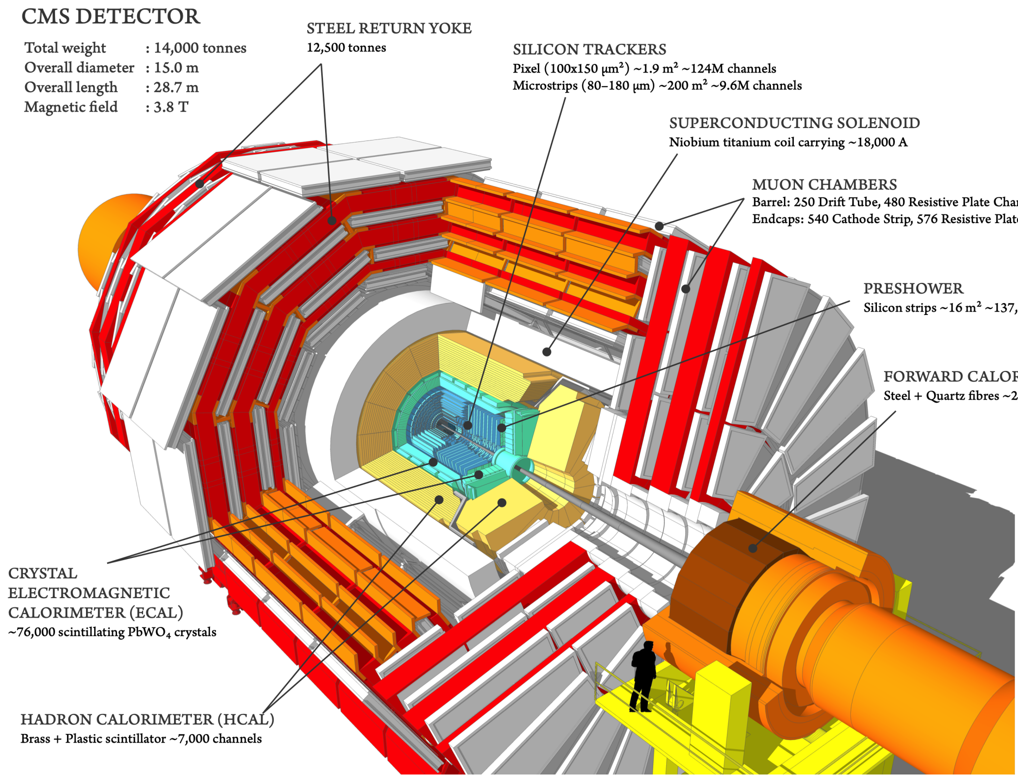


Figure 3.2: The Run 2 CMS detector and its main components [90].

2.1 CMS DETECTOR COMPONENTS

The symmetrical CMS detector [91–94] is built in a traditional barrel design with two endcaps to provide nearly 4π coverage. As the name suggests, the CMS detector has a compact design and is specifically built to provide good muon detection and resolution. The particle identification is aided by the superconducting solenoid that provides a 4 Tesla magnetic field. CMS is composed of a series of layers of sub-detector, each one with a specific function. It is a cylindrical structure symmetric around the interaction point and the beam pipe. The innermost layer is a silicon-based tracker topped by an onion-shaped scintillating crystal Electromagnetic Calorimeter (ECAL). The ECAL is then surrounded by a Hadron Calorimeter (HCAL) followed by the outermost layer, consisting of systems designed for the muon detection. Figure 3.2 introduces the CMS modular design, which is briefly described, from the inside out, in all its main components throughout this section [95].

2.2 INNER TRACKER

The inner tracker [96] is housed around the LHC beam pipe. It is designed to precisely determine the trajectory of charged particles, with high momentum resolution and efficiency better than 98% in the range $|\eta|$ <2.5. Combined with measurements from the dedicated muon detection system, the tracking system contributes to high-resolution muon measurements in CMS. Cylindrical in shape, the tracker consists of two subsystems: a pixel detector that extends radially from 4.4 cm to 10.2 cm away from the beamline, and a silicon strip tracker that covers the radii between 20 cm and 116 cm. The pixel detector has an intrinsic spatial resolution of about 20 μm along the z-axis and up to 10 μm in the $r - \phi$ direction and is placed close to the interaction region to improve the measurement of the impact parameter of charged-particle tracks, as well as the position of secondary vertices. It is divided into 3 barrel layers, placed at radial distances of 4.4 cm, 7.3 cm, and 10.2 cm from the beamline, and 2 endcap disks on each side of the barrel located at $\pm$34.5 cm and $\pm$46.5 cm from the interaction point. The single-pixel cells of a size of $100 \times 150\, \mu m^2$, are arranged in modules for a total of 1440. The pixel tracker is surrounded by the silicon strip tracker composed of the inner strip tracker and the outer strip tracker. The four different subsystems contain in total 15148 detector modules carrying either one thin (320 μm) or two thick (500 μm) silicon sensors. The tracker inner barrel contains four concentric cylinders placed at radii of 25.5 cm, 33.9 cm, 41.9 cm, and 49.8 cm respectively covering the range $|z| < 70$ cm. The tracker inner disks comprise 3 identical disks placed within the segments 80 cm $< |z| <$ 90 cm on each side of the inner barrel, and spanning a radial region from 20 cm to 50 cm. The tracker outer barrel is instead a 218 cm long single mechanical structure with inner and outer radii of 55.5 cm and 116 cm respectively, composed of four identical disks joined by three outer and three inner cylinders. The outer endcaps include 18 disks (9 on each side) extended radially from 22 cm to 113.5 cm and placed from $\pm$124 cm to $\pm$280 cm along the z-direction.

2.3 CALORIMETRY

Electrons, photons, and hadrons will deposit their energies in the calorimeters for accurate energy measured. The inner calorimeter layer, the Electromagnetic Calorimeter [97], is designed to measure electrons and photons' energy deposits with high precision. Strongly-interacting particles and hadrons lose most of their energy in the second layer, the Hadron Calorimeter [98]. Muons deposit only a tiny fraction of their energy in the calorimeters, and thus their detection is obtained by combining the information from the tracking and the muon detector subsystems. Neutrinos escape detection, and we can infer their presence from an apparent energy imbalance in the interaction.

2.4 ELECTROMAGNETIC CALORIMETER (ECAL)

CMS has a very compact scintillating crystal calorimeter that offers excellent energy resolution performance since almost all of the energy of electrons and photons is transferred within the crystal volume. The electromagnetic calorimeter (ECAL) uses 75848 lead-tungstate ($PbWO_4$) crystals with coverage in pseudorapidity up to $|\eta| < 3.0$. The majority of these crystals (about 61200) structure the central ECAL barrel ($|\eta| << 1.48$), while the rest are in the two endcap regions. The scintillation light coming from the interactions with the crystal is detected by silicon avalanche photodiodes (APDs) in the barrel region and vacuum photodiodes (VPTs) in the endcap region. A pre-shower system is installed in front of the endcap ECAL for the π^0 mesons rejection and the detection of photons. It is designed to detect the slightly-overlapping pair of photons from a short-lived π^0 decay and to distinguish these from the shower produced from a single photon interaction in the calorimeter material. The pre-shower detector is composed of lead absorbers and silicon strips sensors and spans the fiducial region of $1.65 < |\eta| < 2.6$.

2.5 HADRONIC CALORIMETER (HCAL)

The ECAL is surrounded by a brass/scintillator sampling Hadron Calorimeter (HCAL) with coverage up to $|\eta| < 3.0$. More precisely, the HCAL is located between the ECAL and the magnet coil (1.77 m $< r <$ 2.95 m), plus an outer hadron calorimeter (HO) installed outside the solenoid to ensure an adequate thickness in the central region. The remaining three components of the HCAL system are the inner barrel (HB), the endcaps (HE), and the forward calorimeter (HF) that provides coverage up to $|\eta| = 5.0$ to ensure full geometric coverage for the measurement. The barrel and endcap component are sampling calorimeters with 50 mm thick copper absorber plates interleaved with 4 mm thick scintillator sheets. The scintillation light is converted by wavelength-shifting (WLS) fibers embedded in the scintillator tiles and channeled to photodetectors via clear fibers. This light is detected by novel photodetectors (hybrid photodiodes, or HPDs) that can provide gain and operate in high axial magnetic fields. In the harsher radiation environment of the forward systems, the steel absorber plates are used and the hadronic showers are seen by radiation-resistant quartz-fibers. The Cerenkov light emitted in the quartz fibers is detected by conventional photomultiplier tubes. The HCAL plays an essential role in the identification and measurement of quarks, gluons, and neutrinos by measuring the energy and direction of jets and missing transverse energy flow in events. The HCAL depth is crucial to contain high-energy jets and to filter out hadrons from jets involving muons. As we will see in the next chapter, the accurate measurements of high-energy jets are a fundamental tool for searches for high mass SM and BSM processes.

2.6 SUPERCONDUCTING MAGNET

At the heart of CMS sits a 13-m-long, 6 m inner diameter, 12,000 ton, 3.8 T superconducting solenoid [99] which in part gives CMS its name. The CMS magnet system consists of a superconducting coil, the magnet yoke (barrel and endcap), a vacuum tank, and ancillaries such as cryogenics, power supplies, and process controls. Inside the bore of the magnet sit the inner tracker and the calorimetry, while outside is the flux return system and muon detector. In order to achieve good momentum resolution with a compact spectrometer, it is necessary to have a combination of a high magnetic field and high precision on the spatial resolution and alignment of the detectors. The momentum measurement of charged particles in the detector is based on the bending of their trajectories. The magnetic field also brings benefits for calorimetry - it enables the detection of many isolated electrons produced by the decays of W's, Z's, and b quarks. Using these electrons, the CMS crystal electromagnetic calorimeter can be calibrated to an accuracy of a fraction of a percent. The magnetic flux is returned via a 1.5 m thick saturated iron yoke instrumented with four stations of muon chambers [100].

2.7 MUON DETECTION

Muons provide a clean signature for a wide range of physics processes. The return field of the solenoid is large enough to saturate 1.5 m of iron, allowing 4 muons stations to be interleaved with the iron return yoke plates of the magnet system. This grants the muon system a reconstruction efficiency better than 98% over the full pseudorapidity range. Each muon station consists of several layers of aluminum drift tubes (DT) in the barrel region and cathode strip chambers (CSCs) in the endcap region, complemented by resistive plate chambers (RPCs). The muon detectors are arranged in concentric cylinders around the beamline in the barrel region, and in disks perpendicular to the beamline in the endcaps. Muon identification is ensured by the large thickness of the absorber material (iron), which cannot be traversed by particles other than neutrinos and muons. The DT and CSC detectors are used to obtain a precise measurement of the position and thus the momentum of the muons, whereas the RPC chambers are dedicated to providing fast information for the Level-1 trigger. Muon measurements taken in the muon stations outside the solenoid are used for fast triggering and standalone muon reconstruction, but can also be combined with information from the inner tracking detectors for more precise reconstruction and momentum measurement.

2.8 TRIGGER AND DATA ACQUISITION SYSTEM

The LHC provides a collision rate of about 100 MHz, producing an incredible amount of data. The CMS trigger system consists of two steps. In the first step, the Level-1 trigger (L1), dedicated hardware is used to reduce at the minimum the dead time and to take a very quick accept/reject decision in order to cut down the data rate to almost 100 kHz. L1 selection is based completely on calorimeter and muon chamber information, which are the fastest available. Their signals are processed with hardware logical circuits and the selection is based on rapid identification of some physical objects (muons, electrons, photons, jets, or missing energy) with the appropriate thresholds. In case of a positive decision, data are temporarily stored and then passed to the second stage, the software-based High-Level Trigger (HLT) that reduces the rate further down to about 1 kHz. It relies on commercial processors, where several dedicated software algorithms select events coming from the L1 trigger. Considering the L1 trigger data occur at 10^5 Hz, and the total storage capacity is 10^2 Hz, a reduction factor of 10^3 is required. The HLT uses the response of all subdetectors to make it possible and to get complete information of the events for offline analysis.

4

SEARCHING FOR HEAVY COMPOSITE MAJORANA NEUTRINO

1 INTRODUCTION

We present the search for Heavy Majorana Neutrinos that stem from a BSM model based on a compositeness scenario. Different works have considered this possibility [77, 101–103], in particular in Ref. [104] a study of the phenomenology for the production and decay of these hypothetical particle at the CERN LHC is presented.

Compositeness of ordinary fermions is one possible scenario beyond the SM. In this context, quarks and leptons are assumed to have an internal substructure that should become manifest at some sufficiently high energy scale, the compositeness scale Λ. Ordinary fermions are thought to be bound states of some as–yet–unobserved fundamental constituents generically referred to as *preons*. Two model–independent properties [105–108] are relevant: the contact interaction, which complements the gauge interaction, and the existence of excited states of quarks and leptons with masses lower than or equal to Λ. The Heavy Composite Majorana Neutrino, N, would be a particular case of such excited states, and recently associated to the leptogenesis mechanism for explaining the matter-antimatter asymmetry in the Universe [70, 71].

The contact interaction represents an effective approach for describing the effects of the unknown internal dynamics of compositeness. It is described by an effective four-fermion Lagrangian of the type:

$$\mathcal{L}_C = \frac{g_*^2}{\Lambda^2}\frac{1}{2}j^{\mu}j_{\mu}, \tag{4.1}$$

with

$$j_{\mu} = \eta_L \bar{\phi_L}\gamma_{\mu}\phi_L + \eta\prime_L \bar{\phi^*_L}\gamma_{\mu}\phi^*_L + \eta\prime\prime_L \bar{\phi^*_L}\gamma_{\mu}\phi_L + h.c. + (L \to R),$$

where $g_*^2 = 4\pi$ and the η factors are usually set to unity [108]. This interaction contributes to the total production amplitude, which is the sum of contact and gauge interactions, as it is shown in Fig. 4.1. The gauge interaction between the SM fermions and the excited fermions is described by a magnetic type coupling

$$\mathcal{L}_G = \frac{1}{2\Lambda}\bar{\phi^*_R}\sigma^{\mu\nu}\left(gf\frac{\overrightarrow{\tau}}{2}\cdot\overrightarrow{W}_{\mu\nu} + g'f'YB_{\mu\nu}\right)\phi_L + h.c., \tag{4.2}$$

where g and g' are the $SU(2)_L$ and $U(1)_Y$ gauge couplings, f and f' are dimensionless couplings, which are simply assumed to be 1 [108], and ϕ and ϕ^* are the

SM and excited fermion fields.

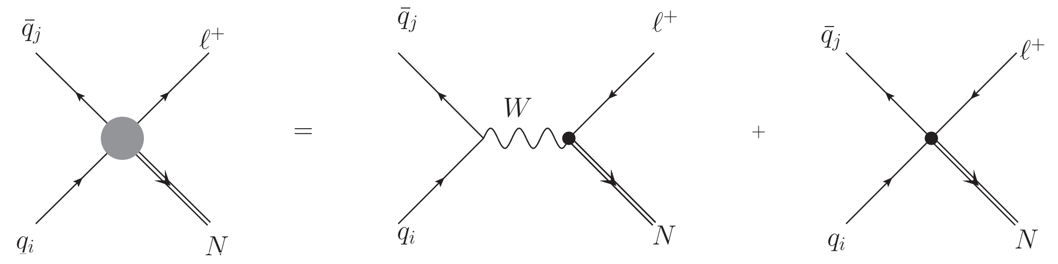


Figure 4.1: The fermion interaction as sum of the gauge and contact contributions.

The Heavy Composite Majorana Neutrino is its own antiparticle and, as such, can be produced and decay either as a neutrino or an antineutrino. It can be produced in association with a lepton, in pp collisions, via quark–antiquark annihilation ($qq \to \ell N$). This process can occur via both gauge and contact interactions, although the latter appears dominant for each value of Λ and mass of N, as shown in Fig. 4.2.

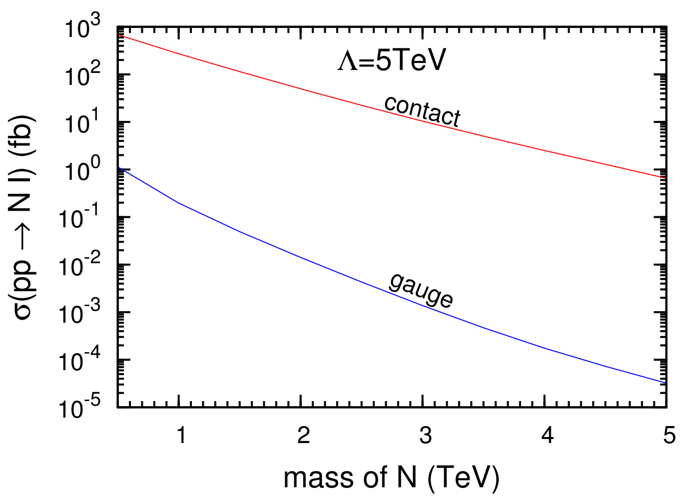


Figure 4.2: Production cross section of the Heavy Composite Majorana Neutrino for gauge and contact interaction at $\Lambda = 5$ TeV.

The Heavy Composite Majorana Neutrino can decay again through both gauge and contact interactions (Fig. 4.3), though in this case the dominant interaction changes depending on Λ and the mass of N, as can be seen from Fig. 4.4.

The possible decays are:

$$N \to \ell qq, \qquad N \to \ell^+\ell^-\nu(\bar{\nu}), \qquad N \to \nu(\bar{\nu})qq.$$

This implies that the possible final states are:

$$\ell\ell qq, \qquad \ell\ell\ell\nu(\bar{\nu}), \qquad \ell\nu(\bar{\nu})qq.$$

In this work the final state $\ell\ell qq$ will be considered, particularly the cases in which ℓ is either an electron or a muon, giving rise to the channels $eeqq$ and $\mu\mu qq$.
In Fig. 4.5 the Feynman diagram of the entire process $pp \to \ell N \to \ell\ell qq$ is shown.

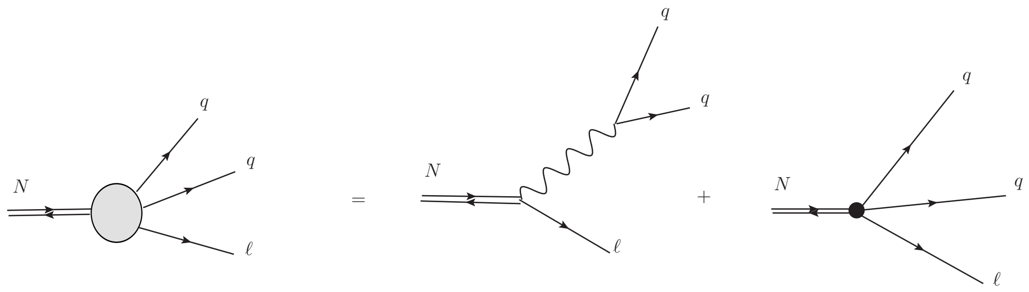


Figure 4.3: The Feynman diagram of the decay of a Heavy Composite Majorana Neutrino to the $\ell\ell qq$ final state.

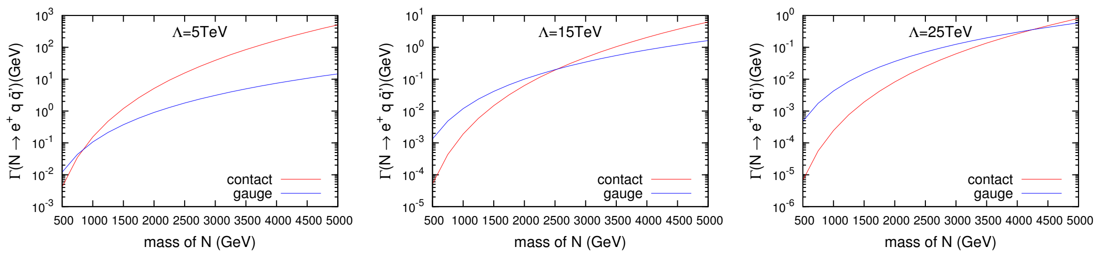


Figure 4.4: Decay amplitude of the Heavy Composite Majorana Neutrino for $\Lambda = 5$ TeV (left), $\Lambda = 15$ TeV (center), and $\Lambda = 25$ TeV (right), as a function of its mass.

A previous search for Heavy Majorana Neutrino was performed at CMS with $2.3\,\mathrm{fb}^{-1}$ at $\sqrt{s}$=13 TeV [109] finding good agreement between the data and the SM expectations. Consequently, a 95% confidence level (CL) upper limit, using the CLs criteria, on the Majorana neutrino mass was placed at about 4.6 TeV for both the electron and muon channel. Moreover, the sensitivity of this search was investigated for two future collider scenarios: the High-Luminosity LHC (HL-LHC), with a centre-of-mass energy of 14 TeV and an integrated luminosity of 3 ab^{-1}, and the High-Energy LHC (HE-LHC), with a centre-of-mass energy of 27 TeV and an integrated luminosity of 15 ab^{-1}. The projection studies, included in the recent Yellow Report CERN publication, have shown the potential of such facilities in reaching

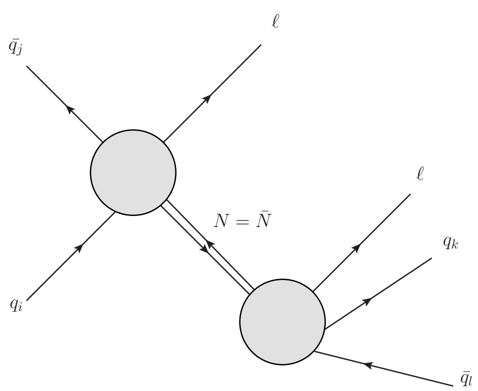


Figure 4.5: The Feynman diagram of the process for the production and decay of a Heavy Composite Majorana Neutrino, according to the decay chain $pp \to \ell N \to \ell\ell qq$.

much higher neutrino masses [80, 81]. Part III of this thesis will be totally devoted to these future collider projections.

The aim of the present work is to extend the previous version of the analysis to the full Run 2 data set, including data taken in 2016, 2017, and 2018. The samples collected in different years of data-taking are analyzed separately and then combined for the final result. Moreover, the distinctive feature of this analysis lies in the joint phenomenological effort, which we introduced in Section 2.2. The perturbative unitarity bounds in composite models serve, for the first time, as a collider-driven theoretical tool for the exploration of the parameter space in this experimental results, more rigorous than the simple relation $M^* \leq \Lambda$. The analysis benefits from these studies, and takes into account a re-optimization of the signal region. For a more exhaustive comprehension of the topic, we briefly take up again the effect of unitarity bounds in the experimental search for heavy composite neutrinos in the following part.

1.1 UNITARITY BOUNDS AND OPTIMIZATION OF THE EXPERIMENTAL SEARCH

In an EFT scenario, a limited set of operators can lead to compelling unitarity violation issues as the energy of the process approaches the cut-off scale of the low-energy theory, here the compositeness scale Λ. In order to avoid unphysical predictions from a certain EFT interaction, while working with a limited number of operators, a fully unitary prescription is necessary. In general, it is not straightforward to implement a unitarization method in experimental analysis. However, an indication about the applicability of the given effective operator(s) can be envisaged [45, 50–52].

In Ref. [110], the authors showed that the study of partial wave unitarity can play a significant role also for excited leptons searches at the LHC. The bounds obtained in the parameter space of dimension-5 and dimension-6 effective operators in Eqs. 4.1,4.2 link the model parameter space, (g^*, M, Λ), together with the partonic collision energy $\sqrt{\hat{s}}$, namely the energy involved in a given process. They found that the unitarity bound, so far overlooked, is quite compelling, and significant portions of the parameter space (M, Λ) become excluded if one insists on working with the leading effective operators. The unitarity bound complements the standard request $M \leq \Lambda$, and possibly suggests a rethinking of this original constraint on the model, not completely justifiable if being agnostic on the UV-complete theory (as for QCD, the emergent resonances may have masses greater than the scale). This is shown in Fig. 4.6 where the observed limits on the heavy composite Majorana neutrino in 2015 Run 2 are shown with unitarity bound curves corresponding to different percentages of MC events satisfying unitarity in the $\sqrt{s} = 13$ TeV proton-proton collisions. It is worth mentioning that a percentage smaller than 100% signals the breakdown of the effective description in terms of the operators in Eqs. 4.1,4.2. Therefore, the safe region is the one above the 100% line. The black dotted line is even more conservative, and it stands for the collision nominal energy $\sqrt{s} = 13$ TeV.

We could therefore conclude that the importance of unitarity bounds at low masses pushes the exploration of the (M, Λ) parameters space towards higher Λ region: this motivated the reconsideration of the signal region used in the 2015 data analysis that we present in Sec. 4.4.

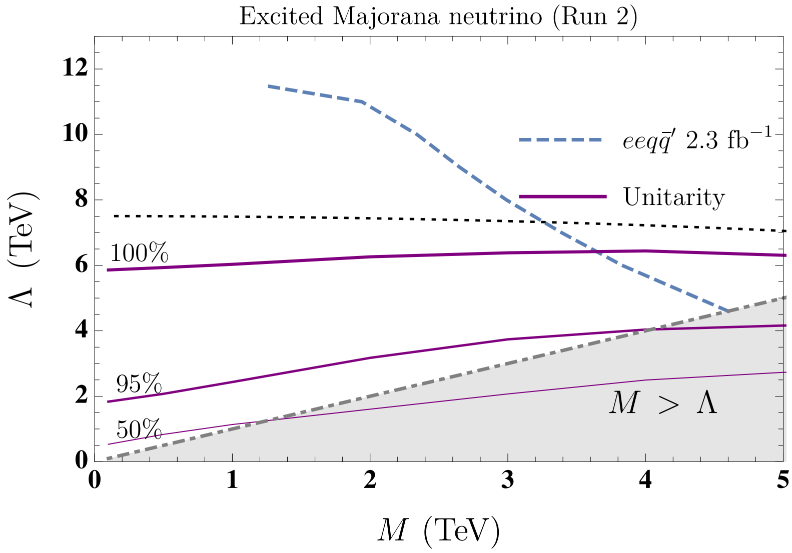


Figure 4.6: The unitarity bounds at different percentage (violet lines) in the (M, Λ) plane compared with the Run 2 exclusion at 95% CL from Ref. [110], dashed line (blue), for the electron final state. The solid (violet) lines with decreasing thickness represent the fraction of MC events that satisfy the unitarity condition in the EFT approximation.

2 DATA AND MONTE CARLO SAMPLES

2.1 DATA SAMPLES

The data used in this analysis were collected with the CMS detector during the 2016, 2017 and 2018 proton-proton runs of the LHC at 13 TeV. The data samples have been reconstructed with the following CMS software releases: CMSSW_9_4_9 (2016), CMSSW_9_4_5_cand1 (2017) and CMSSW_10_2_5_patch1 (2018). The framework used for the data processing runs on the CMSSW_10_2_16_UL release. To indicate the luminosity sections of a specific data run that are considered good to be processed,the CMS Collaboration maintains centralized "good run lists" stored as Java Script Object Notation (JSON) files [111]. For the present analysis, the events were preselected considering the following JSON files:

- 2016:
  Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt,
- 2017:
  Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON.txt,
- 2018:
  Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_PromptEraD_Collisions18_JSON.txt.

The samples considered for the three years are reported in Table 4.1, while on Table 4.2 the list of the trigger paths used for the three years are shown. The full sample corresponds to an integrated luminosity of approximately 35.5 fb^{-1}(2016) + 41.5 fb^{-1}(2017) + 58.9 fb^{-1}(2018) = 136 fb^{-1}.

2.2 MONTE CARLO SAMPLES

Several different Monte Carlo (MC) samples produced by the CMS Collaboration were used in this analysis. The specific detector and LHC run conditions of each year are taken into account in the MC simulations with different production "campaigns". The reference campaigns considered for the heavy composite neutrino search are the following three:

- 2016:
  RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v1,

- 2017:
  RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1,

- 2018:
  RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1(2,3) .

The list of samples for the signal processes, produced at the Leading Order (LO), is reported in Table 4.3 for the benchmark compositeness scenario of $\Lambda = 13$ TeV. The list of samples for the background processes is reported in Table 4.4. The listed names contain information about the generator that were used for the production of the corresponding processes. The $t\bar{t}$ and inclusive DY backgrounds are calculated at the Next-to-Next-to-Leading-Order (NNLO), the WW, WZ and ZZ backgrounds are calculated at the Next-to-Leading-Order (NLO), whilst the cross-sections of the NLO tW samples the LO $WJets$ are multiplied for a k-factor that is defined as inclusive NLO/LO. For further information see [112, 113].

The MC samples are adjusted according to the following corrections:

- Normalization to the integrated luminosity of the data according to the cross-section of the specific process, see Table 4.1,
- **Lepton scale factors weights**, as described in Section 4.3,
- **L1 prefiring correction**: in the 2016 and 2017 dataset, so-called "prefiring" is present. Prefiring is caused by L1 trigger primitives (summary data on the position and energy of particles) in the endcap regions of the ECAL that are assigned to an incorrect earlier bunch crossing due to a timing problem. In most cases, the events to which the trigger is assigned do not pass the HLT selection and are discarded, at which point the trigger rule system prevents the event to which the trigger originally belonged from being recorded. This effect leads to an overall decrease of the trigger efficiency that is not detectable with standard offline measurements. A solution has been developed by parametrizing the probability of a jet or photon the event to cause prefiring in terms of their

Data samples	Luminosity (fb^{-1})
2016 dataset	
/SingleElectron/Run2016B-17Jul2018_ver1-v1/MINIAOD /SingleElectron/Run2016B-17Jul2018_ver2-v1/MINIAOD /SingleElectron/Run2016C-17Jul2018-v1/MINIAOD /SingleElectron/Run2016D-17Jul2018-v1/MINIAOD /SingleElectron/Run2016E-17Jul2018-v1/MINIAOD /SingleElectron/Run2016F-17Jul2018-v1/MINIAOD /SingleElectron/Run2016G-17Jul2018-v1/MINIAOD /SingleElectron/Run2016H-17Jul2018-v1/MINIAOD /SingleMuon/Run2016B-17Jul2018_ver1-v1/MINIAOD /SingleMuon/Run2016B-17Jul2018_ver2-v1/MINIAOD /SingleMuon/Run2016C-17Jul2018-v1/MINIAOD /SingleMuon/Run2016D-17Jul2018-v1/MINIAOD /SingleMuon/Run2016E-17Jul2018-v1/MINIAOD /SingleMuon/Run2016F-17Jul2018-v1/MINIAOD /SingleMuon/Run2016G-17Jul2018-v1/MINIAOD /SingleMuon/Run2016H-17Jul2018-v1/MINIAOD	35.5
2017 dataset	
/SingleElectron/Run2017B-31Mar2018-v1/MINIAOD /SingleElectron/Run2017C-31Mar2018-v1/MINIAOD /SingleElectron/Run2017D-31Mar2018-v1/MINIAOD /SingleElectron/Run2017E-31Mar2018-v1/MINIAOD /SingleElectron/Run2017F-31Mar2018-v1/MINIAOD /SingleMuon/Run2017B-31Mar2018-v1/MINIAOD /SingleMuon/Run2017C-31Mar2018-v1/MINIAOD /SingleMuon/Run2017D-31Mar2018-v1/MINIAOD /SingleMuon/Run2017E-31Mar2018-v1/MINIAOD /SingleMuon/Run2017F-31Mar2018-v1/MINIAOD	41.5
2018 dataset	
/EGamma/Run2018A-17Sep2018-v2/MINIAOD /EGamma/Run2018B-17Sep2018-v1/MINIAOD /EGamma/Run2018C-17Sep2018-v1/MINIAOD /EGamma/Run2018D-PromptReco-v2/MINIAOD /SingleMuon/Run2018A-17Sep2018-v2/MINIAOD /SingleMuon/Run2018B-17Sep2018-v1/MINIAOD /SingleMuon/Run2018C-17Sep2018-v1/MINIAOD /SingleMuon/Run2018D-PromptReco-v2/MINIAOD	58.9

Table 4.1: Datasets and corresponding luminosities used in the analysis.

Year	$\mu\mu$	ee
2016	HLT_Mu50_v* or HLT_TkMu50_v*	HLT_Ele115_CaloIdVT_GsfTrkIdT_v*
2017	HLT_Mu50_v* or HLT_OldMu100_v* or HLT_TkMu100_v*	HLT_Ele35_WPTight_Gsf_v* or HLT_Ele115_CaloIdVT GsfTrkIdT_v* or HLT_Photon200_v*
2018	HLT_Mu50_v* or HLT_OldMu100_v* or HLT_TkMu100_v*	HLT_Ele115_CaloIdVT GsfTrkIdT_v*

Table 4.2: Triggers used during each year of data taking.

transverse momenta and pseudorapidity. In this analysis, per-event prefiring weights w are calculated as:

$$w = 1 - P(\text{Prefiring}) = \prod_{i=\text{photons, jets}} (1 - \epsilon_i^{pref}(\eta, p_T)), \tag{4.3}$$

where ϵ_i^{pref} is the prefiring probability for a single photon or jet.

- **Pile-up reweighting**: MC distributions are reweighted according to the scaling of the number of vertex distribution between data and MC in the different years. In order to account for differences between the distributions of the number of pileup interactions in the simulations compared to the one in real data, the recipe in Ref. [114] is applied. For the MC mixing scenario we used:
  - mix_2016_25ns_Moriond17MC_PoissonOOTPU_cfi for 2016,
  - mix_2017_25ns_WinterMC_PUScenarioV1_PoissonOOTPU_cfi for 2017,
  - mix_2018_25ns_JuneProjectionFull18_PoissonOOTPU_cfi for 2018.

  The effect of this correction is shown in Fig. 4.7, where we show the number of vertex distribution for the MC and the data, together with the weight ratio for the three different years.

Signal MC samples	σ (pb)
ExtendedWeakIsospinModel_$\ell\ell$jj_L9000_M500_CalcHEP	2.6040e-01
ExtendedWeakIsospinModel_$\ell\ell$jj_L9000_M1000_CalcHEP	1.4550e-01
ExtendedWeakIsospinModel_$\ell\ell$jj_L9000_M2000_CalcHEP	4.7540e-03
ExtendedWeakIsospinModel_$\ell\ell$jj_L9000_M5000_CalcHEP	2.4000e-04
ExtendedWeakIsospinModel_$\ell\ell$jj_L9000_M6000_CalcHEP	1.4620e-05
ExtendedWeakIsospinModel_$\ell\ell$jj_L9000_M9000_CalcHEP	3.5980e-08
ExtendedWeakIsospinModel_$\ell\ell$jj_L11000_M500_CalcHEP	1.1330e-01
ExtendedWeakIsospinModel_$\ell\ell$jj_L11000_M1000_CalcHEP	5.8690e-03
ExtendedWeakIsospinModel_$\ell\ell$jj_L11000_M2000_CalcHEP	1.8900e-03
ExtendedWeakIsospinModel_$\ell\ell$jj_L11000_M4000_CalcHEP	1.0760e-04
ExtendedWeakIsospinModel_$\ell\ell$jj_L11000_M7000_CalcHEP	1.2520e-06
ExtendedWeakIsospinModel_$\ell\ell$jj_L13000_M500_CalcHEP	5.709e-03
ExtendedWeakIsospinModel_$\ell\ell$jj_L13000_M1000_CalcHEP	2.819e-03
ExtendedWeakIsospinModel_$\ell\ell$jj_L13000_M2000_CalcHEP	8.207e-04
ExtendedWeakIsospinModel_$\ell\ell$jj_L13000_M5000_CalcHEP	1.425e-05
ExtendedWeakIsospinModel_$\ell\ell$jj_L13000_M8000_CalcHEP	9.127e-08
ExtendedWeakIsospinModel_$\ell\ell$jj_L15000_M500_CalcHEP	3.1970e-03
ExtendedWeakIsospinModel_$\ell\ell$jj_L15000_M1000_CalcHEP	1.5230e-03
ExtendedWeakIsospinModel_$\ell\ell$jj_L15000_M2000_CalcHEP	4.1100e-04
ExtendedWeakIsospinModel_$\ell\ell$jj_L15000_M6000_CalcHEP	1.8950e-06
ExtendedWeakIsospinModel_$\ell\ell$jj_L17000_M500_CalcHEP	1.9310e-03
ExtendedWeakIsospinModel_$\ell\ell$jj_L17000_M1000_CalcHEP	8.9670e-04
ExtendedWeakIsospinModel_$\ell\ell$jj_L17000_M2000_CalcHEP	2.2790e-04
ExtendedWeakIsospinModel_$\ell\ell$jj_L17000_M7000_CalcHEP	2.1960e-07
ExtendedWeakIsospinModel_$\ell\ell$jj_L20000_M500_CalcHEP	1.0060e-03
ExtendedWeakIsospinModel_$\ell\ell$jj_L20000_M1000_CalcHEP	4.5480e-04
ExtendedWeakIsospinModel_$\ell\ell$jj_L20000_M2000_CalcHEP	1.0820e-04
ExtendedWeakIsospinModel_$\ell\ell$jj_L20000_M8000_CalcHEP	1.4910e-08

Table 4.3: Value of the cross-sections of the simulated MC samples for the signal used in the analysis for a compositeness scale values above the saturation of the unitarity bound (at about $\Lambda = 7.6$ TeV [110]) TeV for both the channels ($\ell = e, \mu$).

Background MC samples	σ (pb)	k-factor
2016 dataset		
TT_TuneCUETP8M2T4_13TeV-powheg-pythia8	730.6 (NNLO)	
ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	38.09 (NLO)	0.935
ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	38.06 (NLO)	0.935
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6225.42 (NNLO)	
WJetsToLNu_HT-70To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1352 (LO)	1.21
WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1346 (LO)	1.21
WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	360.1 (LO)	1.21
WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	48.4 (LO)	1.21
WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	12.07 (LO)	1.21
WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5.497 (LO)	1.21
WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.329 (LO)	1.21
WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.03209 (LO)	1.21
WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	50260 (LO)	1.21
WW_TuneCUETP8M1_13TeV-pythia8	118.7 (NLO)	
WZ_TuneCUETP8M1_13TeV-pythia8	47.13 (NLO)	
ZZ_TuneCUETP8M1_13TeV-pythia8	16.523 (NLO)	
2017 and 2018 dataset		
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	88.29 (NNLO)	
ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	34.91 (NLO)	0.935
ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	34.97 (NLO)	0.935
DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8	6225.42 (NNLO)	
WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	1346 (LO)	1.21
WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	407.9 (LO)	1.21
WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	57.48 (LO)	1.21
WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	12.87 (LO)	1.21
WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	5.366 (LO)	1.21
WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	1.074 (LO)	1.21
WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.008001 (LO)	1.21
WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8	52940 (LO)	1.21
WW_TuneCP5_13TeV-pythia8	118.7 (NLO)	
WZ_TuneCP5_13TeV-pythia8	47.13 (NLO)	
ZZ_TuneCP5_13TeV-pythia8	16.523 (NLO)	

Table 4.4: Background MC samples and their cross-sections used in the analysis at the accuracy level of the simulated samples, and the k-factor used for some of them.

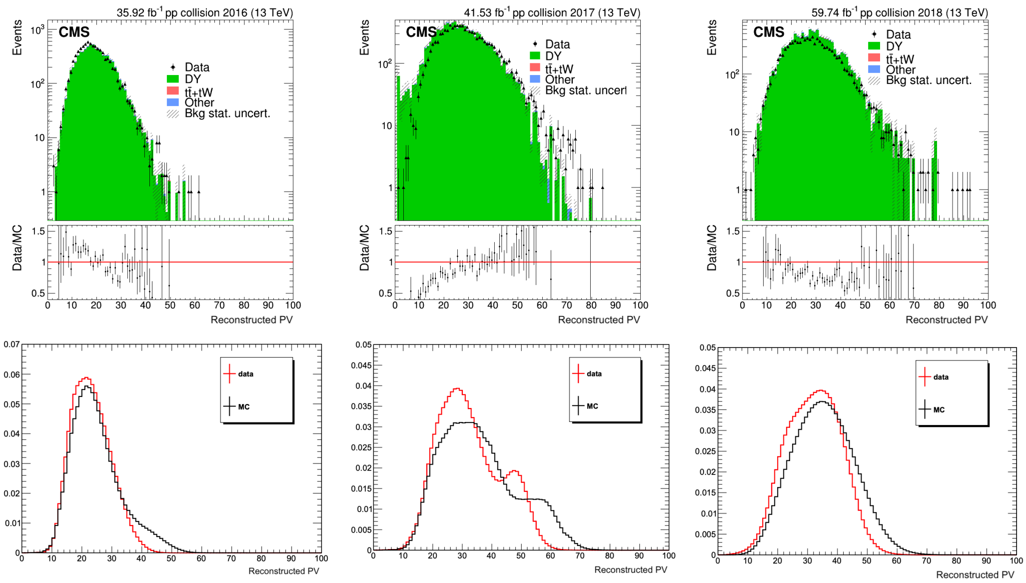


Figure 4.7: Number of reconstructed vertex distribution for data and MC (top row) and weight factors from the pileup reweighting procedure (bottom row). From left to right: 2016, 2017 and 2018.

3 OBJECT RECONSTRUCTION AND IDENTIFICATION

The reconstruction and identification of the objects employed in the analysis are described in the following. Dedicated groups within the CMS collaboration, the POG ("Physics Object Group"), develop different algorithms to get the lowest possible misidentification rate and highest efficiency for a considered reconstructed object in the detector. Specifically, this section contains the description of the objects used in the search, i.e. electrons, muons, and jets, and the main features of their selection for the signal and control regions, as well as the scale factors applied.

3.1 ELECTRONS

The electrons are taken from the collection "slimmedElectrons" [115] and are identified according to the "HEEP V7.0" selection [116], which has been tailored for high-momenta electrons like the ones we expect in our analysis. The electron candidates are required to be within $|\eta|$ <2.4 and with a minimum p_T of 150 (100) GeV for the leading (subleading) electron. We use the data/MC scale factors according to the measurement reported in [117]. In the $\mu\mu qq$ channel we use the "Veto" working point described in [118] to veto the presence of electrons with p_T >20 GeV and $|\eta|$ <2.4.

3.2 MUONS

The muons are taken from the collection "slimmedMuons" and are identified according to the "High-pT" working point and the "Loose" tracker isolation selection recommended by the Muon POG [119]. The muon candidates are required to be within $|\eta| <2.4$ and with a minimum p_T of 150 (100) GeV for the leading (subleading) muon. We use the data/MC scale factors according to the indications reported in [120]. In the *eeqq* channel we rely on the "Loose" working point and Loose tracker isolation described in [119] to veto the presence of muons with $p_T >20$ GeV and $|\eta| <2.4$. To correct biases in muon momentum reconstruction due to a mismodeling of detector alignment and magnetic field we use the Rochester corrections [121] for muons with $p_T < 200$ GeV. For higher momenta muons, the increase of shower probability in the muon system and the extremely small curvature could lead to under-performance in momentum scale and resolution with respect to the performance at low-p_T. Therefore, muons with $p_T \geq 200$ GeV are corrected with the so-called generalized-endpoint (GE) method [120], which corrects biases in the momentum scale and brings the position and width of the $Z \to \mu\mu$ mass peak in simulations and data into better agreement. Moreover, an extra Gaussian smearing is applied to MC to match resolution estimated from data [120].

3.3 JETS

The jets used in this analysis are reconstructed by clustering particle flow candidates in the events using the anti-kT algorithm with a distance parameter $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.8$ (AK8). The AK8 jets, taken from the "slimmedJetsAK8" collection, are identified according to the "Tight" working point recommended in [122–124]. We further perform a cleaning against electrons and muons requesting $\Delta R > 0.8$. Jets should also have $p_T >190$ GeV and $|\eta| <2.4$. The jet candidates are corrected according to the most updated jet energy corrections (JEC) reported in [125] and jet energy resolutions (JER) reported in [126].

4 EVENT SELECTION AND OPTIMIZATION

The signal mediated by a heavy Majorana neutrino requires the presence of two same-flavor leptons and two quarks. Despite the signal being prompted by a self-conjugate resonance, we work within a charge-blind signature hypothesis (for the motivation of this choice see 6.12). Considering the different nature of the two interactions in the decay of the heavy neutrino, we require one or more AK8 jets, which guarantees high signal efficiency for events with two leptons and is suitable for N_ℓ decays through both the gauge and the contact interactions, as explained in Sec. 4.4.1. Subsequently, selection is applied to restrict the invariant mass of the two leptons $M(\ell, \ell)$ GeV to be in a high-mass region, where the background contribution is lower. The constraints on the parameter space given by the unitarity conditions on the Lagrangian of the model presented in Sec. 4.1.1, we said, suggest reconsidering the optimization of the signal region. The target is not only to reach higher Heavy Composite Neutrino (HCN) masses but also to extend the sensitivity towards higher values of Λ (and small neutrino masses), that region being the safest to avoid unphysical predictions from the EFT interactions of the model. Therefore, in the last part of this section, we will report the details of this optimization procedure.

4.1 SIGNAL TOPOLOGY

In this section, we analyze the peculiar decay topology of the HCN and comment on the reason why our selection of the number and radius of the reconstructed jets is optimal for this particular signal.

Figure 4.8 shows the relative importance of the decay amplitudes of the two competing mechanisms in the phase space of the analysis. For a portion of the HCN parameter space, the gauge interactions rule the decay process $N \to \ell q q'$, and the two quarks coming from the W decay tend to overlap, producing one fat jet in the final state. For the parameter space in which, instead, the decay mechanism is dominated by contact interactions, the two quarks, and the lepton are produced without being constrained to a particular direction in the space. This interplay is evident in the generator level distribution of the ΔR between the two quarks in the signal for different mass scenarios (and fixed compositeness scale) in Fig. 4.9: the sharp peak at small ΔR represents the gauge-dominating decays, whether the peak around the $\Delta R = 3$ region can be thought as prompted by contact interactions in decay. Nonetheless, requiring at least one fat jet in the event grants a high acceptance to the quarks in the signal even for the contact interaction decays, as shown in the study of the matching probability reported in Table 4.5.

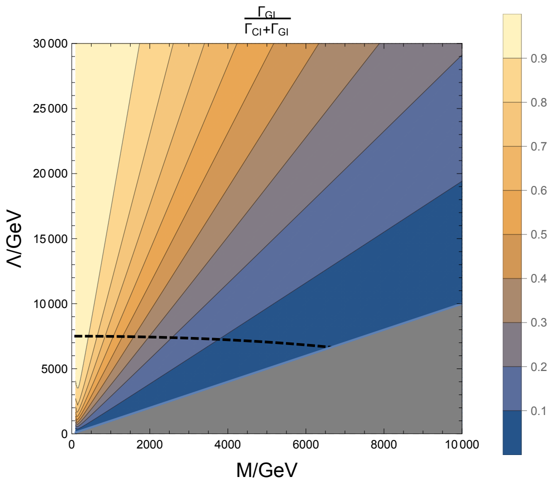


Figure 4.8: Contour plot of the fraction of events decaying with gauge interactions, $\Gamma_{GI}/(\Gamma_{GI}+\Gamma_{CI})$. The gray shaded area represent the region forbidden by the model $M \leq \Lambda$, and the dashed black line remarks the saturation of the unitarity bound in the production mechanism (Section 2.2).

M(N) [GeV]	# FatJet	# events	matched q1	matched q2	matched q1 and q2	matched none
500	1	38114	16 %	1%	75%	8 %
	≥ 2	3366	8 %	2%	22%	68%
5000	1	52120	70 %	7 %	21%	2%
	≥ 2	3908	7 %	68 %	0%	25%

Table 4.5: Number of AK8 jets in the signal samples and percentages of matched signal quarks for two different mass points: 500 GeV (gauge dominated) and 5000 GeV (contact dominated) (and same $\Lambda = 13$ TeV).

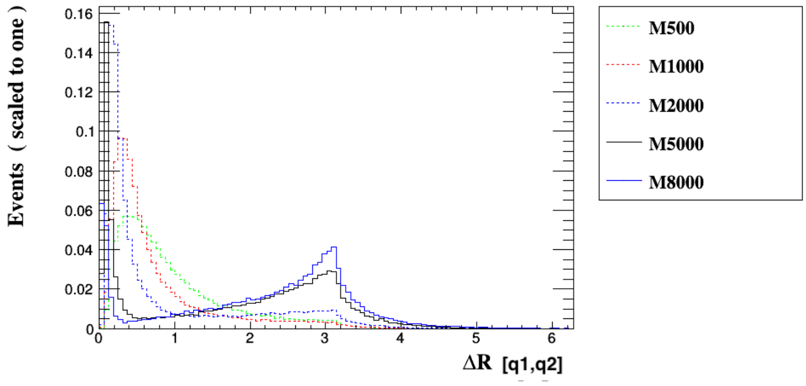


Figure 4.9: The ΔR distribution between the two quarks for different mass hypotheses of the signal with $\Lambda = 13$ TeV (plot obtained with MadAnalysis). The gauge and contact contributions can be easily distinguished.

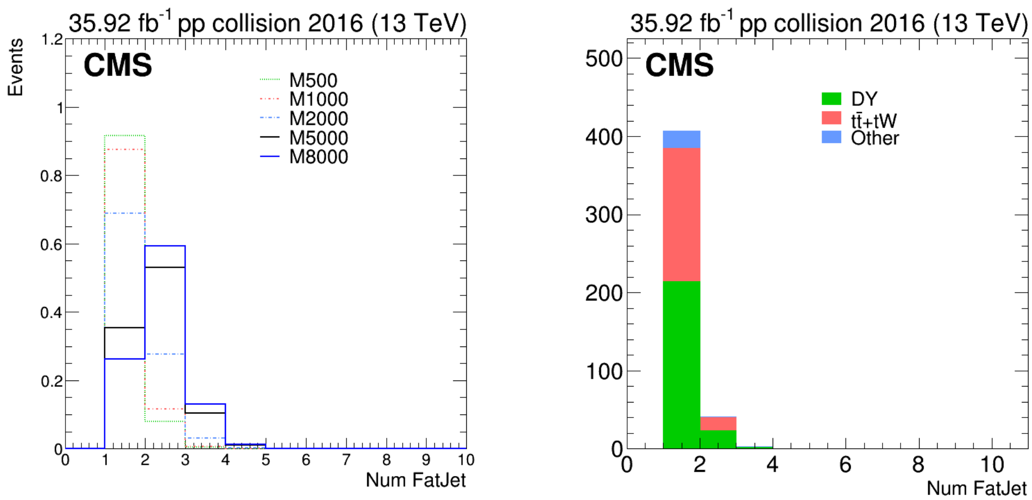


Figure 4.10: The distributions for the number of AK8 jets in the event for the background stacked (right), and the signal samples (left) with different masses and $\Lambda = 13$ TeV, normalized to one.

4.2 THE SIGNAL REGION OPTIMIZATION

Here we comment on the choice of the cuts of the finale-state objects. We report in Figure 4.11 the shapes of the main kinematical variables of events firing the triggers chosen for the signal samples with masses 500 GeV and 1 TeV (and compositeness scale 13 TeV), together with the MC predictions of the backgrounds. The optimization process can be summarized as the result of two main requirements:

- Choosing the p_T threshold for the leptons selection looking at the plateau in the trigger efficiency. As reported in Fig. 4.12 and Fig. 4.13, the expected overall trigger efficiency for the di-lepton events in both channels, considering a signal sample of Heavy Composite Majorana Neutrino with $\Lambda = 13$ TeV and the background processes, is between 90 and 100 % for $p_T(\ell) > 100$ GeV. We only take a single Λ point as a reference since the overall efficiency of the heavy composite neutrino sample does not depend on the compositeness scale, as already demonstrated in the previous version of this analysis (and in the HL/HE-LHC projection studies [80]). The same approximation holds for the statistical analysis in Section 4.8, where we compute upper limits on the cross-section of the reference Λ signal samples as a function of the HCN mass.

- Optimization of the figure of merit $\sqrt{(2((S+B)ln(1+S/B)-S))}$ for smaller mass samples in the region favored by the unitarity constraint. Results are shown in Tab. 4.6 and 4.7 for a few benchmark points (taking 2018 as reference year).

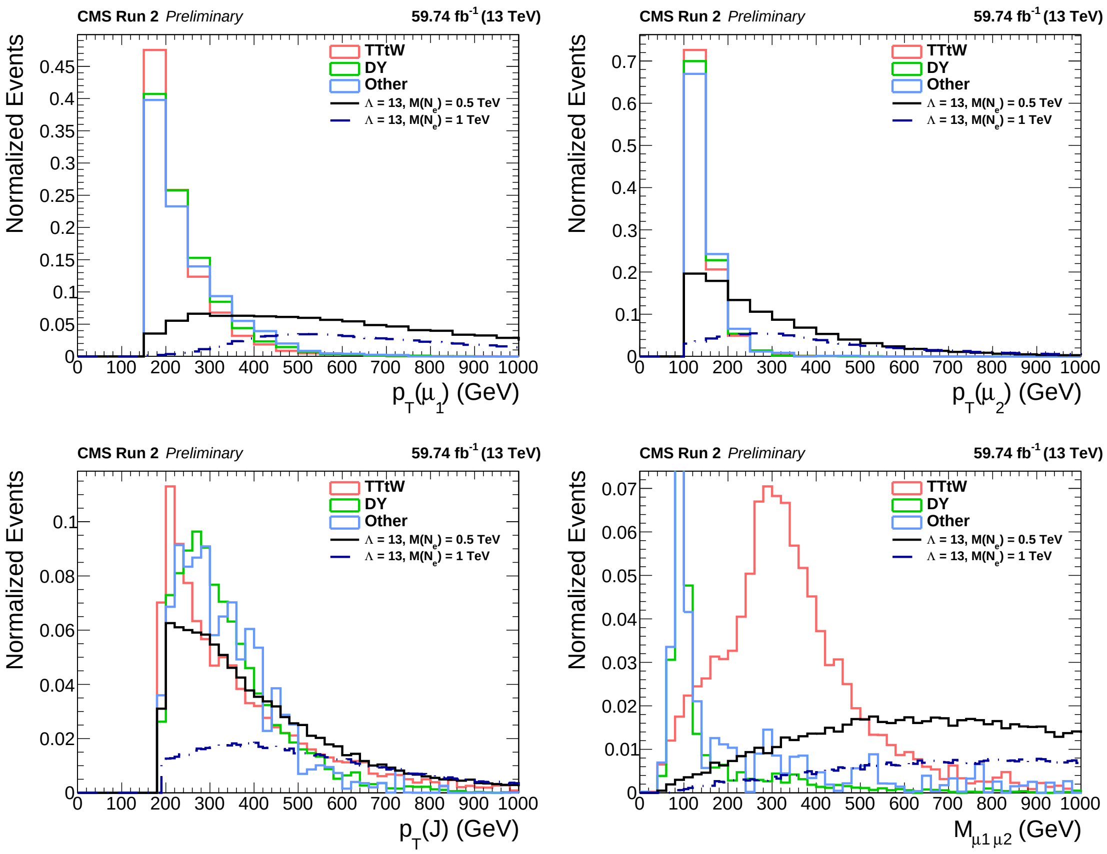


Figure 4.11: Kinematical variables of the main reconstructed objetcts in the analysis for the muonic channel at the preselection level for the simulated backgrounds, and for two signal hypotheses with masses 0.5 and 1 TeV (and same Λ=13 TeV). The histograms are normalized to unitary area, and MC samples are chosen from 2018 production.

Request	$\sqrt{(S+B)} = \sqrt{(2((S+B)ln(1+S/B)-S))}$				
	L13 M500	L13 M1000	L13 M2000	L13 M5000	L13 M8000
lead lepton $p_T > 150$ GeV	0.55	0.40	0.13	0.0022	0.000014
sublead lepton $p_T > 100$ GeV	0.91	0.73	0.24	0.0042	0.000026
$m_{ll} > 300$ GeV	3.54	2.94	0.98	0.0177	0.000111
fat jet $p_T > 190$ GeV	3.54	2.94	0.98	0.0177	0.000111

Table 4.6: Figure of merit $S/\sqrt{(S+B)}$ for various signal reference points for the muon channel.

Request	$\sqrt{(S+B)} = \sqrt{(2((S+B)ln(1+S/B)-S))}$				
	L13 M500	L13 M1000	L13 M2000	L13 M5000	L13 M8000
lead lepton $p_T > 150$ GeV	0.53	0.39	0.12	0.0021	0.000013
sublead lepton $p_T > 100$ GeV	0.85	0.68	0.22	0.0038	0.000024
$m_{ll} > 300$ GeV	3.49	2.90	0.95	0.0160	0.000106
fat jet $p_T > 190$ GeV	3.49	2.90	0.95	0.0169	0.000106

Table 4.7: Figure of merit $S/\sqrt{(S+B)}$ for various signal reference points for the electron channel.

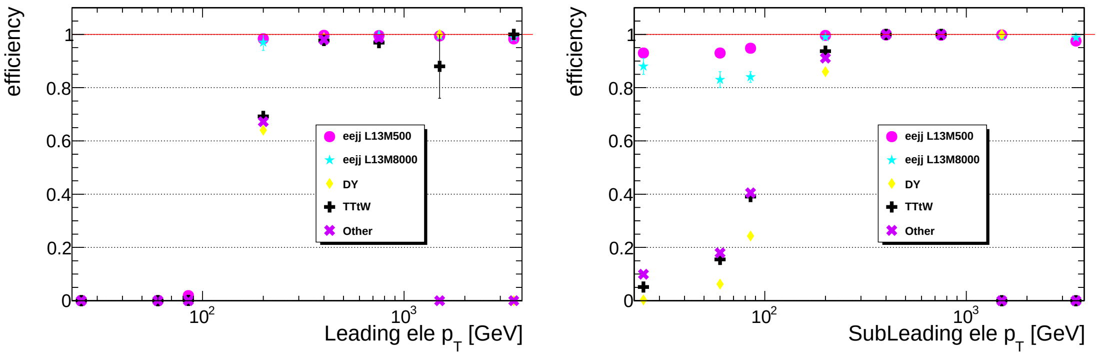


Figure 4.12: The trigger efficiency for the electron channel for the leading (left) and the sub-leading lepton (right) reported for all the backgrounds and a two points of the signal MC samples of Heavy Composite Majorana Neutrino with mass of 0.5 and 8 TeV.

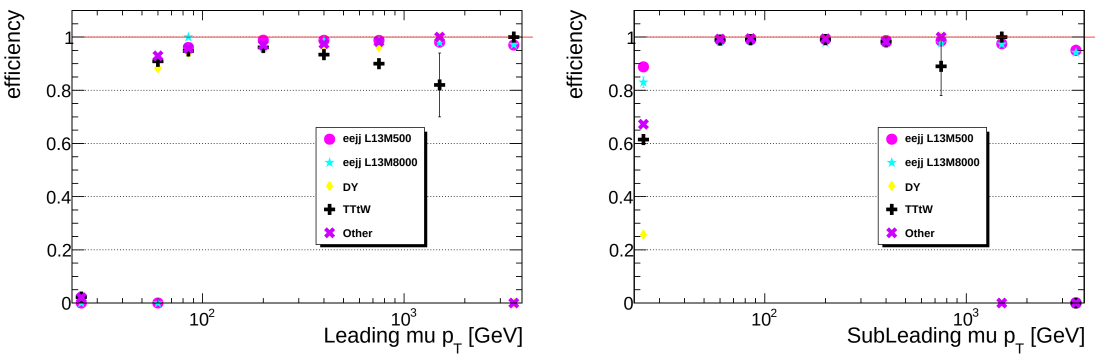


Figure 4.13: The trigger efficiency for the muon channel for the leading (left) and the sub-leading lepton (right) reported for all the backgrounds and a two points of the signal MC samples of Heavy Composite Majorana Neutrino with mass of 0.5 and 8 TeV.

The final selections of the signal region for the $eeqq$ and $\mu\mu qq$ channels are summarized in Table 4.8. The effect of the selection cuts on the reference signal samples is shown in Fig. 4.14 with the cumulative efficiency for both channels. The distribution used in the statistical analysis is the invariant mass of the leptons and the leading fat jet, $M(\ell\ell J)$. Besides providing a good discrimination between the signal and the SM background contributions (Fig. 4.15), this variable is strongly correlated with the mass of the HCN, especially when the signal quarks are matched in the leading AK8 jet [104]. If we write the invariant mass of the final state particles $\ell\ell qq$ as a convolution of parton density functions:

$$\frac{d\sigma}{dM^2_{\ell\ell jj}} = \sum_{ab} \int^1_{M^2_{\ell jj}} \frac{dx}{x} f_a\left(x, q^2\right) f_b\left(\frac{M^2_{\ell\ell jj}}{sx}, q^2\right) \times \int_0^{M^2_{\ell\ell j}} \frac{d\mathcal{Q}^2}{\pi} \frac{\sqrt{\mathcal{Q}^2}\hat{\sigma}_{q_a q_b \to \ell N^*}\left(M_{\ell\ell jj}, \mathcal{Q}\right)\Gamma_{N^* \to \ell jj}(\mathcal{Q})}{\left(\mathcal{Q}^2 - M_N^2\right)^2 + \left(M_N \Gamma_{\text{tot}}(\mathcal{Q})\right)^2}, \tag{4.4}$$

where q is the QCD factorization scale and $\mathcal{Q}$ is the virtual momentum of the resonantly produced heavy neutrino N, we can distinguish two factors. One such factor, the first on the right-hand side of Eq. 4.4, is a (dimensionless) parton distribution luminosity factor that vanishes at very large invariant masses $M(\ell\ell qq) \sim \sqrt{s}$, while the second factor in the right hand side of Eq. 4.4, is an integral over the virtuality of the produced neutrino, $\mathcal{Q}$, and vanishes for small values of the invariant mass or $M(\ell\ell qq) \ll M(N)$. Therefore, in general, we expect an invariant mass distribution characterized by a peak for $M(\ell\ell qq) \geq M(N)$. Such a picture is of course not altered by the relative importance that contact and gauge interactions may have in the decay process. Figure 4.15 presents this invariant mass distribution for two HCN masses, and the expected SM backgrounds from MC simulations. The expected dominant background sources are given by the DY and the "TTtW" processes, the latter containing both the $t\bar{t}$ and tW. The other backgrounds from multiboson processes ($WJets$, WW, WZ, ZZ), are expected to contribute less and will be referred to altogether as "Other".

Event Selection	eeJ	$\mu\mu J$
lepton 1	$p_T(e_1) > 150$ GeV and $\|\eta\| < 2.4$	$p_T(\mu_1) > 150$ GeV, and $\|\eta\| < 2.4$
lepton 2	$p_T(e_2) > 100$ GeV and $\|\eta\| < 2.4$	$p_T(\mu_2) > 100$ GeV and $\|\eta\| < 2.4$
M(l1,l2)	$M(e_1, e_2) > 300$ GeV	$M(\mu_1, \mu_2) > 300$ GeV
Veto extra lepton	$p_T > 20$ GeV and $\|\eta\| < 2.4$	$p_T > 20$ GeV and $\|\eta\| < 2.4$
N(jetAK8)>0	$p_T > 190$ GeV and $\|\eta\| < 2.4$	$p_T > 190$ GeV and $\|\eta\| < 2.4$

Table 4.8: Summary of signal event selection requirements

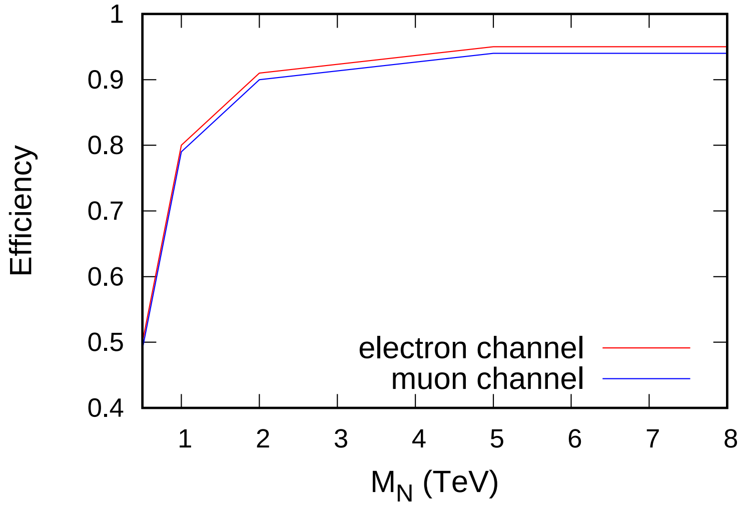


Figure 4.14: The cumulative efficiency in the signal region for the electron channel and the muon channel, as a function of the HCN mass, for the benchmark compositeness scale value $\Lambda = 13$ TeV. The result is consistent for the three years.

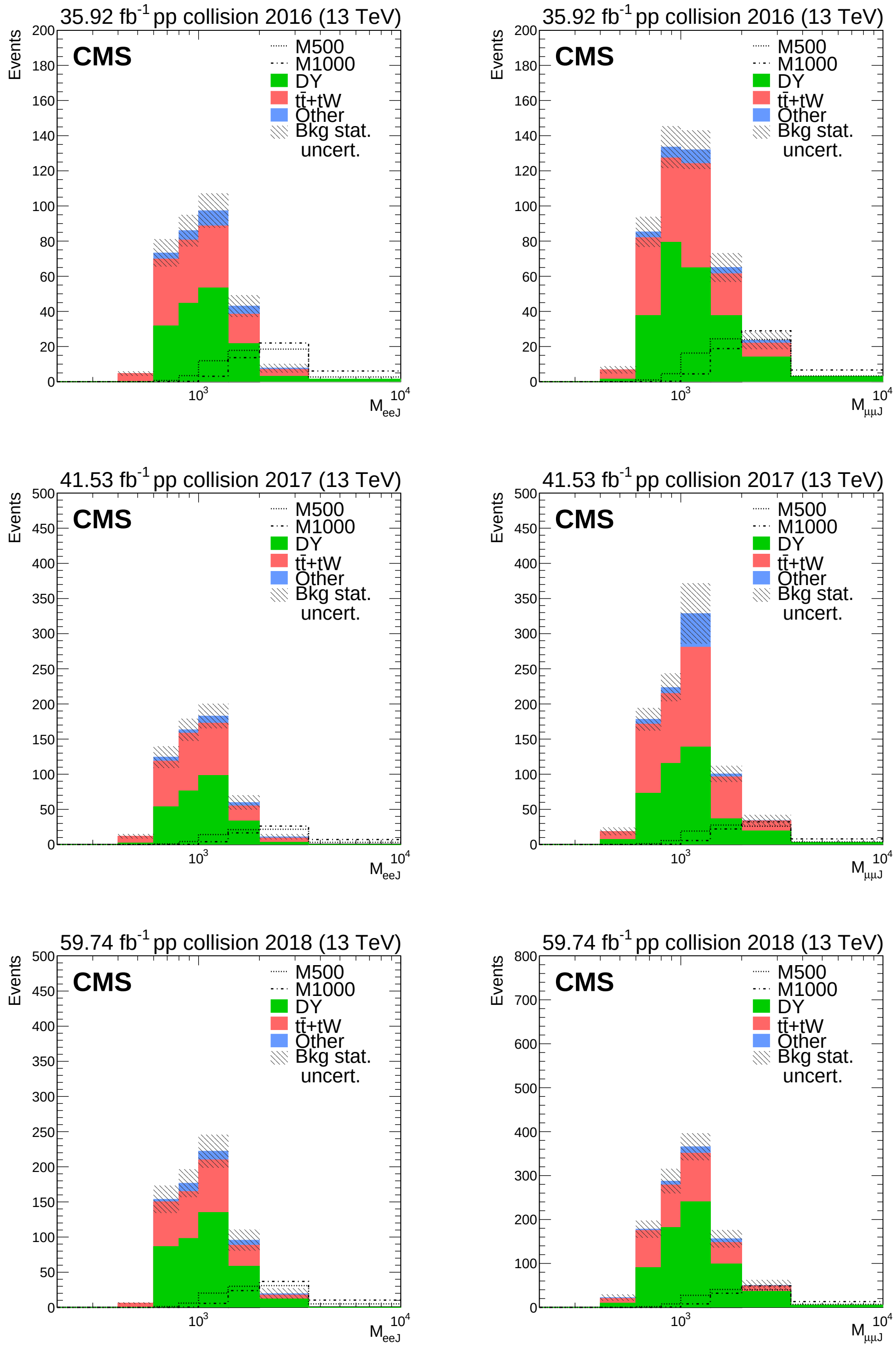


Figure 4.15: Distribution of the variable $M(\ell, \ell, J)$ of backgrounds (stack plots) as estimated in Sec. 4.5 and signal expected from simulation (lines) in the signal region, having considered the signal parameters $\Lambda = 13$ TeV and $M(N) = 500$ GeV and $M(N) = 1000$ GeV, for the *eeqq* (left) and the $\mu\mu qq$ (right) channels. The three years are reported separately, normalized to the corresponding luminosity, and applying the correction described in Sec. 4.3: 2016 on the top, 2017 on the middle, and 2018 on the bottom.

5 BACKGROUND ESTIMATION

In this section, we describe the methods used for the estimate of the most relevant standard model backgrounds (Drell-Yan, $t\bar{t}$, and QCD) for the search. For the DY and $t\bar{t}$, we define two enriched regions, whose definition is illustrated in Fig. 5.13.

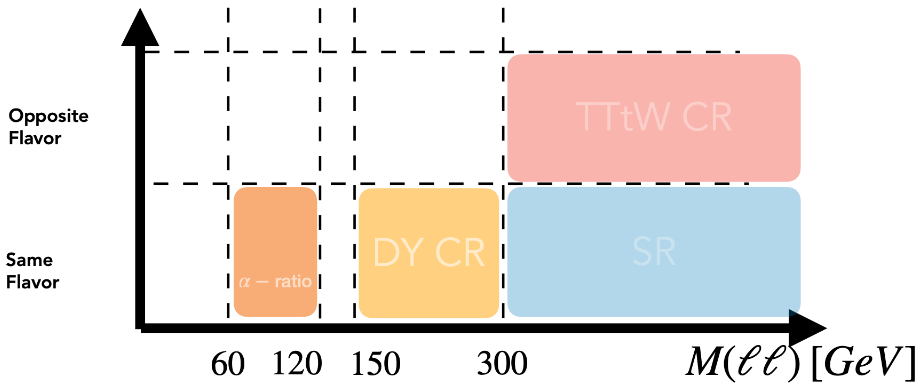


Figure 4.16: Pictorial representation of the phase space of the HCN search, for the signal and control regions.

All the remaining minor background sources, called "Other", are directly taken from MC simulations.

5.1 ESTIMATION OF *DY* BACKGROUND

The most significant background is due to the irreducible *DY* process, and is estimated with a semi data-driven technique, the so-called "α–ratio method" [94]. The same method is applied for the three years and to both electron and muon final states. Before discussing the extrapolation of the DY to the signal region, we comment on the k-factors applied on the DY MC sample used. For the signal and control regions, a large statistical power, as well as precise predictions are necessary, especially in the high-mass region that we select. Therefore, the LO simulation samples for the DY backgrounds are reweighted using higher-order corrections separately corresponding to NLO QCD and NLO EW terms. The theory-based corrections of the DY MC samples are usually parameterized as a function of the generator-level p_T of the respective Z boson. The correction factors for the NLO EW are obtained from Ref. [127]. Scale factors corresponding to NLO QCD corrections to the Z production are obtained from central CMS samples produced in the “Fall17” campaign, in which large LO and NLO samples are available. Both samples are generated using MadGraph5_amc@NLO. The LO samples are identical to the ones used in the analysis and are generated with up to four partons in the matrix element. The NLO samples are generated with up to two additional partons in the matrix element calculation at NLO accuracy, and the third parton at LO accuracy. Further jet multiplicities are handled by the parton shower, which in both cases is performed using Pythia8 with tune CP5. The scale factors are derived by obtaining the generator-level boson p_T distribution in both samples, normalizing the distributions to their respective cross-sections, and then dividing them as SF = NLO / LO. The SFs are shown in Fig.4.17 for both QCD NLO (left) and EW NLO (right). Figures 4.18, 4.19 and 4.20, 4.21 show the effect of these corrections to the data-MC ratios, for the phase space of the present analysis for both the channels. We compare the DY LO samples corrected with EW and QCD NLO SF, to the DY NLO samples available for 2016 and 2017 (the 2018 is not included here for the low statistics) in Figs. 4.22 and 4.23.

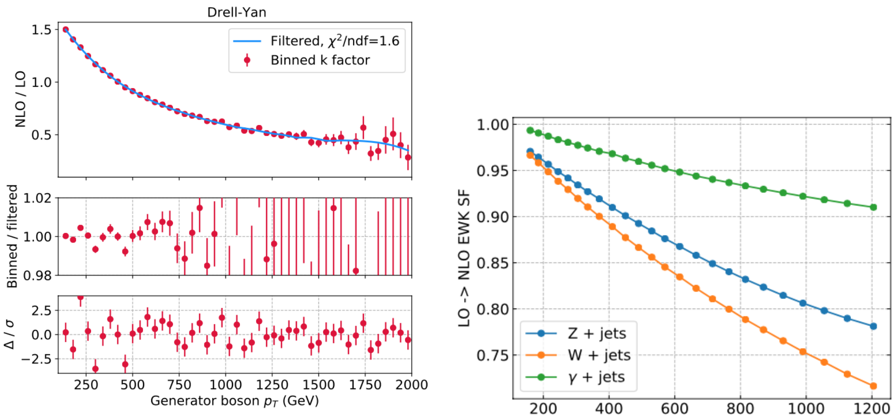


Figure 4.17: Left: QCD NLO correction functional form as a function of the Z boson generator level momentum. Right: EW NLO scale factors for DY, W and photon production as a function of the generator level momentum $p_{T,V}$. The derivation of these scale factors is documented in Ref. [128] for the same MC samples adopted in this analysis.

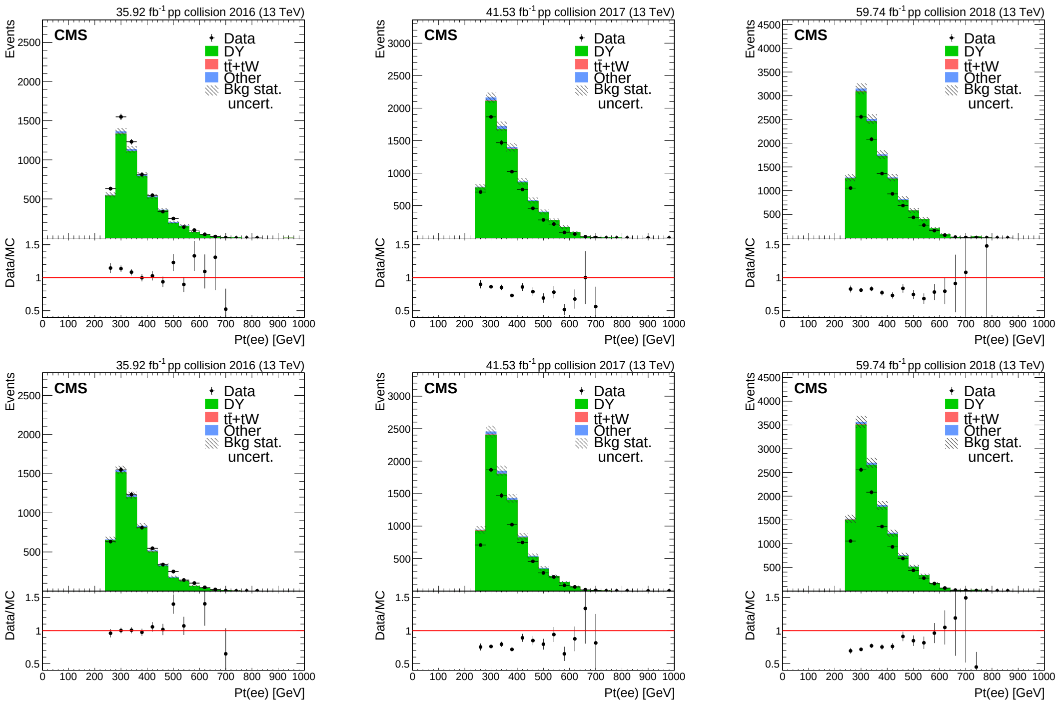


Figure 4.18: From left to right: distributions for Data and MC of the $p_T(ee)$ variable in the Z peak, with $60 < M(\ell\ell) < 120$ GeV, for 2016 (left), 2017 (center) and 2018 (right) data taking. On the first line, we show the MC distribution without LO to NLO corrections. On the second line, the same distributions, with the EW and QCD NLO weights applied.

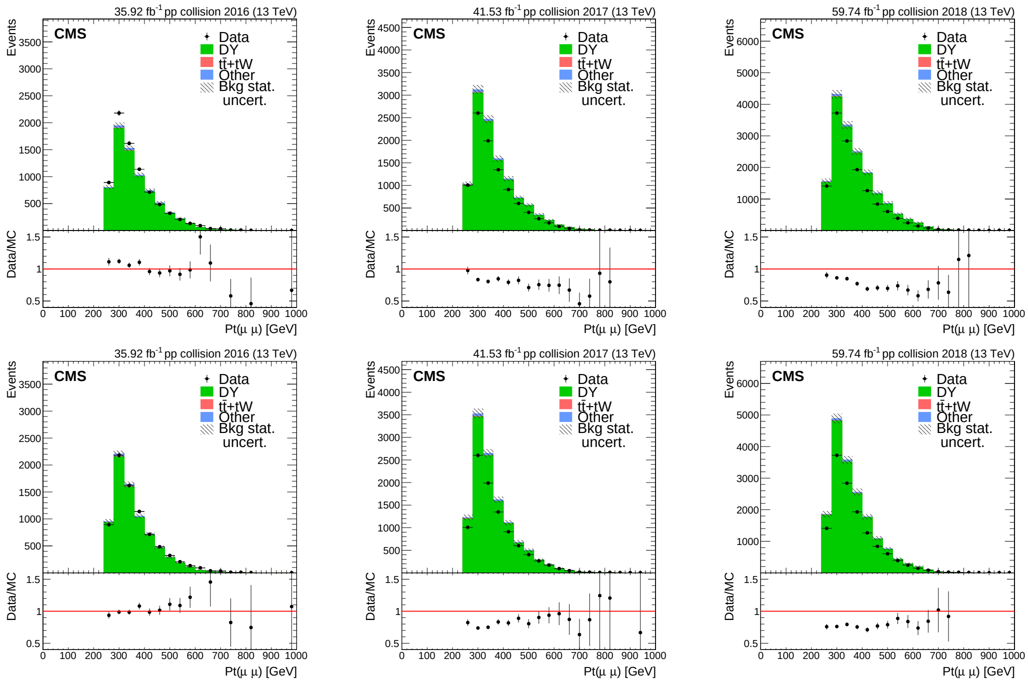


Figure 4.19: From left to right: distributions for Data and MC of the $p_T(\mu\mu)$ variable in the Z peak, with $60 < M(\ell\ell) < 120$ GeV, for 2016 (left), 2017 (center) and 2018 (right) data taking. On the first line, we show the MC distribution without LO to NLO corrections. On the second line, the same distributions, with the EW and QCD NLO weights applied.

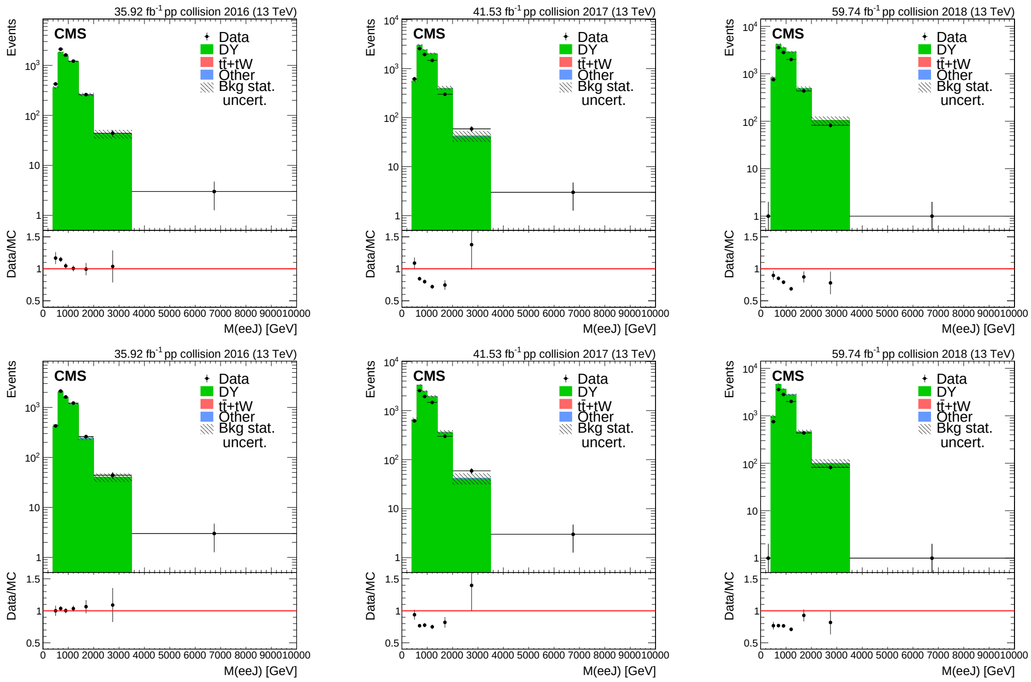


Figure 4.20: From left to right: distributions for Data and MC of the $M(eeJ)$ variable in the Z peak, with $60 < M(\ell\ell) < 120$ GeV, for 2016 (left), 2017 (center) and 2018 (right) data taking. On the first line, we show the MC distribution without LO to NLO corrections. On the second line, the same distributions, with the EW and QCD NLO weights applied.

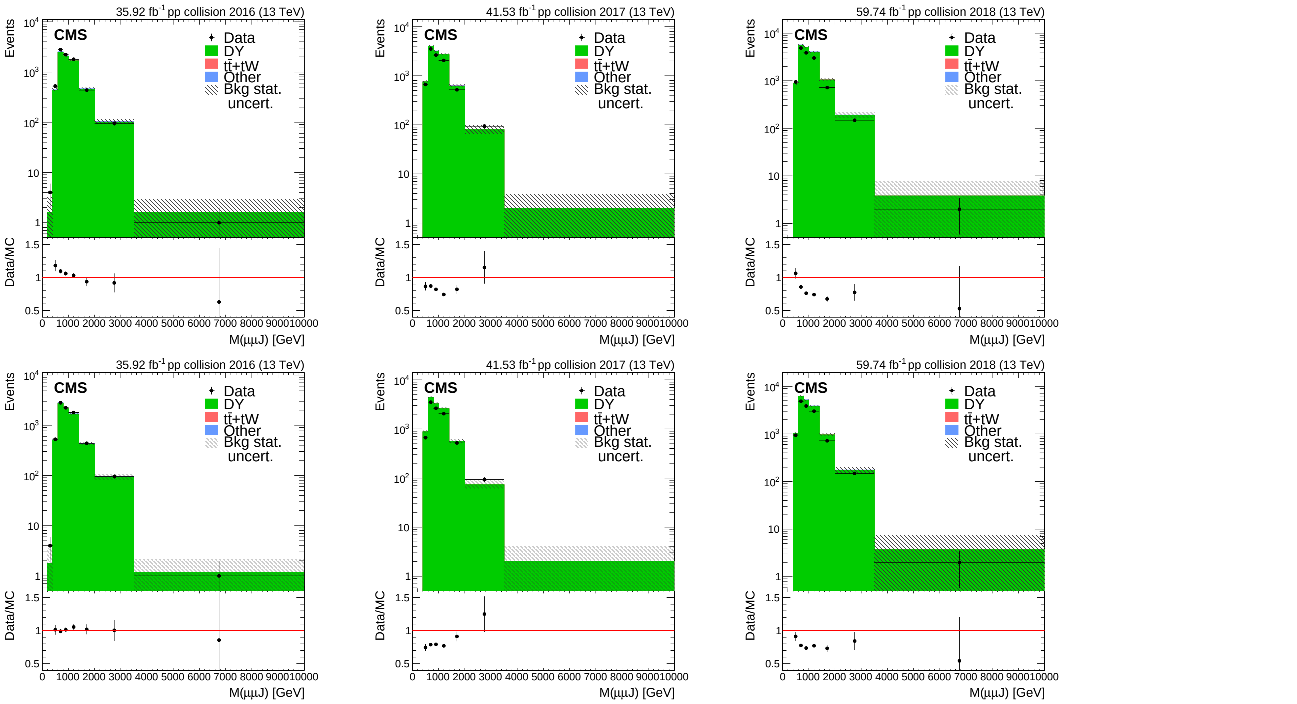


Figure 4.21: From left to right: distributions for Data and MC of the $M(\mu\mu J)$ variable in the Z peak, with $60 < M(\ell\ell) < 120$ GeV, for 2016 (left), 2017 (center) and 2018 (right) data taking. On the first line, we show the MC distribution without LO to NLO corrections. On the second line, the same distributions, with the EW and QCD NLO weights applied.

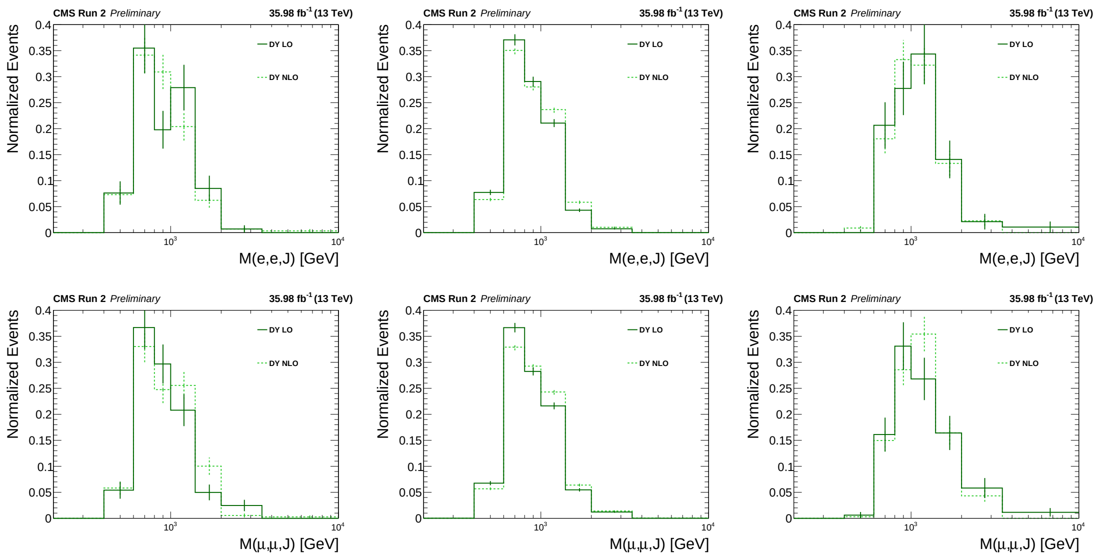


Figure 4.22: Comparison between DY NLO samples (/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3_ext2-v1/MINIAODSIM) and the DY LO samples used in this analysis for the MC 2016 samples for electron (first line) and muon channel (second line). On the left, distribution for $M(\ell\ell J)$ in the DY control region (with $150 < M(\ell\ell) < 300$ GeV). On the center, distribution for $M(\ell\ell J)$ around the Z peak (with $60 < M(\ell\ell) < 120$ GeV). On the right, for the signal region.

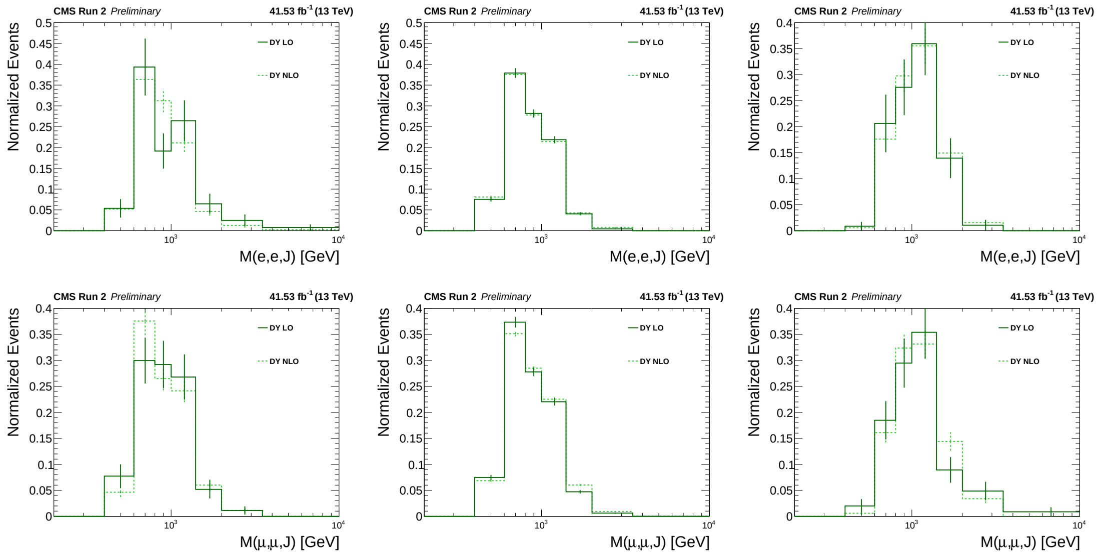


Figure 4.23: Comparison between DY NLO samples (/DYJetsToLL_M-50_ TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM) and the DY LO samples used in this analysis for the MC 2017 samples for electron (first line) and muon channel (second line). On the left, distribution for $M(\ell\ell J)$ in the DY control region (with $150 < M(\ell\ell) < 300$ GeV). On the center, distribution for $M(\ell\ell J)$ around the Z peak (with $60 < M(\ell\ell) < 120$ GeV). On the right, for the signal region.

5.1.1 *The extraction of the alpha-ratio from the Z region*

The DY contribution to the signal region (N_{DY}^{SR}) is estimated by normalizing the MC simulation (N_{MC}^{SR}) to the data, in a region defined with the same cuts used for the signal region, but around the Z peak, i.e. $60 < M_{\ell\ell} < 120$ GeV.
A scale factor (SF_{DY}) is measured in this region as the ratio of the events given by (data-nonDY)/DY, where "nonDY" and DY indicate the "Other" plus "TTtW" and the DY expectations taken from MC simulation, respectively. The formula used to evaluate this background is thus:

$$N_{DY}^{SR} = N_{MC}^{SR} \times SF_{DY}$$

Fig. 4.24 shows the results of the comparison Data/MC for the variable $M(\ell\ell)$ for the three years. However, since this analysis employs the variable $M(\ell\ell J)$ as a discriminant, we will extract the alpha-ratio from the distributions in $M(\ell\ell J)$,(and $60 < M_{\ell\ell} < 120$ GeV). The alpha ratio will be extracted from the plots shown in the bottom row of Figs. 4.20 and 4.21 (considering the DY samples with the LO to NLO EWK and QCD corrections). The alpha ratio extracted is then binned in $M(\ell\ell J)$ (instead of being a single number), see Table.4.9.

5.1.2 *Definition of the CR for DY background estimate and systematics*

This analysis will probe a kinematic region much higher than the one used to extract the alpha-ratio. In order to better understand any residual dependence on the kinematics of the event, we compare the data distribution both in $M(\ell\ell)$ and the

year	channel	α-ratio $\pm$ (stat.)
2016	ee	{-,-,1.00 ± 0.08, 1.03 ± 0.04, 1.00 ± 0.04, 1.03 ± 0.05, 1.07 ± 0.11, 1.09 ± 0.21, - }
	$\mu\mu$	{-,-,1.01 ± 0.08, 0.99 ± 0.04, 1.01 ± 0.04, 1.06 ± 0.04, 1.02 ± 0.08, 1.00 ± 0.16, 0.86 ± 1.10 }
2017	ee	{-,-,0.94 ± 0.08, 0.76 ± 0.03, 0.77 ± 0.03, 0.75 ± 0.04, 0.82 ± 0.09, 1.43 ± 0.43, - }
	$\mu\mu$	{-,-,0.74 ± 0.06, 0.79 ± 0.03, 0.79 ± 0.03, 0.76 ± 0.03, 0.91 ± 0.08, 1.26 ± 0.28, - }
2018	ee	{-,-,0.77 ± 0.06, 0.76 ± 0.03, 0.78 ± 0.03, 0.71 ± 0.03, 0.93 ± 0.10, 0.82 ± 0.20, - }
	$\mu\mu$	{-,-,0.91 ± 0.07, 0.77 ± 0.02, 0.77 ± 0.02, 0.77 ± 0.03, 0.71 ± 0.05, 0.84 ± 0.14, 0.54 ± 0.66 }

Table 4.9: Summary of the results for the α-ratio method for the three years and two channels. $M(\ell\ell J)$ variable bin per bin scale factors are reported with their statistical uncertainty.

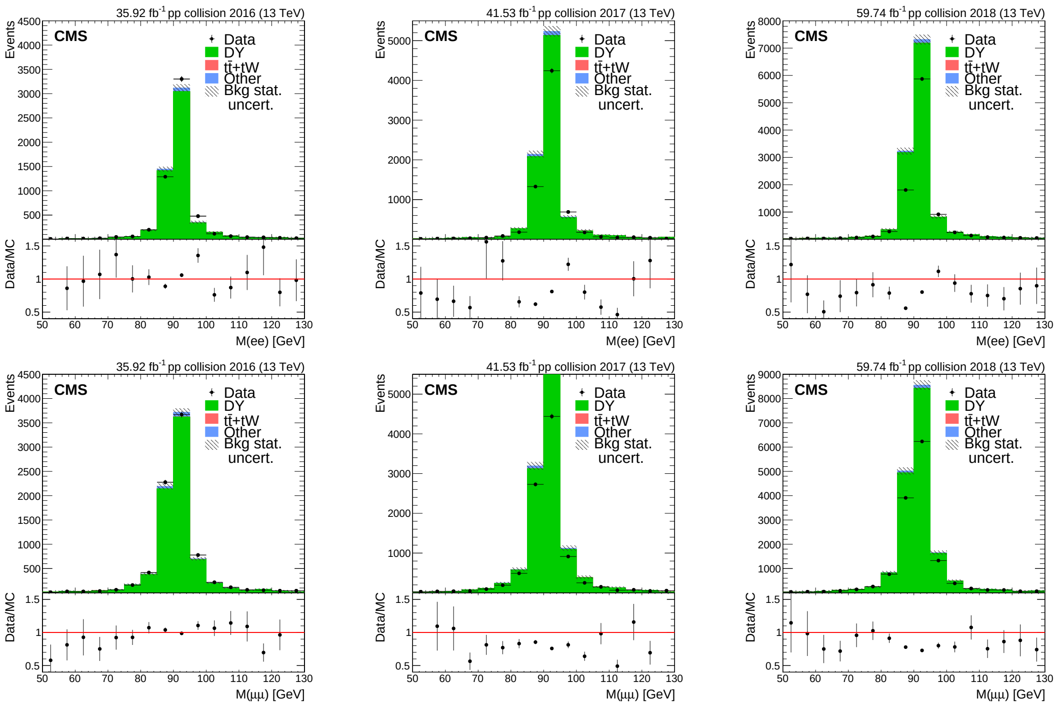


Figure 4.24: Data/MC comparison of $M(\ell\ell)$ in the Z peak region, $60 < M(\ell\ell) < 120$ GeV, for the $eeqq$(top) and $\mu\mu qq$ (bottom) channel for the years 2016 (left), 2017 (middle), 2018 (right). Events are selected with signal region requirements, except for $M(\ell\ell)$. DY MC distributions are corrected using the NLO QCD and EW k-factors defined in Fig. 4.17 .

discriminant variable $M(\ell\ell J)$ to the expected background in a region above the Z mass, with $150 < M(\ell\ell) < 300$ GeV, but still below our signal region, which will be our new Control Region for DY background estimate. In Figs. 4.25 we show the $M(\ell\ell)$ distribution for Data/MC in the region $150 < M(\ell\ell) < 300$ GeV for the electron and muon channel. We find that the data/MC comparison remains rather constant in the full range for this variable. In Fig. 4.26 we show the $M(\ell\ell J)$ comparison for data and MC(without alpha-ratio) in the region with $150 < M(\ell\ell) < 300$ GeV. We do see a dependence of the ratio data/MC for the $M(\ell\ell J)$ variable.

In Fig. 4.27 we show the Data/MC distribution in $(M\ell\ell J)$ variable in the region $150 < M(\ell\ell) < 300$ GeV with the MC corrected with the binned alpha-ratio. As expected the alpha-ratio does not fully correct the dependence as the statistic from the Z peak into this kinematical region runs out quickly. The solution proposed, as suggested also by the sub-conveners, is then to use the plot of the DY MC in $M(\ell\ell J)$

from events with $150 < M(\ell\ell) < 300$ GeV, corrected for the alpha-ration binned in $M(\ell\ell J)$, as an input to the combine fit as a DY Control Region. The uncertainties will be taken care as systematics in the final evaluation by the combine fit.

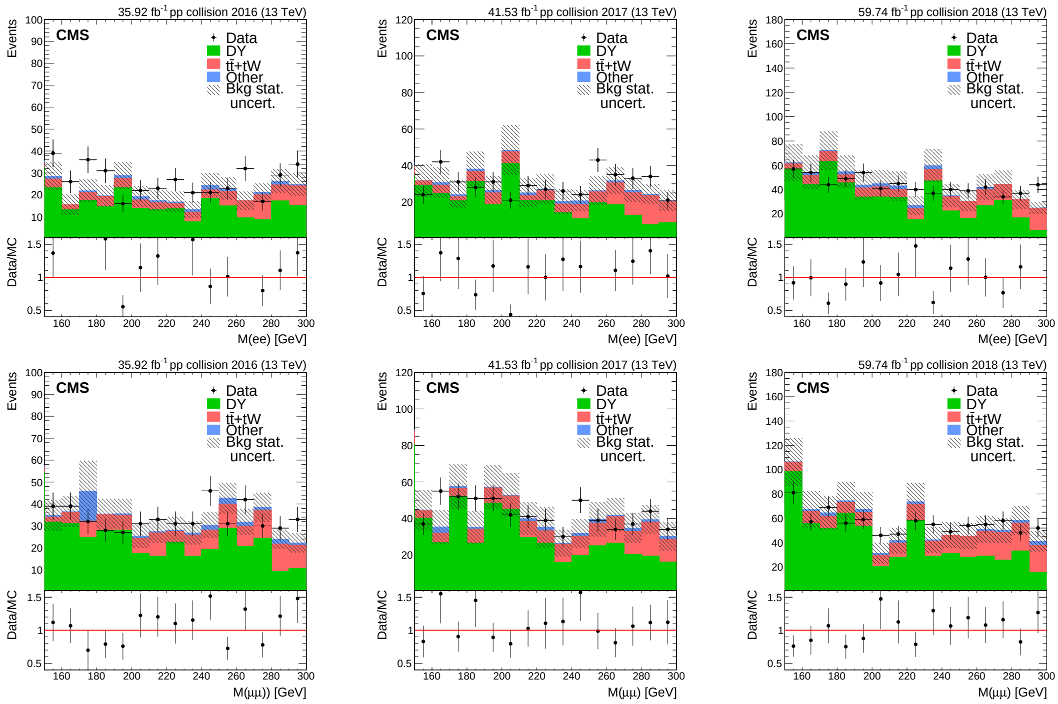


Figure 4.25: Data to MC comparison in the control region for $150 < M_{\ell\ell} < 300$ GeV for the $eeqq$ (top line) and $\mu\mu qq$ channel (bottom lines) shown for 2016 (left), 2017 (middle), 2018 (right). The data/MC ratio points are fitted with a line with a slope. The DY MC distributions are corrected by the NLO k-factor presented above. No Alpha-ratio applied here

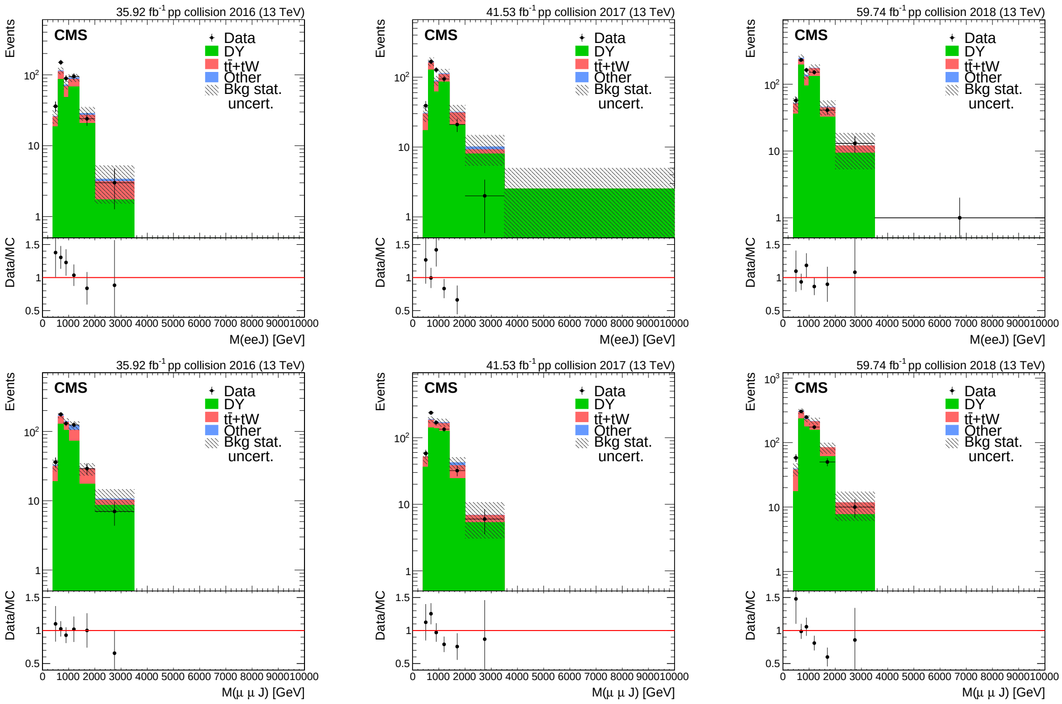


Figure 4.26: Data to MC comparison of $M(\ell\ell J)$ in the control region for $150 < M_{\ell\ell} < 300$ GeV for the $eeqq$ (top line) and $\mu\mu qq$ channel (bottom lines) shown for 2016 (left), 2017 (middle), 2018 (right). The data/MC ratio points are fitted with a line with a slope. The DY MC distributions are corrected using the NLO k-factor presented above. No Correction with alpha-ratio.

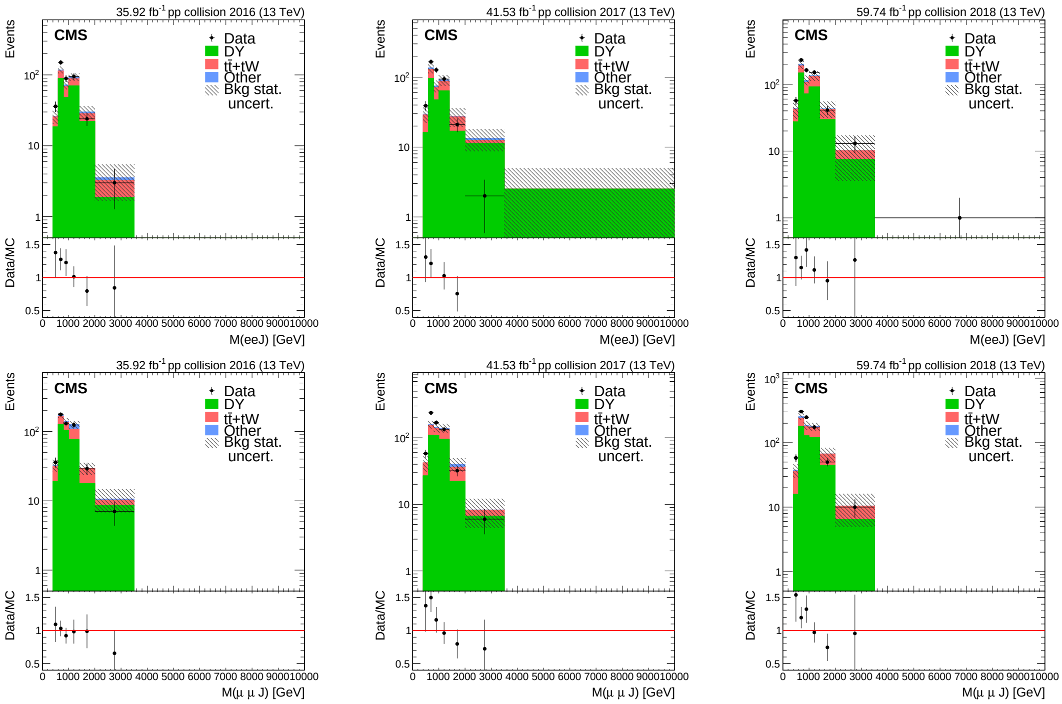


Figure 4.27: Data to MC comparison of $M(\ell\ell J)$ in the Control Region for $150 < M_{ee} < 300$ GeV for the $eeqq$ (top line) and $\mu\mu qq$ channel (bottom lines) shown for 2016 (left), 2017 (middle), 2018 (right). The MC DY distribution is corrected with the alpha-ratio factor extracted above. This is the plot used for the final fit of the DY background estimate.

5.2 ESTIMATION OF THE $t\bar{t}$ AND tW BACKGROUNDS

The second most important background is due to leptons from $t\bar{t}$ and single-top production and is estimated directly from MC simulation. All these processes are flavor-symmetric and have a branching ratio to a pair of leptons of different flavor, $e\mu$, twice as large as the branching ratio to ee. This means the $e\mu$ mass spectrum provides an excellent probe to validate the Monte Carlo predictions for these backgrounds. The control region to check the data-MC agreement has been defined starting from the request of one muon and one electron. The leading lepton is required to have $p_T > 55$ GeV, while for the subleading $p_T > 35$ GeV is requested. Furthermore a veto is applied on events for which an electron and a muon with $p_T > 5$ GeV are found to be in a cone of $\Delta R < 0.1$, with $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. Fig. 4.28 shows the invariant mass spectrum of $e\mu$ events in data and MC spectrum for the 2016, 2017 and 2018 datasets. After having checked that the simulations describe the sample of $e\mu$ events well, the contribution of these backgrounds is estimated using the MC simulated sample with its cross section, adopting the full set of cuts of the signal region, only changing the flavor of the leptons (as represented in Fig.5.13).

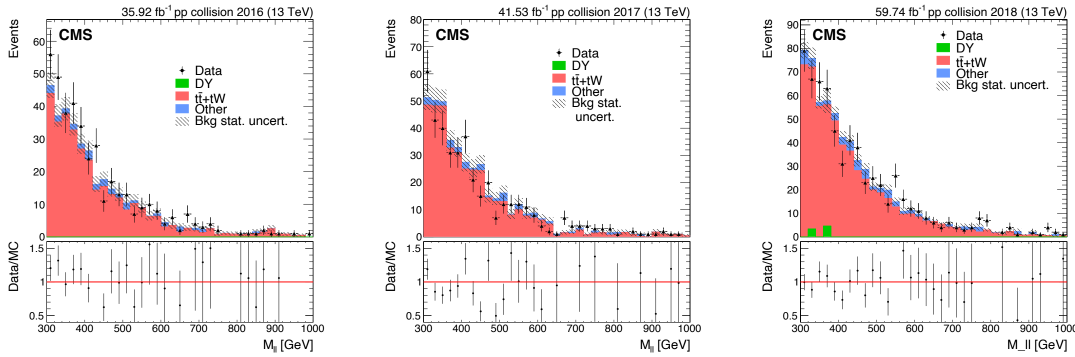


Figure 4.28: Comparison of data and MC estimate for $t\bar{t} + tW$ in the top enriched control region. From left, years: 2016, 2017, 2018.

Moreover, we have verified that there is no significant dependence from the mis-modeling of the top p_T in the Data and MC comparison in the control region for different kinematical variables that could be more sensitive to this effect with respect to the di-lepton mass (Figure 4.29).

5.3 ESTIMATION OF QCD MULTIJET BACKGROUND

We evaluate the QCD multijet background from poorly identified, non-isolated lepton candidates selected from data and weighted by a correction factor which allows the final contribution to be extrapolated into the signal region. We rely on the method developed by the $Z' \to \ell\ell$ groups [129] [130] and their scale factors that cover a similar p_T range as this analysis. The used formula is

$$QCD_{SR} = QCD_{CR} \times W^{i}_{SR/CR} \times W^{j}_{SR/CR}$$

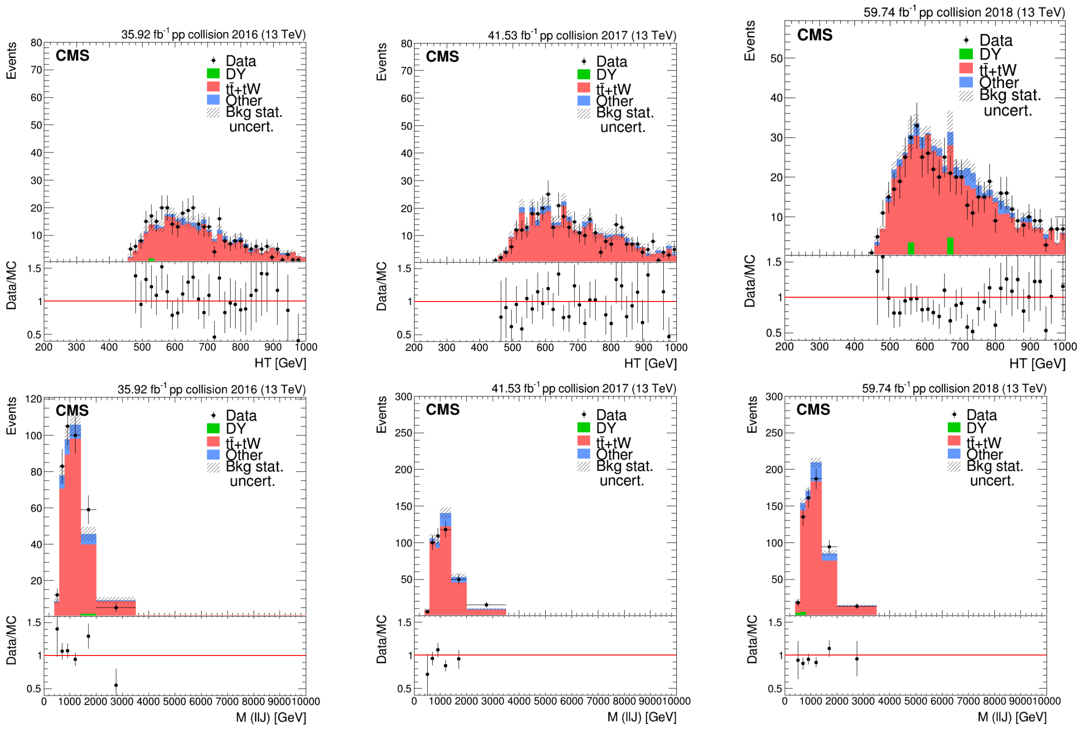


Figure 4.29: Distributions for the HT and $M(e\mu J)$ in the top control region: no significant dependence from the mis-modeling of the top p_T comparing Data and MC is shown.

where QCD_{SR} is the multijet contribution estimated in the signal region; QCD_{CR} is taken from the data, applying the signal region selection on leptons selected with a looser selection (fake-rate preselection) defined in Tab.4.10 and Tab.4.12. $W_{SR/CR}$ is a weight used to correct QCD_{CR} in order to estimate QCD_{SR}, and i, j refer to the two lepton candidates. In particular $W_{SR/CR}$ is defined as $FR/(1-FR)$, where FR is the p_T, η dependent fake-rate measured by the $Z' \to \ell\ell$ groups.

5.3.1 *Estimation in the eeqq channel*

In the $eeqq$ channel the fake-rate preselection is defined as in Table 4.10, while the pT,η dependent FR factor in Table 4.11.

variable	barrel	endcap
$\sigma_{i\eta i\eta}$	< 0.013	< 0.034
H/E	< 0.15	< 0.10
nr. missing hits	≤ 1	≤ 1
$\lvert d_{xy} \rvert$	< 0.02	< 0.05

Table 4.10: Fake rate preselection in the $eeqq$ channel [129].

We estimate QCD_{SR} to be equal to 0.14 ± 0.02 events, corresponding to less than 1% of the total expected background and thus we decide not to include it in the final distribution of $M(e, e, J)$.

region	E_T range (GeV)	functional form
barrel	$35.0 \leq E_T < 131.6$ GeV	$0.106 - 0.0025 \times E_T + 2.26e^{-05} \times E_T^2 - 7.11e^{-08} \times E_T^3$
	$131.6 \leq E_T < 359.3$ GeV	$0.0139 - 0.000104 \times E_T + 3.6e^{-07} \times E_T^2 - 4.13e^{-10} \times E_T^3$
	$E_T \geq 359.3$ GeV	$0.00264 + 3.38e^{-06} \times E_T$
barrel	$35.0 \leq E_T < 131.5$ GeV	$0.106 - 0.00252 \times E_T + 2.28 \times 10^{-5} \times E_T^2 - 7.21 \times 10^{-8} \times E_T^3$
	$131.5 \leq E_T < 355.0$ GeV	$0.0138 - 0.000103 \times E_T + 3.62 \times 10^{-7} \times E_T^2 - 4.25 \times 10^{-10} \times E_T^3$
	$E_T \geq 355.0$ GeV	$0.00279 + 2.43 \times 10^{-6} \times E_T$
endcap $\|\eta\| < 2.0$	$35.0 \leq E_T < 122.0$ GeV	$0.117 - 0.0013 \times E_T + 4.67 \times 10^{-6} \times E_T^2$
	$122.0 \leq E_T < 226.3$ GeV	$0.0345 - 4.76 \times 10^{-5} \times E_T$
	$E_T \geq 226.3$ GeV	$0.0258 - 9.09 \times 10^{-6} \times E_T$
endcap $\|\eta\| > 2.0$	$35.0 \leq E_T < 112.5$ GeV	$0.0809 - 0.000343 \times E_T$
	$E_T \geq 112.5$ GeV	0.0423

Table 4.11: Functional form of the fake-rate FR used for electrons in the QCD multijet estimation [129].

5.3.2 *Estimation in the $\mu\mu qq$ channel*

In the $\mu\mu qq$ channel the fake-rate preselection is defined in Table 4.12, while the pT,η dependent FR factor in Table 4.13. The coefficients a_i, b_i, c_i, and d_i are derived from a fit procedure detailed in Ref. 4.12.

variable	cut value
is GlobalMu and is TrackerMu	true
$\|d_z\|$	< 1.0
$\|d_{xy}\|$	< 0.2
Nb. of Tracker Layers with Measurement	> 5
Nb. of Valid Pixel Hits	> 0
Matching with HLT object (Mu50 || TkMu50)	true

Table 4.12: Fake rate preselection in the $\mu\mu qq$ channel [130].

Eta region	p_T range (GeV)	functional form
Barrel	$53 < p_T \leq 200$	$a_1 + b_1 \times p_T + c_1 \times p_T^2$
Barrel	$200 < p_T \leq 800$	$a_2 + b_2 / [c_2 + p_T]$
Barrel	$p_T > 800$	d_1
Endcap	$53 < p_T \leq 250$	$a_3 + b_3 \times p_T + c_3 \times p_T^2$
Endcap	$p_T > 250$	d_2

Table 4.13: Functional form of the fake-rate FR, defined in the text, used for muons in the QCD multijet estimation [130].

We estimate QCD_{SR} to be equal to $\simeq 0.055 \pm 0.003$. As it is lower than the other backgrounds, we decided to neglect it.

6 SYSTEMATIC UNCERTAINTIES

Systematic effects on the background estimations and the simulated MC signals have to be considered before dealing with the interpretation of the results.

For the uncertainty on the luminosity we consider a value of: 2.5% for 2016 [131], 2.3% for 2017 [132], and 2.5% for 2018 [133] on the normalization of both backgrounds and MC signals.

For the shape uncertainties related to the corrections applied on the MC samples, and listed in Section 4.2.2, we proceed as follows:

1. We take the central value of our measurement given by the distribution $M(\ell, \ell, J)$, which contains the backgrounds estimated according to the methods described in Section 4.5.

2. For every systematic effect we repeat the whole analysis twice, considering the deviations up and down related to systematics, and producing two distributions accordingly: $M(\ell, \ell, J)_{Up}$ and $M(\ell, \ell, J)_{Down}$.

3. We supply these two additional histograms for each systematic uncertainty to the "combine" tool [134] for the statistical analysis, as described in the next section.

This approach is used for the following systematic uncertainties:

- **Pileup**: pileup weights are re-calculated by varying the value 69000 μb of minBiasXsec of $\pm 5\%$.

- **Lepton Scale Factors**: discrepancies in the lepton reconstruction, identification, and isolation efficiencies between data, simulation and trigger are corrected by applying a scale factor to all the simulated samples. The scale factors, which depend on the p_T and η, are varied by $\pm\sigma$ and the change in the yield in the signal region is taken as the systematic. Scale factors are taken from [135] for electrons and [120] for muons.

- **Lepton momentum scale and resolution**: the lepton momentum scale uncertainty is computed by varying the momentum of the leptons by their uncertainties. For muons with $p_T < 200$ GeV the systematic has been evaluated by applying/not applying the Rochester corrections [121]. For muons with $p_T >$ 200GeV,the generalized–endpoint (GE) method [120] was applied, and the uncertainties on the muon curvature bias are taken from a Gaussian distribution. Moreover, the systematic uncertainty associated to the smearing applied to correct high-momentum muon resolution has been considered, as recommended in [120]. For electrons, we used the MiniAOD V2 energy corrections [136], and corresponding uncertainties.

- **Jet energy scale and resolution**: in order to have the resolution in the simulation match that in the data the momentum of the jets is smeared as:

$$p_T \rightarrow max[0, p_T^{gen} + c_{\pm 1\sigma} \cdot (p_T = p_T^{gen})] \tag{4.5}$$

in which $c_{\pm 1\sigma}$ are the data/MC scale factors shifted by $\pm\sigma$ according to the prescriptions in [125, 126]

- **Background Systematics**: the uncertainty on the "α-rato" method is derived in Section 4.5.1.

Moreover, we include an estimation of the systematics for the PDF uncertainty of the signal sample generation. Being a new massive particle in the TeV scale, the Parton Density Functions (PDFs) are employed in regions where available experimental constraints are limited, notably close to the hadronic threshold where $x \to 1$. For the HCN sample generation, the NNPDF3.0 [137] set was used. To assess the impact of these systematic uncertainties on the signal, we have applied the standard "PDF4LHC" recommendation [138], using MC replicas with up and down variations of the PDFs. The effect of these uncertainties has been evaluated on the acceptance of our signal selection, and a "LogNormal" uncertainty from 12% to 48% was derived, depending on the HCMN mass. Figure 4.30 shows the PDF errors for both the channels and for the HCN masses explored in this analysis. The uncertainty is considered fully correlated across the years.

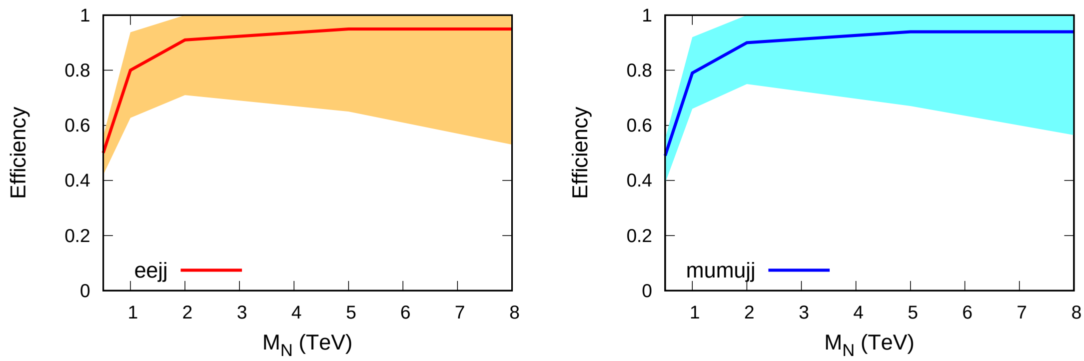


Figure 4.30: The effects of the PDF uncertainties on the cumulative efficiency in the signal region for the electron channel (left) and the muon channel (right) for the benchmark compositeness scale value $\Lambda = 13$ TeV. The result is presented for the 2018 samples, but is consistent for the three years.

Source	Bkgd./Signal process	Type	ee bkgd. (%)	ee signal (%)	$\mu\mu$ bkgd. (%)	$\mu\mu$ signal (%)
Integrated luminosity	All bkgd./Signal	LogN	2.3–2.5	2.3–2.5	2.3–2.5	2.3–2.5
Jet energy resolution	All bkgd./Signal	Shape	0.5–1.4	0–0.9	0.2–1.2	0–0.7
Jet energy scale	All bkgd./Signal	Shape	1.9–4.2	0–0.4	2.1–3.4	0–0.1
Muon SF (trigger, reco., ID)	All bkgd./Signal	Shape	-	-	0.4–1.0	5.8–36.8
Muon momentum scale/resolution	All bkgd./Signal	Shape	-	-	10-20	0.1–0.2
Electron (trigger, reco., ID)	All bkgd./Signal	Shape	1.0–1.6	0.8–1.4	-	-
Electron energy scale/resolution	All bkgd./Signal	Shape	0.5–10		-	-
Pileup modeling	All bkgd./Signal	Shape	2–18	0.1–1.1	2–18	0.1–1.9
DY bkgs estimation (α-ratio stat. unc.)	DY	Shape	10	-	10	-
PDF error	Signal	LogN	-	15	-	15

Table 4.14: Summary of the relative systematic uncertainties in signal region (effect estimated on the $M(\ell\ell J)$ variable for the last three bins). The numbers for signal is obtained for $M(N_\ell) = 1$ TeV and for the two most relevant background sources.

7 UNBLINDING OF THE RESULTS

In this section, we unblind the selected LHC Run 2 data in the signal region. In Figures 4.31, 4.32, we show the post-fit distribution of the variable $M(\ell, \ell, J)$ for the estimated SM backgrounds, and a benchmark signal point (with parameters $\Lambda = 13$ TeV and $M(N) = 1000$ GeV) for both the electron and muon channel for the full Run 2 and three years considered separately. The distributions are obtained with a simultaneous fit of the signal region with the DY and top control regions, as explained in the next following sections. In Table 4.15 we present the estimated background yields and the total number of observed events for each channel. The uncertainties on the estimated background events are obtained post-fit and are the sum of the statistical and systematic uncertainties.

The observations are in decent agreement with the SM predictions and will be examined further as part of the future development of this data analysis. In case no evidence of new physics is found in all the mass spectrum, upper limits can be measured on the production cross-section of the heavy Majorna neutrino in the di-lepton di-jets final state.

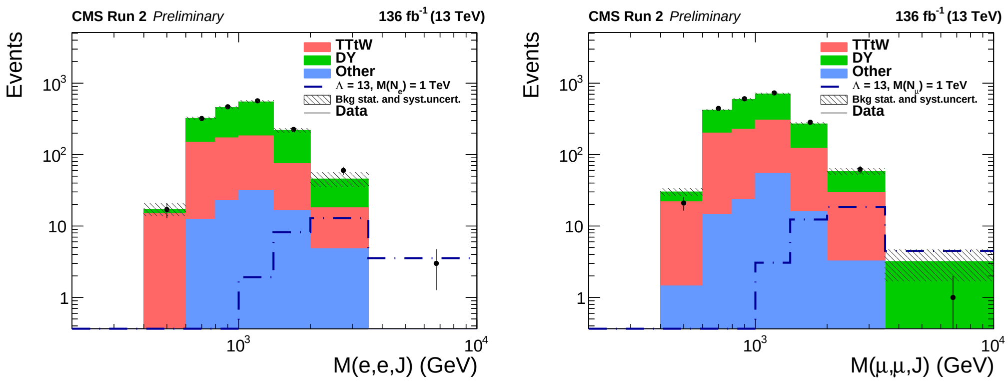


Figure 4.31: Post-fit distributions of the variable $M(\ell, \ell, J)$ of the data, the backgrounds (stack plots) as estimated in Sec. 4.5 and signal expected from simulation (lines) in the signal region, having considered the signal parameters $\Lambda = 13$ TeV and $M(N) = 1000$ GeV, for the $eeqq$ (top) and the $\mu\mu qq$ (bottom) channels. The three years of Run 2 data taking are combined.

Year	Channel	DY	TTtW	Others	Total background	Data
2016	eejj	39± 6.8	20.2±1.4	5.5±0.44	64.3 ± 6.9	70
	$\mu\mu$jj	50.5± 5.5	30.3±1.0	5.30±0.22	86.1 ± 5.6	86
2017	eejj	50 ± 10	17.9±1.4	7.08±0.93	75 ± 10	84
	$\mu\mu$jj	57.4± 6.1	43.1±1.4	4.98±0.18	105.5 ± 6.2	117
2018	eejj	83± 13	33.5±1.7	8.7±1.1	125 ± 13	134
	$\mu\mu$jj	70.4± 9.1	59.7±1.8	8.89±0.46	139.0 ± 9.3	144
Combined	eejj	172± 18	71.6±2.6	21.3±1.5	265 ± 18	288
	$\mu\mu$jj	178± 12	133.2 ±2.5	19.18±0.54	331 ± 12	347

Table 4.15: Summary table of expected (and observed) events in the signal region for the last three bins in $M(\ell\ell J)$.

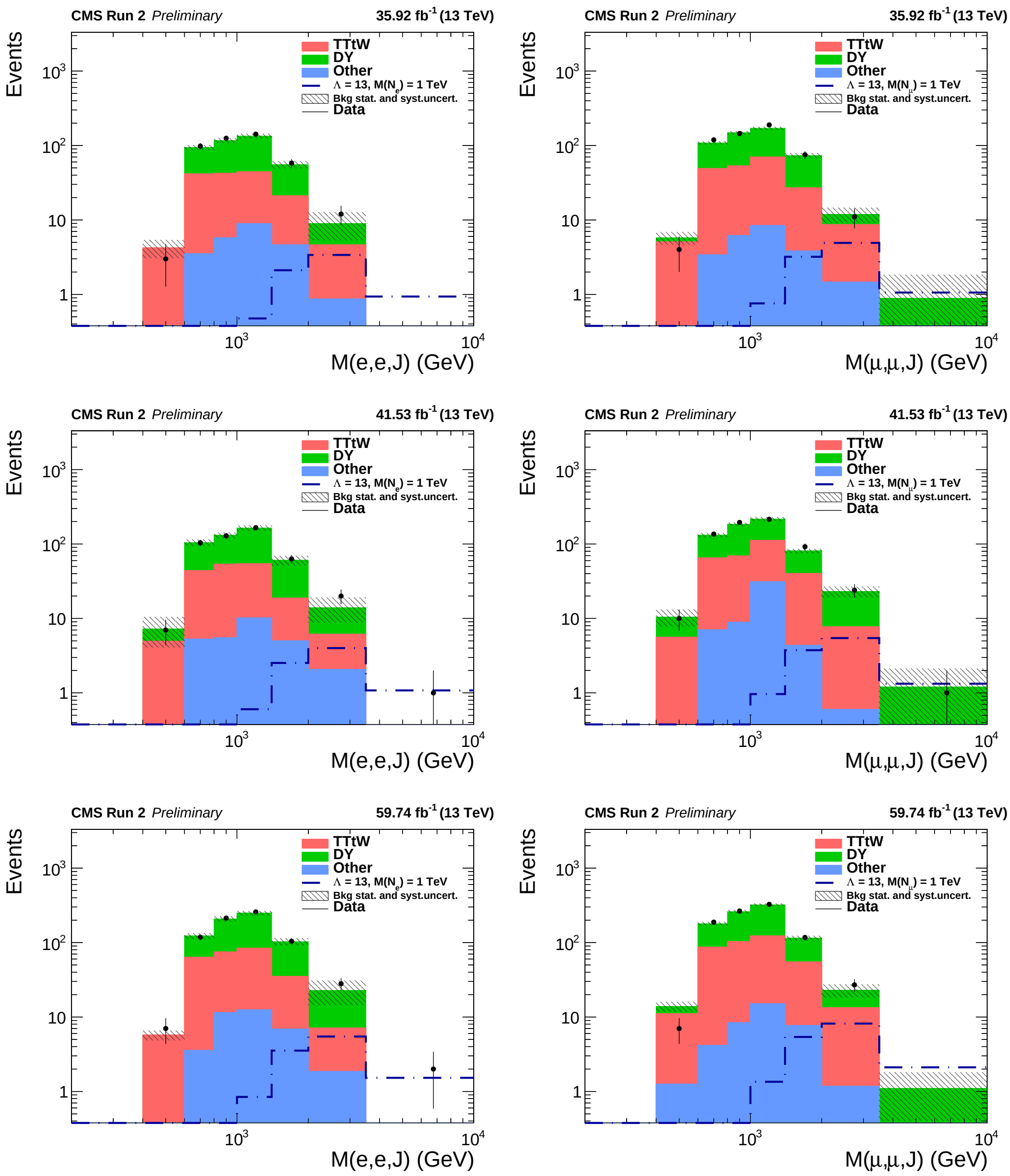


Figure 4.32: Post-fit distributions of the variable $M(\ell, \ell, J)$ of the data, the backgrounds (stack plots) as estimated in Sec. 4.5 and signal expected from simulation (lines) in the signal region, having considered the signal parameters $\Lambda = 13$ TeV and $M(N) = 1000$ GeV, for the *eeqq* (left) and the $\mu\mu qq$ (right) channels. The three years are reported separately, normalized to the corresponding luminosity, and applying the correction described in Sec. 4.3: 2016 on the top, 2017 on the middle, and 2018 on the bottom.

8 STATISTICAL INTERPRETATION OF THE RESULTS

In this section we give a statistical interpretation of the expected results from our MC predictions, setting upper limits on the existence of a composite Majorana neutrino.

The CL_s technique [139, 140] is used to extract the upper limits on the product of the cross section and branching fraction of the signal process. This provides a quantitative assessment about the incompatibility of the signal $-$ plus background model and the only-background hypotheses, provided no excess of data relative to background-only expectation is observed.

The probability P to observe k events under a model that predicts λ events is computed by means of the Poisson statistic as:

$$P(k \mid \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

In this analysis λ is the sum of expected background yields. since not only event yields are considered but also the shape of the invariant mass distribution (or any quantity), P must be promoted to the Poisson likelihood,where S and B are the signal and background estimates respectively. The Poisson likelihood thus reads:

$$\mathcal{L}(k \mid \mu, \theta) = \prod_{i=1}^{N_{bins}} \frac{[\mu S(\theta)_i + B(\theta)_i]^{k_i} \, e^{[\mu S(\theta)_i + B(\theta)_i]}}{k_i!}.$$

A test statistic, t_μ, is defined as the ratio of likelihoods so that $\hat{\theta}$ and $\hat{\mu}$ maximize the numerator and the denominator, respectively

$$t_\mu = -2 \ln \left[\frac{\mathcal{L}\left(s \mid \mu, \hat{\theta}_\mu\right)}{\mathcal{L}(s \mid \hat{\mu}, \hat{\theta})} \right]$$

where the numerator represents the best agreement to the observed data for a fixed-size signal, while the denominator is the configuration of signal plus-background models that best fits the observed data. An array of t_μ is produced by generating a large number of pseudo-experiments, in which t_μ are determined by considering the pseudo-data as k. Pseudo-experiments are randomly drawn from a pool based on the average expectations derived from the analysis, which are then shifted up and down as a function of the uncertainties. The resulting distribution is compared to the observed value, t^{obs}, which is calculated from the observed data. The CL_s is the ratio of p -values for the signal-plus-background (p_{s+b}) and the background-only (p_b) hypothesis,

$$CL_s = \frac{p_{s+b}}{p_b}$$

where

$$\begin{aligned} p_{s+b} =& P\left[t_\mu > t_\mu^{obs} \mid \mu S\left(\hat{\theta}_\mu^{obs}\right) + B\left(\hat{\theta}_\mu^{obs}\right)\right] \\ p_b =& P\left[t_\mu > t_0^{obs} \mid B\left(\hat{\theta}_0^{obs}\right)\right] \end{aligned}$$

The confidence level (CL) is given by $(1 - CL_s)$. Consequently this procedure is reiterated for varying μ until $CL_s = 0.05$ in order to obtain a limit at the 95% CL.

8.1 UPPER LIMIT EXTRACTION

In the section, we extract the expected upper limits on the existence of Heavy Composite Majorana Neutrino in the di-lepton and large-radius jets channel, without using the observed events. The upper limits have been evaluated using the CMS high-level tool "combine" [134] and computing limits with toys data generation, providing as inputs the $M(\ell, \ell, J)$ mass shape of estimated backgrounds, the expected signal, and the data. The fit procedure includes both the top control region $M(e\mu J)$ distribution (Figure 4.33), and the DY control region (in the mass window $150 < M(\ell\ell) < 300$ GeV) $M(\ell\ell J)$ distributions.

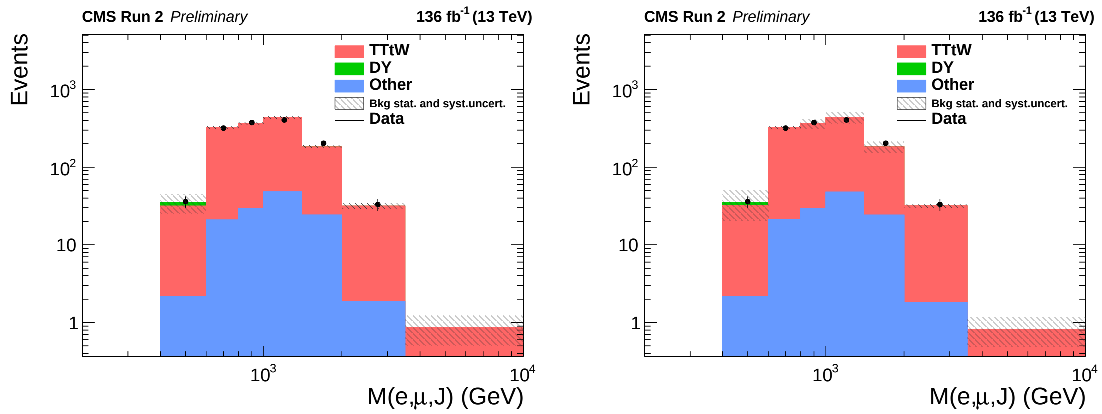


Figure 4.33: Post-fit distribution of $M(e, \mu, J)$ in the top enriched phase space defined in Section 4.5, for the electron flavor on the left, and the muon flavor on the right.

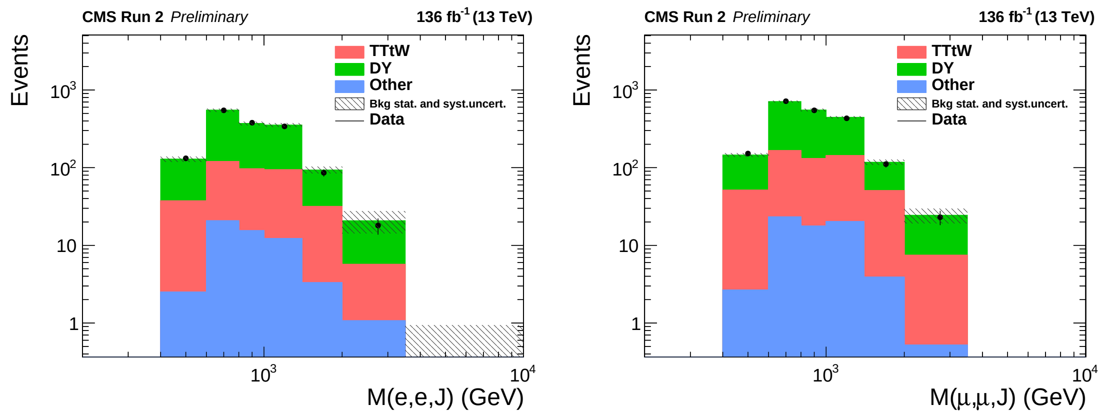


Figure 4.34: Post-fit distribution of $M(\ell, \ell, J)$ in the Drell-Yan enriched phase space defined in Section 4.5, for the electron flavor on the left, and the muon flavor on the right.

We incorporate the systematic uncertainties listed in Sec. 4.6, which are treated as uncorrelated among the different data-taking years. The expected limits from the full Run 2 data set for the *eeqq* channel and the $\mu\mu qq$ channel, displayed in Fig. 4.35, provide an exclusion for the Heavy Composite Majorana Neutrino up to a mass of 3.8 (2.8) TeV, for the reference value of $\Lambda = 13$ TeV in the *eeqq* ($\mu\mu qq$) channel.
The limits on the process cross-section can be also presented in the 2-dimensional plane $(M(N), \Lambda)$ for a more practical comparison with the unitarity restrictions

on the parameters space. The results are shown in Fig. 4.36 for both the channels. The accessible compositeness scale domain is almost double than that in the 2015 search (order of 20 TeV for the full Run 2 versus the order of 10 TeV for 2015), whether the mass reach is increased by a factor of 0.5. To conclude, we assess the effect of the various systematic uncertainties presented in Section 4.6 on the signal strength measurement of this search. This is often referred to as the calculation of the "impact" of each uncertainty, i.e. we determine the shift in the signal strength, for the best-fit, that is induced if a given nuisance parameter is shifted by its $\pm 1\sigma$ post-fit uncertainty values. If the signal strength shifts a lot, it tells us that it has a strong dependency on this systematic uncertainty. In Figure 4.37, we present the impact plots for both the channels, considering the reference HCN signal with mass 1000 GeV and compositeness scale of 13 TeV.

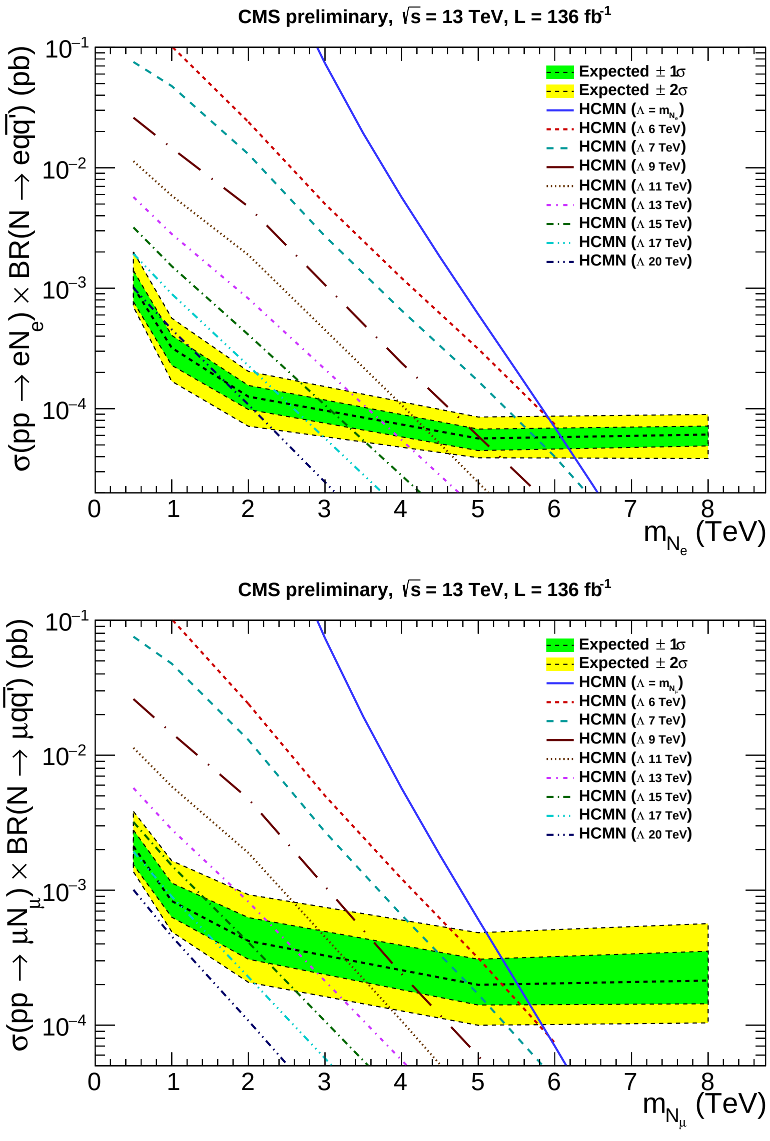


Figure 4.35: Exclusion limit for the combination of the full Run 2 data set for the $eeqq$ (top) and $\mu\mu qq$ (bottom) channels. Error bars account for both statistical and systematic uncertainty.

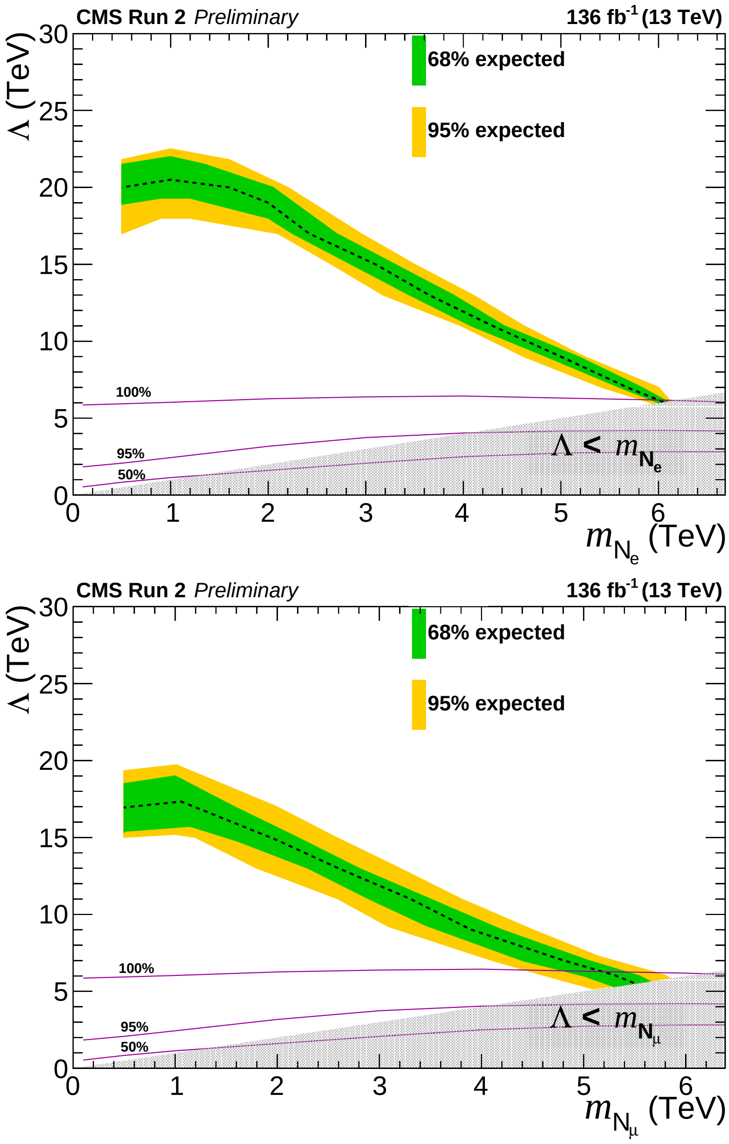


Figure 4.36: Exclusion limit for the $eeqq$ (top) and $\mu\mu qq$ (bottom) channel for the combination of the three years of data taking, 2016, 2017, and 2018, in the 2D plane. The solid (violet) lines represent the fraction of MC events that satisfy the unitarity condition in the EFT approximation [110] with the various percentages considered.

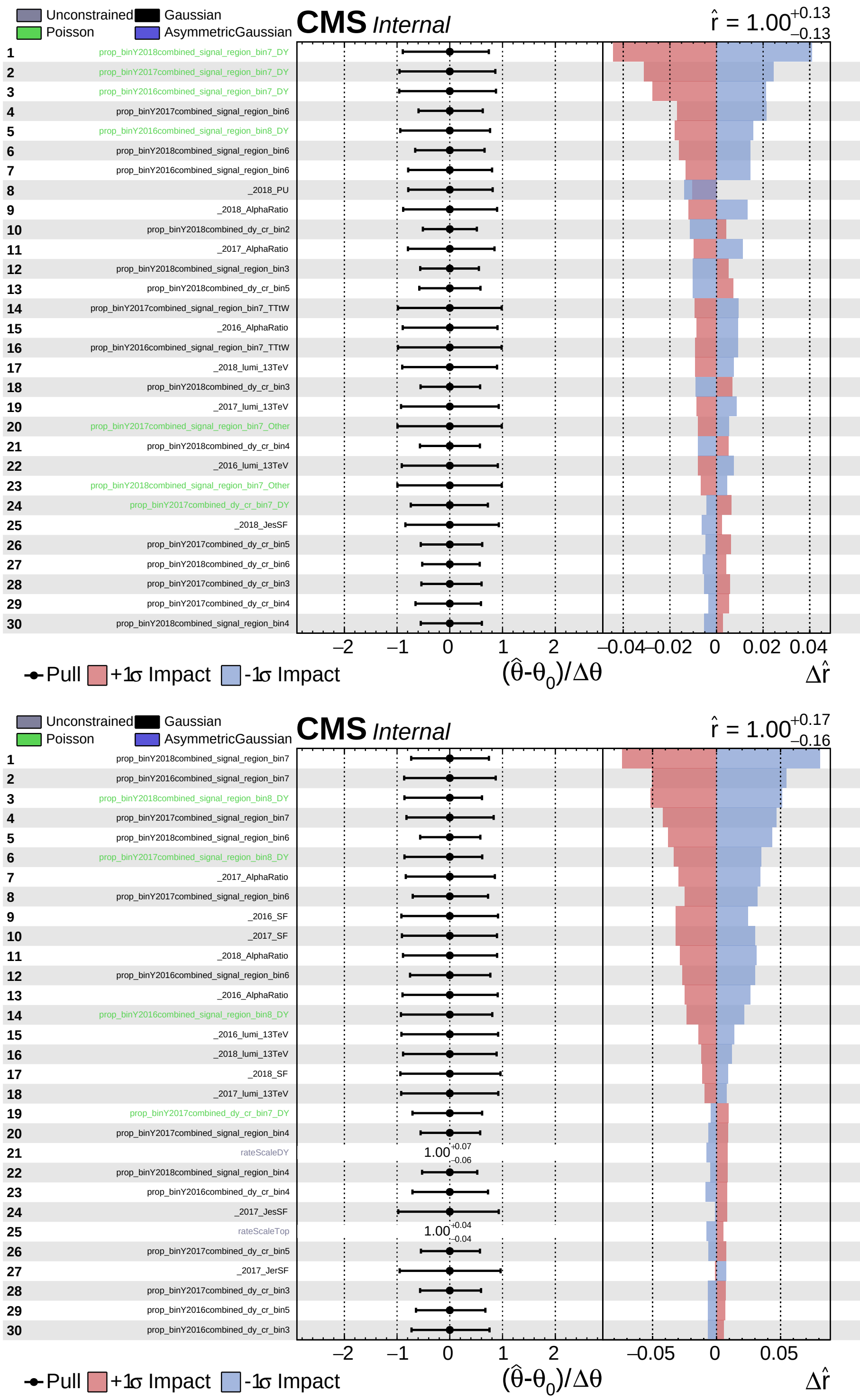


Figure 4.37: Impact plots for the main systematic uncertainties, considering the reference HCN mass of 1000 GeV and compositeness scale of 13 TeV. Top plot: electron channel. Bottom plot: muon channel. The plots are obtained using the higher-level script "combineTool.py" [134]. The nuisances, ordered by importance, are indicated on the left panels. Nuisances labeled as "prop_bin*"account for the statistical fluctuation of a specific bin in the input histogram. Among these, the nuisances highlighted in green consider bins with a low number of events (~ 10).

9 SUMMARY OF THE RESULTS

A search for new physics beyond the standard model has been performed considering a framework of composite fermions that introduces a violation of the lepton number through the presence of a Heavy Majorana Neutrino [104]. The model was previously used for an experimental analysis with CMS Run 2 data collected for pp collisions at $\sqrt{s} = 13$ TeV and an integrated luminosity of 2.3 fb^{-1}, observing no excesses in $\ell\ell q\bar{q}'$ ($\ell = e, \mu$) channel with respect to SM predictions.
Here we have addressed the same scenario to the full LHC Run 2 data set, taking into account for the first time unitarity bounds on the model parameter space for the optimization of the analysis fiducial region. The study has been carried out considering a heavy composite Majorana neutrino produced in association with a lepton and decaying into a same-flavor lepton plus two quarks and requiring two leptons and at least one AK8 jet in the signal region. The data is in good agreement with the SM expectations, and we have computed the expected upper limit at 95% CL on the on $\sigma(pp \to \ell N_\ell) \times \mathcal{B}(N_\ell \to \ell q\bar{q}')$, for both an electron and muon HCN. The present search represents a considerable improvement of the sensitivity to heavy composite Majorana neutrinos at the LHC over a much larger parameter space compared to the previous Run 2 result. The parameters explored are also expected to lie within the region of validity of the EFT that describes the interactions of the model.

5 VECTOR BOSON SCATTERING IN THE ZV SEMILEPTONIC CHANNEL

1 INTRODUCTION TO MULTIBOSON SEARCHES AT THE LHC

The main target of the LHC is to probe the nature of the electroweak symmetry breaking. The discovery of the Higgs boson and then the measurements of its properties are a great success, but they are not sufficient for an exhaustive picture of the mechanism. The SM predicts interaction vertices with three (triple gauge coupling) or four (quartic gauge couplings) bosons. Several collider experiments explored these processes in the past [141]. The Large Electron-Positron (LEP) collider accessed the WW and WZ channels and studied their production cross section as a function of the center-of-mass energy. This brought the first confirmation of the correctness of the SM, critically limiting the possible divergent behavior of the WWWW and WZWZ vertices, which we mentioned in the context of unitarity bounds in Section 2.1.1. The CDF and D0 experiments at Tevatron also measured diboson final states, before the beginning of the extensive ATLAS and CMS research programs at the LHC.

There are a variety of theoretical motivations for testing the structure of multi-boson processes at the LHC. While this was done extensively for the QCD-mediated interactions, there is still much ognoing effort on the study of multiple EW gauge boson interactions. Thanks to the enormous progress made in precision calculations, many QCD processes are now estimated up to Next-to-Next-to-Leading-Order (NNLO), and their implementation in MC generators has reached a high level of agreement with data in the strong sector. Within this context, we present here the implementation of the very first of the analysis for the $\mathcal{O}(\alpha^6)$ electroweak ZVjj production mediated by a Vector Boson Scattering (VBS) mechanism.

In the following part of this chapter, we will focus on the measurement with the full Run 2 data of the electroweak ZVjj semileptonic process, collected with the CMS detector. The entire analysis workflow is presented in detail as described in the following[1]. In Section 5.2, we underline the main features of this channel, which shares the main final states objects (plus the typical VBS tag jets from initial state radiation) with the search for heavy composite Majorana neutrino widely discussed in the previous chapters of this thesis. In Section 5.3, we mention all the data sets and Monte Carlo samples processed and analyzed, and all the corrections implemented. A brief review of the objects reconstructed here is shown in Section 5.4. Then follows a detailed description of the event selection, comparing several alternatives, the estimation of the background sources (Section 5.7), and the treatment of the systematic uncertainties (Section 5.8) before the outcomes of the search.

1 The final results containing the data cannot be shown at the moment, as the analysis is still undergoing the CMS Physics Review.

2 SEARCH FOR VECTOR BOSON SCATTERING PRODUCTION OF ZV DECAYING SEMILEPTONICALLY

The Vector Boson Scattering (VBS) process represents an ideal test bench for the spontaneous symmetry breaking mechanism at hadron colliders, giving primary access to the non-Abelian gauge structure of the electroweak interactions through quartic gauge couplings [141]. Experimentally, VBS is a purely electroweak signature wherein a constituent of each proton emits a boson that interacts with each other producing two bosons in the final states. The remnant of the proton emission leads to two forward/backward jets in the beam direction characterized by large rapidity separation and a large dijet mass, often called "tag jets". Figure 5.3 shows a pictorial representation of the topology of a VBS event with respect to the CMS sensing volume coordinates. Both ATLAS and CMS have recently claimed the very first observation of electroweak production of same-sign W boson pairs in proton-proton collisions at 13 TeV [142, 143], but still, no evidence is shown in the semi-leptonic channel in which a vector boson, V (V = W, Z), decays in leptons and the other V decays in quarks.

In this section, we presented a search for electroweak production of the VZ bosons at $\mathcal{O}(\alpha^6)$, where V is the vector boson decaying hadronically (V=W, Z) and Z decays in charged leptons (e, μ), plus two jets in proton-proton collisions at 13 TeV (Figure 5.1). The data correspond to an integrated luminosity of 136 fb^{-1} collected from 2016 to 2018.

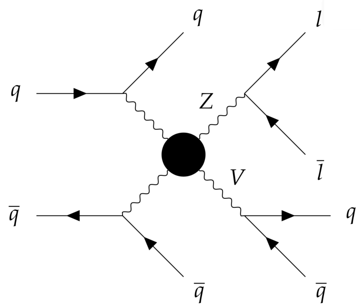


Figure 5.1: The Feynman diagram of the VBS process with an hadronically decaying gauge boson V, a Z. A dark blob in the vertex represents the possible inclusion of anomalous couplings contribution.

The semi-leptonic final state profits of the fact that the $V \to q\bar{q}'$ branching fraction is much larger than the leptonic one, however, it suffers from larger background contamination from other processes, such as the irreducible $\mathcal{O}(\alpha^4\alpha_s^2)$ QCD background as is shown in Fig. 5.2, and the DY+jets processes $\mathcal{O}(\alpha^2\alpha_s^4)$. Interference of the EW and QCD amplitudes contributes at $O\left(\alpha^5\alpha_{\mathrm{s}}\right)$, and is shown in Ref. [144] to account for $\mathcal{O}(1\%)$ contribution in the typical VBS fiducial region, and hence are neglected for the moment.

Moreover, the electroweak signal due to the quartic boson self-interaction, interferes with the diagrams of the background processes, making a complete clean separation intrinsically impossible. Two distinct categories are considered to enlarge the signal acceptance, depending on the topology of the V hadronic decay: the "boosted category" where we reconstruct a single object (an AK8 jet), and the

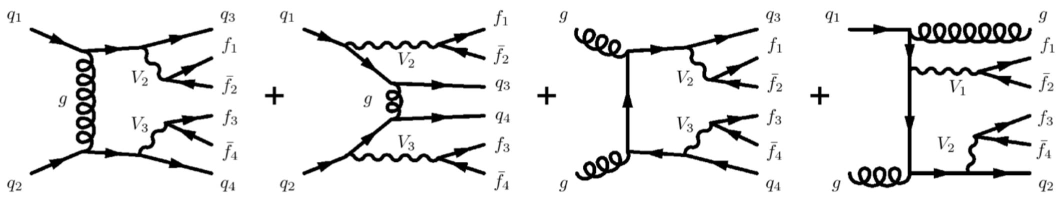


Figure 5.2: The Feynman diagrams of the VBS process QCD induced contribution at $\mathcal{O}(\alpha^4\alpha_s^2)$.

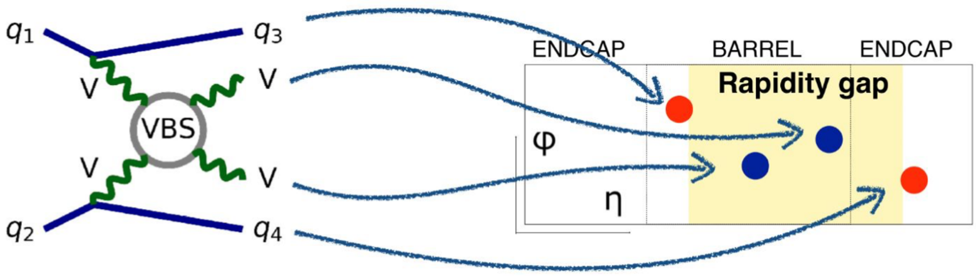


Figure 5.3: A sketch of the VBS topology of the event in the CMS detector parametrization.

"resolved category", when two jets (AK4) are distinctly reconstructed. The complete jets categorization is summarized in Section 5.5. To enhance the analysis sensitivity, we have also tested two distinct multivariate approaches: one based on a BDT, and one making use of explainable artificial intelligence techniques. A detailed description of these methods and the comparison with the "cut-based" analysis strategy is reported in Sec. 5.9.

Besides the pure EW measurement, it is well known that the study of the VBS process gives direct access to the non-Abelian gauge structure of the electroweak interactions through quartic gauge couplings [141]. Therefore, VBS analyses are sensitive to new physics at energy scales beyond the collider direct reach through an effective field theory (EFT) description of the standard model (SM) [144–147]. The second step of the VBS semileptonic analysis consists of the measurement of aQGCs. Such an approach consists of adding higher dimensional operators to the SM Lagrangian:

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \sum_i \frac{g_i \mathcal{O}_i}{\Lambda^{\Delta_i - 4}}, \tag{5.1}$$

where g_i are the so called Wilson coefficients, Λ is the energy scale of the correction and $\Delta_i \leq 4$ the dimensions of the operators, in order to parametrize small deviations from the original SM prediction. In this context, the VBS process can be modified by anomalous quartic gauge couplings (AQGCs), resulting in an enhancement of the production cross-section at large masses of the di-boson system and high-transverse momenta. The current most stringent limits on the Wilson coefficients have been obtained combining the results of the WV and ZV semileptonic channels with a data sample corresponding to an integrated luminosity of 35.9 fb^{-1} collected with the CMS detector in proton-proton collisions at $\sqrt{s} = 13$ TeV [144]. With the full Run 2 data sample, we could narrow the confidence limits on the Wilson coefficients by up to a factor two. This can be achieved, on one hand, with the innovations to the current analysis presented in this proposal, like the introduction of additional selection on new topological variables (such as the Zeppenfeld variable [148]). On the other hand, the improved statistic is shown to reduce the systematic uncertainties, such as the one regarding the normalization of the dominant background source (DY+jets), which was approximately 16%. From a more phenomenological standpoint, it is important to note that the bounds obtained in the semileptonic final state, while being the most stringent, are limited to the interpretation in terms of dimension-8 operators [149]. However, recently the implication of dimension-6 operators in the aQGCs of the VBS [150] has shown the relevance of their inclusion in a more adequate interpretation of the experimental results. Moreover, the implication of perturbative unitarity bounds on EFT couplings has shown the relevance of their inclusion in a more adequate interpretation of the experimental results[2].

3 DATA AND MONTE CARLO SAMPLES

In this section we present the data and Monte Carlo samples used for this search for VBS scattering process at the LHC.

3.1 DATA SAMPLES

The data used in this analysis were collected using the during the 2016, 2017 and 2018 proton-proton runs, and were pre-selected for good luminosity sections using the following "Golden JSON" files:

- 2016:
  Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt ,
- 2017:
  Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON.txt ,
- 2018:
  Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_PromptEraD_Collisions18_JSON.txt,

  Cert_314472-325175_13TeV_PromptReco_Collisions18_JSON.txt

Centrally produced NANOAOD dataformat of "Nano2Apr2020" production (also known as "NANOAODv7") is used. The list of primary datasets and triggers used in the analysis is reported in Tables 5.1 and 5.2, respectively. Events firing at least one of the listed triggers are used in the analysis. Technical details on triggers and their efficiencies are documented in AN-2019/125 [151]. The samples and the HLT paths considered for the three years are reported in Table 5.2. The full sample corresponds to an integrated luminosity of approximately 137 fb^{-1} (35.87 fb $^{-1}$ + 41.53 fb−1 + 59.74 fb^{-1}).

2 This part of the work is still ongoing and concerns the extension of the interpretation of aQGCs measurement in VBS ZV semileptonic in terms of both dimension-6/8 and unitarized dimension-6/8 operators. Given the relevance of a hypothetical measurement of the EW process in the VBS semileptonic channel, the bounds on the EFT/unitarized EFT from the same channel have been given less priority for the moment.

Data	2016	2017	2018
Dataset	MuonEG SingleMuon SingleElectron DoubleMuon DoubleEG	MuonEG SingleMuon SingleElectron DoubleMuon DoubleEG	MuonEG SingleMuon EGamma DoubleMuon
Run	Run2016B-Nano1June2019-ver2-v1 Run2016C-Nano1June2019-v1 Run2016D-Nano1June2019-v1 Run2016E-Nano1June2019-v1 Run2016F-Nano1June2019-v1 Run2016G-Nano1June2019-v1 Run2016H-Nano1June2019-v1	Run2017B-Nano1June2019-v1 Run2017C-Nano1June2019-v1 Run2017D-Nano1June2019-v1 Run2017E-Nano1June2019-v1 Run2017F-Nano1June2019-v1	Run2018A-Nano2Apr20202019-v1 Run2018B-Nano2Apr20202019-v1 Run2018C-Nano2Apr20202019-v1 Run2018D-Nano2Apr20202019-v1

Table 5.1: Dataset and run periods used for 2016, 2017, and 2018.

Year (lumi)	Data Set	Run range	Trigger path
2016 (35.9/fb)	MuonEG	BF(-278273)	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL
		F(278273-)H	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
	SingleMuon	BH	HLT_IsoMu24 HLT_IsoTkMu24
	SingleElectron	BH	HLT_Ele27_WPTight_Gsf HLT_Ele25_eta2p1_WPTight_Gsf
	DoubleMuon	BH(-281613)	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL
		H(281613-)	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ
	DoubleEG	BH	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
2017 (41.53/fb)	MuonEG	B	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
		CF	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
	SingleMuon	BF	HLT_IsoMu27
	SingleElectron	BF	HLT_Ele35_WPTight_Gsf
	DoubleMuon	B	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ
		CF	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8
	DoubleEG	BH	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL
2018 (59.74/fb)	MuonEG	AD	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
	SingleMuon	AD	HLT_IsoMu24
	DoubleMuon	AD	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
	SingleElectron	AD	HLT_Ele32_WPTight_Gsf
	DoubleElectron	AD	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL

Table 5.2: Datasets and corresponding trigger paths used in the analysis.

This analysis considers final states with 2 leptons, be either electrons or muons, therefore a combination of both single and double lepton triggers is used. The trigger efficiency is calculated from the data and applied to the data directly, so there is no need for MC corrections.

3.2 MONTE CARLO SAMPLES

The analysis uses also Monte Carlo (MC) samples for the simulation of signal and the SM backgrounds. We consider centrally produced MC samples during the following campaigns:

- 2016: RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8_ext2-v1,
- 2017: RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1,
- 2018: RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1.

The EWK VBS signal is simulated at the leading order (LO) with six EW and zero QCD vertices with the MadGraph5_aMC@NLO 2.6.5 generator requiring a final state with two leptons coming from the Z decay, and two quarks from the V boson decay. The QCD mediated production of two gauge bosons, with two opposite sign same flavour leptons and two quarks is considered as a background process. The MC for this QCD background source is produced with the MadGraph5_aMC@NLO 2.6.5 generator at the LO. The complete list of the VBS EW and QCD MC samples and their cross sections can be found in Table 5.3.

The background samples considered are top quark pair production ($t\bar{t}$), single top quark production tW, the Drell-Yan (DY) process, W+jets, diboson and triboson production (VVV, VZ), and other minor backgrounds (like gluon-gluon fusion to WW, Vγ associated production, and vector boson fusion process VBF-V). The simulations for the $t\bar{t}$ and for the tW, are performed, respectively at next-to-next-to-the-leading order (NNLO) and at next-to-the-leading order (NLO) with POWHEG (2.0) [152], while the Drell–Yan (DY) and the W+jets samples are generated with MadGraph5 generator at LO accuracy with the MLM matching scheme [153]. The WW, WZ, and ZZ processes are produced with PYTHIA (v8.2) [154] and are normalized to NLO.

The parton distribution functions (PDFs) used to generate the simulated background samples for 2016 (2017 and 2018) is the NNPDF3.0 (NNPDF3.1 NNLO) [155, 156]. The generators used for signal and background processes are interfaced with PYTHIA 8.226 (8.230) to simulate parton showering and hadronization in 2016 (2017 and 2018).
All MC samples use the PYTHIA [157] program with the CUETP8M1 tune [158] for 2016 simulation, and CP5 tune [159] for 2017 and 2018 samples to describe parton showering and hadronization. Additional collisions in the same or adjacent

Process	Dataset name	cross section (pb)
VBS EW samples		
$WmTo2J\_ZTo2L$	/WminusTo2JZTo2LJJ_EWK_LO_SM*	0.0298
$WpTo2J\_ZTo2L$	/WplusTo2JZTo2LJJ_EWK_LO_SM*	0.0540
$ZTo2L\_ZTo2J$	/ZTo2LZTo2JJJ_EWK_LO_SM	0.0159
VBS QCD process		
$WpTo2J\_WmToLNu_QCD$	/WplusTo2JWminusToLNuJJ_QCD_LO_SM*	5.561
$WpToLNu\_WmTo2J_QCD$	/WplusToLNuWminusTo2JJJ_QCD_LO_SM*	5.544
$WmToLNu\_ZTo2J_QCD$	/WminusToLNuZTo2JJJ_QCD_LO_SM*	1.302
$WpToLNu\_ZTo2J_QCD$	/WplusToLNuZTo2JJJ_QCD_LO_SM*	2.162
$WmTo2J\_ZTo2L_QCD$	/WminusTo2JZTo2LJJ_QCD_LO_SM*	0.3862
$WmToLNu\_WmTo2J_QCD$	/WminusToLNuWminusTo2JJJ_QCD_LO_SM*	0.03774
$WpTo2J\_ZTo2L_QCD$	/WplusTo2JZTo2LJJ_QCD_LO_SM*	0.6409
$WpToLNu\_WpTo2J_QCD$	/WplusToLNuWplusTo2JJJ_QCD_LO_SM*	0.08642
$ZTo2L\_ZTo2J_QCD$	/ZTo2LZTo2JJJ_QCD_LO_SM*	0.3756

Table 5.3: EW VBS ZVjj semileptonic signals and backgrounds used in this analysis. 2016: *_MJJ100PTJ10_TuneCUETP8M1_13TeV-madgraph-pythia8, 2017, 2018: *_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8.

bunch crossings (pileup) are taken into account by superimposing simulated minimum bias interactions onto the hard scattering process, with a number distribution matching that observed in data. Simulated events are propagated through the full GEANT4 based simulation [160] of the CMS detector.
The list of samples for the background processes is reported in Tables 5.4 and 5.5. The listed names contain also the information about the MC generator that were used for the production of the corresponding processes.

3.3 MC CORRECTION FACTORS

MC simulated samples are further given the following event-by-event weights to better approximate the collision data in terms of kinematics and object identification efficiency:

- **Pileup reweighting**: the ratio between the normalized distributions of number of inelastic collisions for each year and the MC pileup profile is used as an event weight. The proton inelastic cross section value of (69.2 ± 4.6%) mb is used to derive the distribution of number of collisions. The pileup reweighting factor is a function of the true number of generated inelastic collisions in the event.
- **Lepton identification efficiency scale factor**: Data / MC scale factor of lepton identification efficiencies is applied to the event. Details of the scale factors can be found in the supporting CMS note AN-2019/125 [151]. Functions of the lepton momentum and pseudorapidity.
- **Heavy-flavor tagging efficiency scale factor**: For each jet in the event where the b-tagging discriminator is evaluated, a scale factor provided by the

Process	DAS name	Xsec. (pb)	Xsec. computed order
DY ($10 < m(\ell\ell) < 50$GeV)	/DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	18610	NLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	6077.22	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-70to100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	208.977	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	181.30	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	50.4177	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	6.98394	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	1.68141	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	0.775392	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	0.186222	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	0.00438495	NNLO
$t\bar{t}$ (semi-leptonic)	/TTToSemilepton_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8/	365.34	NNLO
$t\bar{t}$ (leptonic)	/TTTo2L2Nu_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8/	88.29	NNLO
ttW	/TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/	0.2043	NLO
ttW	/TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/	0.4062	NLO
ttZ	/ttZJets_13TeV_madgraphMLM-pythia8/	0.5407	NLO
Single-top, s-channel	/ST_s-channel_4f_leptonDecays_13TeV_PSweights-amcatnlo-pythia8/	3.36	NLO
Single-top, tW-channel	/ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1/	19.20	NNLO
Single-top, tW-channel	/ST_tW_top_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1/	19.20	NNLO
Single-top, t-channel	/ST_t-channel_antitop_4f_inclusiveDecays_13TeV_PSweights-powhegV2-madspin/	80.95	NLO
Single-top, t-channel	/ST_t-channel_top_4f_inclusiveDecays_13TeV_PSweights-powhegV2-madspin/	136.02	NLO
W+Jets	/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	61526.7	NNLO
W+Jets	/WJetsToLNu_HT-70To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	1637.1	NNLO
W+Jets	/WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	1627.45	NNLO
W+Jets	/WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	435.237	NNLO
W+Jets	/WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	59.1811	NNLO
W+Jets	/WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	14.5805	NNLO
W+Jets	/WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	6.65621	NNLO
W+Jets	/WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	1.60809	NNLO
W+Jets	/WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	0.0389136	NNLO
WW	/WW_TuneCUETP8M1_13TeV-pythia8/	118.7	NLO
WZ	/WZ_TuneCUETP8M1_13TeV-pythia8/	47.13	NLO
ZZ	/ZZ_TuneCUETP8M1_13TeV-pythia8/	16.523	NLO
WWW	/WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8/	0.2086	NLO
WWZ	/WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/	0.1651	NLO
WZZ	/WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/	0.05565	NLO
ZZZ	/ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/	0.01398	NLO

Table 5.4: The list of MC samples and corresponding cross-sections used in 2016 analysis.

Process	DAS name	Xsec. (pb)	Xsec. computed order
DY ($10 < m(\ell\ell) < 50$GeV)	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/	18610	NLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/	6077.22	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-70to100_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	208.977	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	181.30	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	50.4177	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	6.98394	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	1.68141	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	0.775392	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	0.186222	NNLO
DY ($m(\ell\ell) > 50$GeV)	/DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	0.00438495	NNLO
$t\bar{t}$ (semi-leptonic)	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/	365.34	NNLO
$t\bar{t}$ (leptonic)	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/	88.29	NNLO
W+Jets	/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/	61530	NNLO
ttW	/ttWJets_TuneCP5_13TeV_madgraphMLM_pythia8/	0.4611	NLO
ttZ	/ttZJets_TuneCP5_13TeV_madgraphMLM_pythia8/	0.5407	NLO
Single-top, s-channel	/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8/	3.36	NLO
Single-top, tW-channel	/ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/	19.20	NNLO
Single-top, tW-channel	/ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/	19.20	NNLO
Single-top, t-channel	/ST_t-channel_antitop_5f_TuneCP5_13TeV-powheg-pythia8/	80.95	NLO
Single-top, t-channel	/ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8/	136.02	NLO
W+Jets	/WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8/	1637.1	NNLO
W+Jets	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/	1627.45	NNLO
W+Jets	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/	435.237	NNLO
W+Jets	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/	59.1811	NNLO
W+Jets	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/	14.5805	NNLO
W+Jets	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/	6.65621	NNLO
W+Jets	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/	1.60809	NNLO
W+Jets	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/	0.0389136	NNLO
WW	/WW_TuneCP5_PSweights_13TeV-pythia8/	118.7	NLO
WZ	/WZ_TuneCP5_PSweights_13TeV-pythia8/	47.13	NLO
ZZ	/ZZ_TuneCP5_13TeV-pythia8/	16.523	NLO
WWW	/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/	0.2086	NLO
WWZ	/WWZ_TuneCP5_13TeV-amcatnlo-pythia8/	0.1651	NLO
WZZ	/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/	0.05565	NLO
ZZZ	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/	0.01398	NLO

Table 5.5: The list of MC samples and corresponding cross-sections used in 2017 and 2018 analysis.

B-tagging Physics Object Group (BTV POG) depending on the jet momentum, η, flavor, and the b-tagging discriminator value is applied to the event. The scale factors are designed to make the distribution of the discriminator value in the simulation close to that found in collision data.

- **Pileup jet ID scale factor**: preliminary scale factors computed by the Jet and Missing Transverse Energy Physics Object Group (JME POG) [161] for the pileup jet identification at the loose working point are applied to the jet. The total weight per event is a product of the scale factors for all jets with $p_T > 30$ GeV, $|\eta| < 4.7$ in the event.
- **L1 prefiring correction**: for MC simulations for the 2016 and 2017 analyses, additional trigger efficiency scale factors are considered as presented in the previous chapter, in order to compensate the inefficiency due to the L1T prefiring effect.
- **Top p_T reweighting**. During Run 1 and Run 2 of the LHC, the momentum spectrum of top quarks in $\bar{t}t$ data was significantly softer than those predicted by the various MC simulations, based on either LO or NLO matrix elements interfaced with parton showers. The latest NNLO (+ NLO EW) calculations integrate our most current theoretical knowledge of the SM $t\bar{t}$ production and can be used, in specific cases, to correct the top quarks p_T spectra in the $t\bar{t}$ MC. A dedicated top p_T reweighting function is also derived from the top enriched region and applied across the analysis while monitoring the agreement of other distributions as a result of this reweighting.

4 OBJECT RECONSTRUCTION IN VBS SEMILEPTONIC ANALYSES

Here we describe the reconstruction and identification of the objects employed in the analysis [3].

Electrons, muons, and jets are basic elements of the event reconstruction for the analysis, and below we briefly mention the collections we access for each object used in this search and how we identify them. In particular, it is worth to point out that the reconstruction and selection of leptons are optimized separately for each year of the LHC Run2 following the prescriptions of the POGs.

4.1 ELECTRONS

Several strategies are used in CMS to identify promptly isolated electrons present in the signal, and to separate them from background sources, mainly originating from photon conversions, jets misidentified as electrons, or electrons from semileptonic decays of b and c quarks. Simple and robust algorithms have been developed to apply sequential selections on a set of discriminants taken from calorimetric observables, isolation variables, tracking quality variables, and conversion rejection

3 The definition, reconstruction, efficiencies, and systematics of the objects used in the present analysis are shared with the CMS "Higgs to WW" Run-2 analysis group and are fully documented in Ref. [151]

variables. A set of cuts on various variables for achieving a particular signal efficiency is commonly referred to as "Working Point (WP)". Another set of selection comes from the multivariate (MVA) approach, where a single discriminator variable, on the basis of the Boosted Decision Tree (BDT) algorithm, is computed. In the present analysis, electrons are selected with this second method, i.e. by using the so-called "mva_90p_Iso2016" for 2016 samples ("mvaFall17V1Iso_WP90" for 2017 and 2018) [151].

4.2 MUONS

The muon selection used in the analysis relies on the "TightHWW" algorithm, based upon the recommendation from the Muon POG group for the Tight selection, with a series of extra cuts:

- $p_T > 10$ GeV
- $|\eta| < 2.4$
- $|d_{xy}| < 0.01$ cm for $p_T < 20$ GeV and $|d_{xy}| < 0.02$ cm for $p_T > 20$ GeV, where d_{xy} is the impact parameter w.r.t. the primary vertex,
- $|d_z| < 0.1$ cm, where d_z is the longitudinal distance of the tracker track w.r.t. the primary vertex.
- For the muon isolation criteria we are using the recommended Muon POG tight PF-based combined relative isolation with $\Delta\beta$ correction in a cone size of $\Delta R < 0.4$.

On top of these selections, to further reduce the fake leptons contribution. Finally, to eliminate a possible bias in the muon momentum originated from detector misalignment, software reconstruction, or uncertainties in the magnetic field, we apply the corrections developed by the Rochester group [162].

4.3 JETS

A jet is a collection of hadrons originating from a common parton. Jets are confined to a relatively small cone due to the large momentum magnitude of the parent particle and the relatively small transverse momentum associated with the fragmentation process. Jets considered in this analysis are reconstructed as AK4 PF jets with the Charged Hadron Subtraction (CHS) method, meaning that the jets are clustered from the particle flow candidates using the anti-kT algorithm with a distance parameter of $R = 0.4$.

Moreover, as we will see in the next section, hadronically decaying vector boson candidates are reconstructed using an anti-kT clustering algorithm with a distance of $R = 0.8$ and using the PUPPI algorithm [163]. The V-jets candidates are required to have $p_T > 200$ GeV and $|\eta| < 2.4$. The boosted jet mass is computed after employing the softdrop algorithm [164] and is required to be between 40 GeV and 250 GeV. A cut on N-subjettiness [165] variable $\tau_{21} < 0.45$ is used to identify jets from W or Z decays. Moreover, V-jets candidates are excluded if their distance from lepton candidates is $\Delta R < 0.8$. Anti-kT $R = 0.4$ jets within $\Delta R < 0.8$ from the boosted jet candidates are excluded from the event.

An important task for analyses characterized by two vector bosons in the final state is the suppression of the top quark background. The main processes belonging to this category are $t\bar{t}$ production, characterized by the presence of two b jets in the final state, and tW production, which is characterized by having only one top quark, decaying to a W and a b quark, produced together with a W boson. The usage of b-tagging algorithms to efficiently tag jets that are likely coming from B hadron decays is thus a fundamental tool to reject top background events. B tagging is a reconstruction technique that takes advantage of the relatively long lidetime and large mass of B hadrons and assigns to each jet a likelihood that it contains a B hadron. The b-tagging can be based on track information, secondary vertex information, soft lepton information, or some combination of the above. A variety of b-tagging algorithms based on different input information has been developed by CMS. A common feature is that all these algorithms produce an output discriminator that describes how likely it is that a given jet is a b jet, i.e. the higher is the discriminator value, the bigger is the probability that the jet is a real b jet. In this analysis, the loose WP of the DeepCSV ("Combined Secondary Vertex") algorithm [166] is used for b-tagging on AK4 jets.

5 EVENT TOPOLOGY AND ANALYSIS STRATEGY

In this section, we focus on the analysis phase space, presented in Figure 5.4. First, we describe the preselection on our samples, which is common for both control regions and signal region. Then, we highlight the details of the signal region.

5.1 PRESELECTION ON THE EVENTS

The VBS semileptonic selection aims to identify events with two leptons, and four quarks in the final state, which are detected as jets by the showering and hadronization processes and whose number depends also on the jet clustering and pairing algorithms used.
The selection for the lepton candidates in the event considers two isolated light leptons (either muons or electrons), with same flavor (opposite flavor for the top enriched region) and opposite charge and $p_T > 25(15)$ GeV, based on the trigger thresholds. The invariant mass of the leptonic system is also required to lie in a 15 GeV window centered around the nominal Z boson mass.

Figure 5.5 shows the categorization scheme of jets in the event. The two forward "tag-jets" (or "VBS-jets") from the initial state radiation are identified by applying a selection criteria based on the invariant mass of the objects. They are selected as the pair of AK4 jets with the highest invariant mass in the event, with the additional cuts $m_{jj} > 200$ GeV, and $|\Delta\eta_{jj}| > 2.0$.

To maximize the acceptance to the V-boson decay products, we further define two orthogonal categories for the remaining jets in the event, the "boosted" and

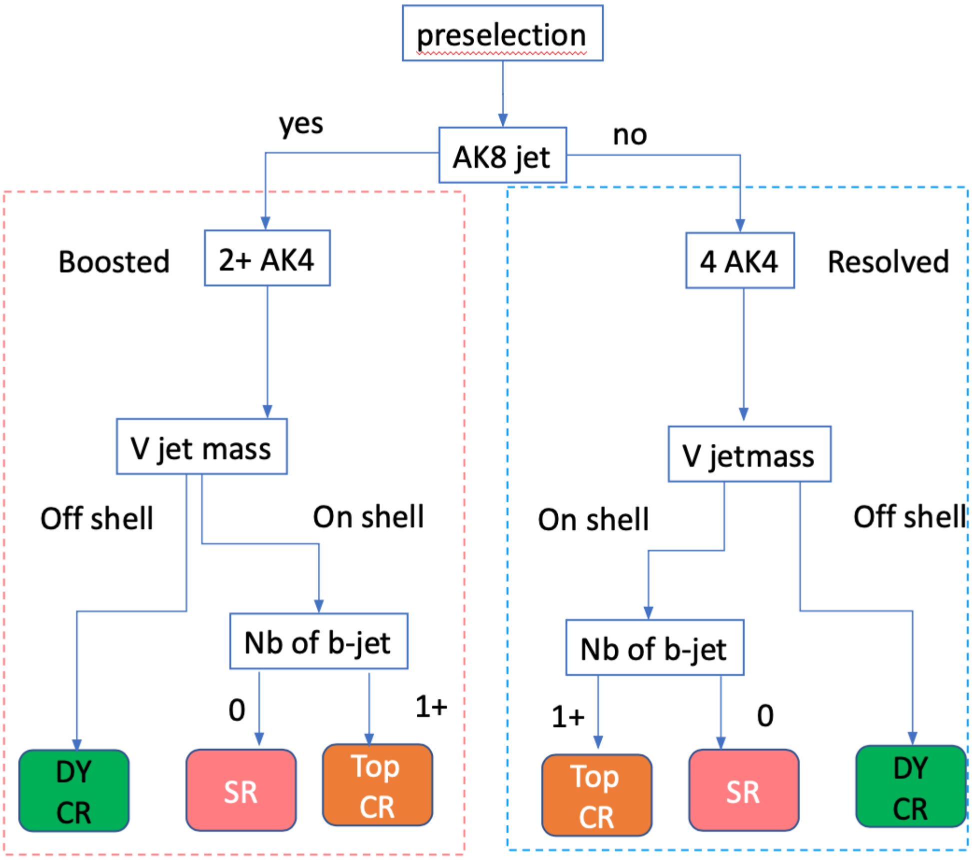


Figure 5.4: Scheme of the VBS ZVjj analysis phase space.

the "resolved", depending on the specific decay topology. Events with two VBS-jets and at least one-large radius jet (AK8), i.e. with a distance parameter $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.8$, fall within the boosted category. This corresponds to the case in which the decay products of the V-boson are collimated in the laboratory system and the resulting jets are merged into a single larger jet (AK8) with mass compatible with the W or Z bosons. When the event category is instead "resolved", there are no AK8 jets and at least two more AK4 jets. In case of more than two V-jets candidates, the pair with the invariant mass closest to the W or Z mass is selected. In both cases, we can define a V-jet mass: either identified with the softdrop mass for the boosted topology, or with the invariant mass of the two AK4 jets. To summarize, the complete list of preselection requirements for the present analysis is listed in Table 5.6.

Once we have defined the common baseline for all our samples, we use the mass of the V-boson to define the signal region and the background enriched regions, that are presented in detail in the next Section 4.5. Since we expect the V-boson in the VBS signal to be generated on-shell, we select our signal in $65 < m_V < 105$ GeV. This condition is instead inverted for the DY+jets region, which is responsible of the major background contamination.

We finally conclude with a remark on the choice of the two specific algorithms on the VBS/V-jets pairings, which are motivated by studies on parton-jet matching presented in Ref. [167], where the efficiency of several categorization algorithms was

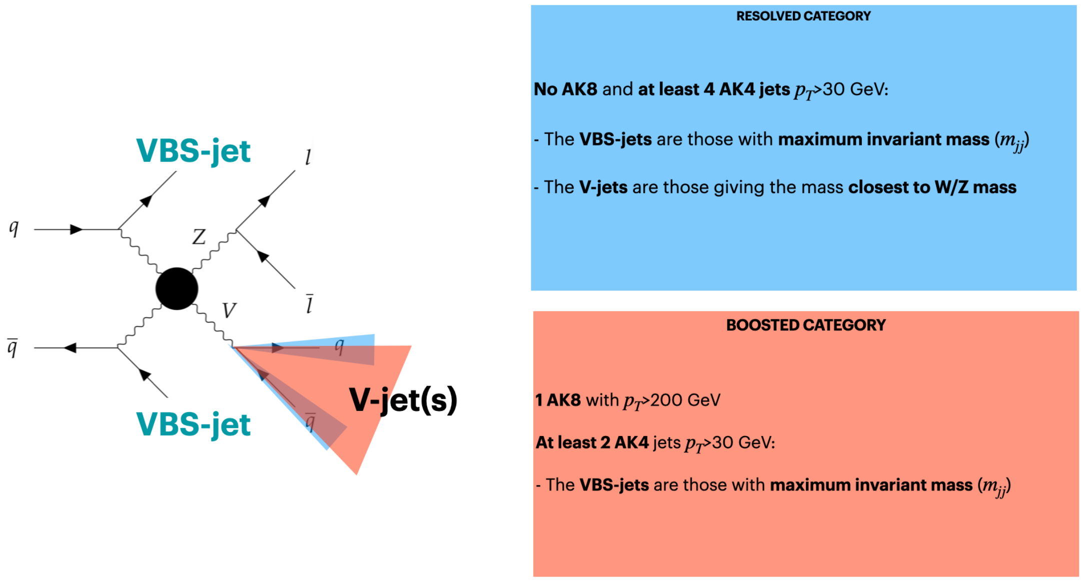


Figure 5.5: The Feynman diagram of the VBS process with an hadronically decaying gauge boson V, a Z. Dark blobs in the vertex represent the possible inclusion of anomalous couplings contribution. The blue and red boxes on the right highlights the jets categorization in the search.

compared. For the sake of completeness, we report in Table 5.7 a summary on the performances of the various jets pairings, to underline the goodness of our choice.

5.2 CUT-BASED ANALYSIS STRATEGY

In this part, we present a possible strategy for the statistical analysis that takes advantage of a series of cuts performed on the main kinematical variables on top of the minimal signal region outlined in the previous Section 5.5.

This tighter selection will be employed in Section 5.9 for the statistical analysis in the extraction of the signal significance for the VBS ZVjj process, and will be compared with a multivariate-analysis approach, presented in the Section 5.6.

For the identification of a series of cuts, we present in Figure 5.6 the comparison of the shape between signal and backgrounds for several kinematical distributions in the signal region with pre-selection cuts. The tighter signal region select events according to the following requirements:

- transverse momentum p_T (sub)leading > 40 (20) GeV,
- invariant mass of the "VBS jets" $m_{jj} > 250$ GeV. This cuts reduces the contamination in the SR of the major background due to DY process,
- (resolved category) leading V-jet momentum $p_T j_1 > 80$ GeV.

In Figure 5.7 we show a comparison between the $\Delta\eta_{jj}$ distributions for the signal region and the signal region with the "tighter" selection. The rejection of the major background source, the DY+jets, is noticeably improved.

Object	Variables	Selection
leptons	N.of Leptons	== 2
	m_{ll}	$[76; 106]$ GeV
	$p_T(\ell_1)$	> 25 GeV
	$p(\ell_2)$	> 15 GeV
VBS jets	$p_T j_{1,2}$	> 30 GeV
	m_{jj}	> 200GeV
	$\|\eta_j\|$	< 5
	$\Delta\eta_{jj}$	> 2.0
V-jet (boosted)	$p_T J$	> 200 GeV
	$\|\eta J\|$	< 2.5
V-jets (resolved)	$p_T j$	> 30 GeV
	$\|\eta j\|$	< 2.5

Table 5.6: Summary of the pre-selection used in the present analysis for the leptons and the VBS jets. A veto on AK4 jets with $p_T < 30$ GeV is also applied to each event.

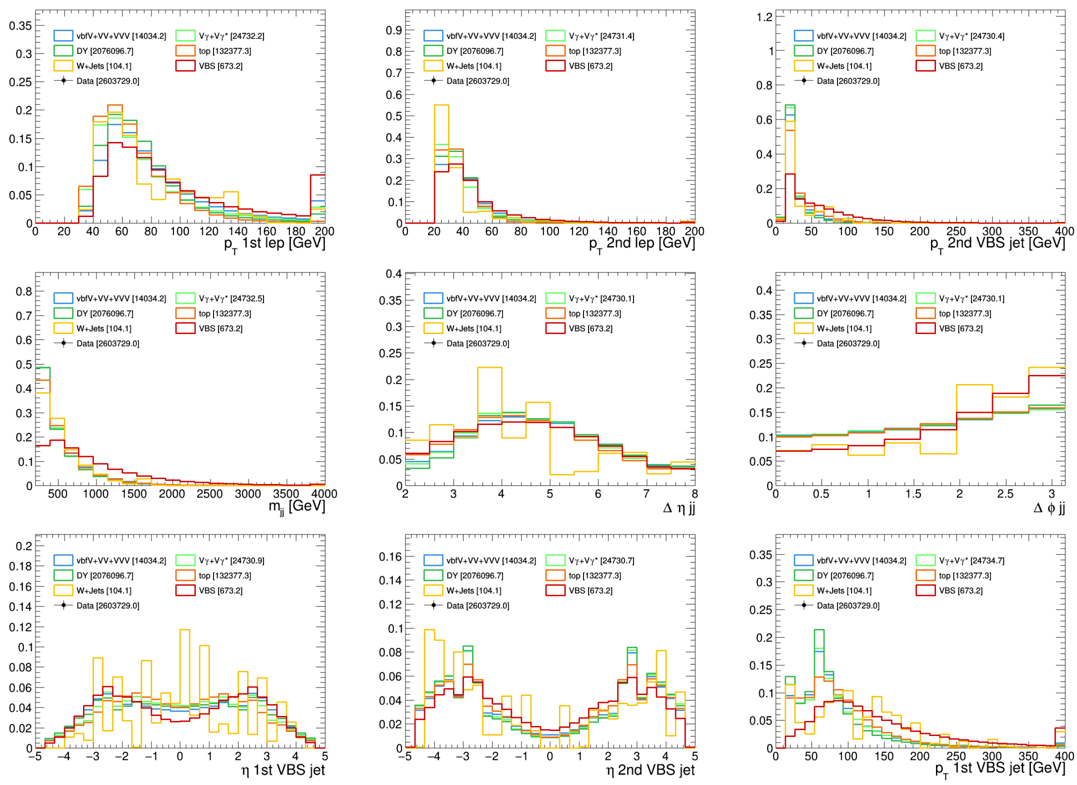


Figure 5.6: Shape comparison of the major backgrounds and signal process (red line) for the final state for 2018 luminosity.

VBS Selection	Algorithm	Efficiency
Max m_{jj}	$\max_{i,j}\left(\sqrt{p^{\mu}p_{\mu}}\right)$	$64,4\%$
Max $\Delta\eta$	$\max_{i,j}\left(\lvert\eta_i-\eta_j\rvert\right)$	$49,5\%$
Max ΔR	$\max_{i,j}\left(\sqrt{(\eta_i-\eta_j)^2+(\phi_i-\phi_j)^2}\right)$	$47,9\%$

V Selection	Algorithm	Efficiency
Min $\Delta\eta$ / Nearest W	$\min_{i,j}\left(\frac{\lvert 80-\sqrt{p^{\mu}p_{\mu}}\rvert}{80}+\lvert\eta_i-\eta_j\rvert\right)$	$76,7\%$
Nearest W	$\min_{i,j}\left(\lvert 80-\sqrt{p^{\mu}p_{\mu}}\rvert\right)$	$76,5\%$
Min $\Delta\eta$/ Nearest W or Z	$\min_{i,j}\left(\min\Delta\eta+\min_{i,j}\left(\frac{\lvert 80-\sqrt{p^{\mu}p_{\mu}}\rvert}{80},\frac{\lvert 91-\sqrt{p^{\mu}p_{\mu}}\rvert}{91}\right)\right)$	$76,3\%$
Nearest W or Z	$\min_{i,j}($ Nearest W, Nearest Z$)$	$73,7\%$
Nearest Z	$\min_{i,j}\left(\lvert 91-\sqrt{p^{\mu}p_{\mu}}\rvert\right)$	$65,6\%$
Min $\Delta\eta$	$\min_{i,j}\left(\lvert\eta_i-\eta_j\rvert\right)$	$50,2\%$
Min ΔR	$\min_{i,j}\left(\sqrt{(\eta_i-\eta_j)^2+(\phi_i-\phi_j)^2}\right)$	$44,6\%$
Max p_T	$\max_{i,j}\left(\sqrt{p_x^2+p_y^2}\right)$	$14,3\%$

Table 5.7: Summary of the Gen-Level studies on the efficiency of several selection algorithms for V-jets and VBS-jets selection presented in Ref. [167]. The efficiency is the ratio between number of correct parton-jet matching normalized to the total number of events in the signal sample.

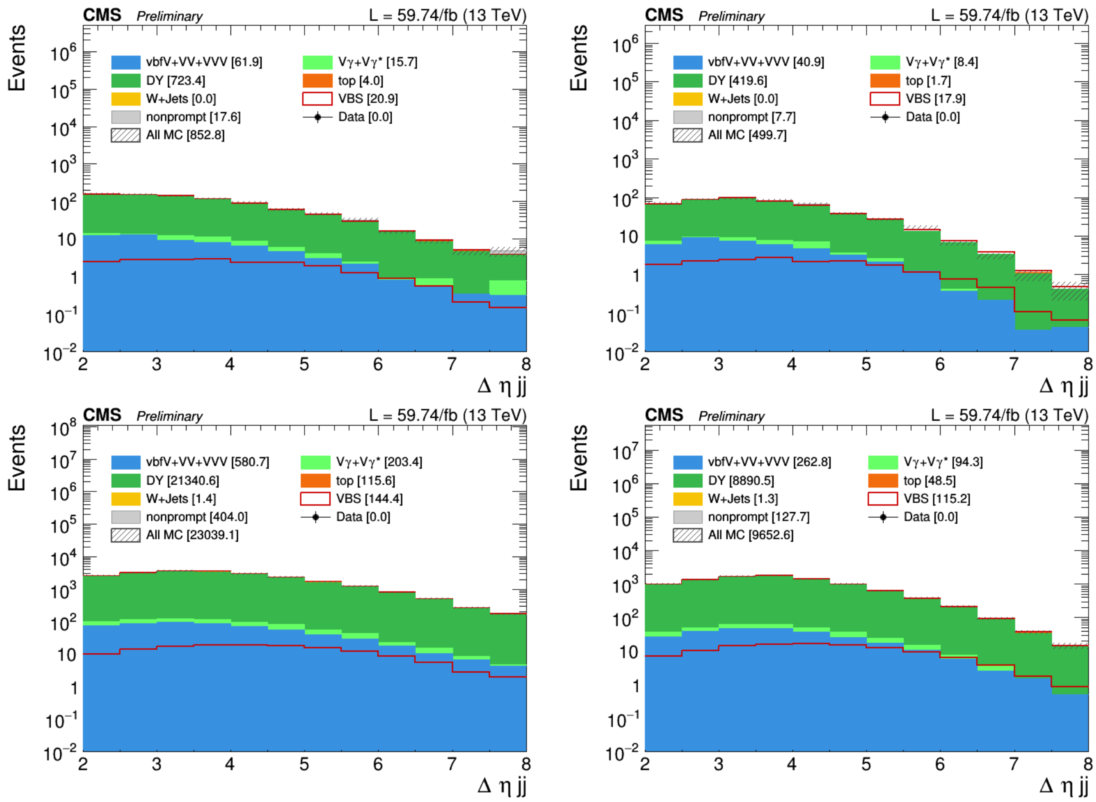


Figure 5.7: Distributions of pseudorapidity gap between the VBS-jets, for the MC VBS ZVjj signal and the backgrounds in the signal region (left) and in the "tight" signal region (right). On the first line: the boosted topology. On the second line: the resolved topology.

6 STUDIES ON A MULTIVARIATE APPROACH

In this Section, we analyze a possible strategy for the extraction of this rare signal process, taking advantage of the multivariate analysis strategy.

As we have seen in Sec. 5.5.2, some of the variables show differences in their distribution between signal and backgrounds, but none can discriminate enough the VBS electroweak from irreducible backgrounds without losing too many signal events. A possible way to improve the statistical significance of the signal in our analysis is to implement a multivariate analysis (MVA) selection. The best performance is achieved by choosing the better of the two approaches: Boosted Decisions Trees (BDT) and feed-forward deep neural network (DNN).

For both methods, the strategy is the same: a single model is trained for each category (Boosted or Resolved) to discriminate the signal against all backgrounds at the same time, providing a binary output as a floating-point number in [0,1] range. This is achieved by combining all simulation samples into a single dataset containing the information on the cross-section, selection, and scale factor coefficients applied as a weight on a per-event basis. The training is performed in a slightly looser Signal Region from 5.5.1. We keep the phase-space, defining cuts but keep the full range for the VBS dijet mass and η separation.

For the models to learn more efficiently, the events are then reweighted to avoid unbalance between signal and background events. Therefore, a scaling of the weights is applied so that the sum of weights for signal events equals the sum of weights for all backgrounds, keeping the ratio between backgrounds equal. This is done so that the models do not learn to discriminate against minor backgrounds while losing discriminating power against main backgrounds. The input variables are also standardized to avoid biases induced by the difference of scale between variables.

The dataset is split into training and testing (or validating) parts on an 80%-20% basis. The testing dataset is used to monitor the models' training and make sure overtraining is kept to the minimum. To be able to use all events in the final fit, we also perform 5-fold cross-validation on the training dataset for optimization (splitting the data into 5 subsets and using iteratively one of them as validation) and performing a Kolmogorov-Smirnov (KS) test on the models' output for training and testing sets. A model failing this test is discarded.

We assess the performances of models trained on only the 2018 dataset of the full Run 2. The 2018 dataset was chosen for the first studies, as it contains the largest amount of events. The number of events are reported in Table 5.8

Table 5.8: Number of events in Boosted and Resolved 2018 Signal Region datasets.

Total number of events	Background events	Signal events
Resolved topology		
385812	343790	42022
Boosted topology		
55809	50024	5785

6.1 BOOSTED DECISIONS TREE

The XGBoost library [168] is chosen for the BDT approach. Two different models were trained for Resolved and Boosted categories. The input variables for each category are shown in Table 5.9.

Table 5.9: List of the inputs variables used for Resolved and Boosted categories models, respectively.

Input variable	Description
Resolved	
lepton pt 1	leading lepton transverse momentum
lepton pt 2	subleading lepton transverse momentum
lepton η 1	leading lepton pseudorapidity
lepton η 2	subleading lepton pseudorapidity
Zlepton 1	Zeppenfeld $Z_{lep1} = \frac{\eta_{lep1} - \bar{\eta}}{\Delta\eta_{VBS}}$
Zlepton 2	Zeppenfeld $Z_{lep2} = \frac{\eta_{lep2} - \bar{\eta}}{\Delta\eta_{VBS}}$
VBS jet pt1	leading VBS jet transverse momentum
VBS jet pt2	subleading VBS jet transverse momentum
VBS jet η1	leading VBS jet pseudorapidity
VBS jet η2	subleading VBS jet pseudorapidity
V jet pt1	leading V jet transverse momentum
V jet pt2	subleading V jet transverse momentum
V jet η1	leading V jet pseudorapidity
V jet η2	subleading V jet pseudorapidity
V jet mass	reconstructed V jet invariant mass
m_{jj}	VBS dijet invariant mass
$\Delta\eta_{jj}$	VBS dijet η separation
Boosted	
lepton pt 1	leading lepton transverse momentum
lepton pt 2	subleading lepton transverse momentum
lepton η 1	leading lepton pseudorapidity
lepton η 2	subleading lepton pseudorapidity
Zlepton 1	Zeppenfeld $Z_{lep1} = \frac{\eta_{lep1} - \bar{\eta}}{\Delta\eta_{VBS}}$
Zlepton 2	Zeppenfeld $Z_{lep2} = \frac{\eta_{lep2} - \bar{\eta}}{\Delta\eta_{VBS}}$
VBS jet pt1	leading VBS jet transverse momentum
VBS jet pt2	subleading VBS jet transverse momentum
VBS jet η1	leading VBS jet pseudorapidity
VBS jet η2	subleading VBS jet pseudorapidity
Fat jet pt	AK8 jet momentum
Fat jet η	AK8 jet pseudorapidity
m_{jj}	VBS dijet invariant mass
$\Delta\eta_{jj}$	VBS dijet η separation

Due to the task being a binary classification, the objective function used is the binary logistic. To be sure to contain overtraining, the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve and the "log loss" were monitored during training. An early stopping that cuts training when the AUC doesn't improve for five consecutive steps was also implemented. L1 and L2 regularizations are also applied.

The maximal tree depth, learning rate, and both regularizations parameters are optimized with a grid search and cross-validation. The optimal values are shown in Tab. 5.10.

Table 5.10: Optimized values for BDT hyperparameters.

Parameter	Optimization range	Optimum (Resolved)	Optimum (Boosted)
Max tree depth	4-10	4	3
Learning rate	[0.01,0.05,0.1,0.2,0.3]	0.1	0.1
Subsample ratio of the training instances	not optimized	0.8	0.8
Subsample ratio of columns	not optimized	0.8	0.8
l1 regularization	[0.01, 0.1,1,10]	10	10
l2 regularization	[0.01, 0.1,1,10]	10	10

6.1.1 *BDT training monitoring*

To monitor the training of the BDT, and to reject overfitted models, the dataset is divided into two subsets, one used for the training and optimization of the model, called *training dataset*, and one used to control whether the model generalizes well on unseen data called *testing dataset*. Moreover, the optimization of the BDT uses five-fold cross-validation. That means that we further split the training dataset into five subsets, and then performing the training five times, each one using one of the subsets to validate the training performed on the other four. This method allows using the full training dataset for validation.

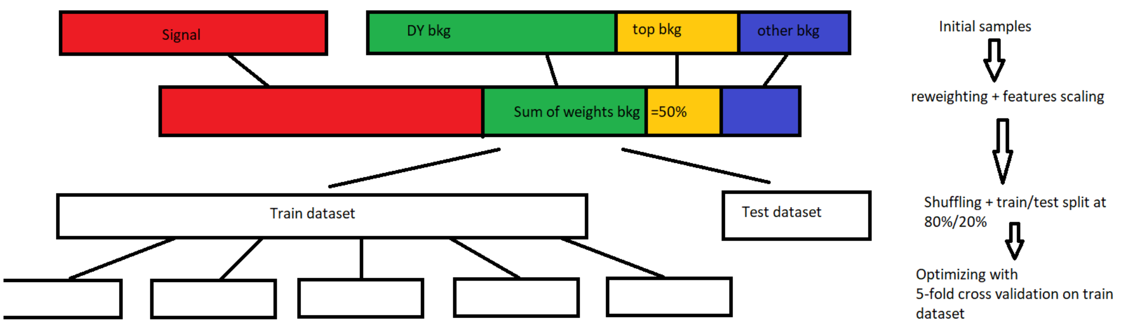


Figure 5.8: The workflow for samples division into subsets for training and testing.

A set of metrics is chosen to see the behavior of the models on the dataset. The first one is the logarithm of the loss ("log loss") function of the model. If the log loss of the training dataset continues to decrease while the validation is stable or increases, it is a clear sign of overtraining. The ROC AUC is also used. The ROC curve plots the number of background events correctly labeled as a function of the number of signal events correctly labeled for different values of thresholds on the BDT outputs. The AUC then measures the area under this curve. A better AUC points to better performances, but if the training and testing AUC starts to diverge it is a sign of overtraining. These metrics are showcased in Fig.5.9.

To optimize the BDT's parameters, a scoring function was defined on the cross-validation results as the test AUC penalized by the absolute value of the difference between test and train AUC. The α coefficient is set to 1.

$$[AUC_{test} - \alpha \times |AUC_{test} - AUC_{train}|] \tag{5.2}$$

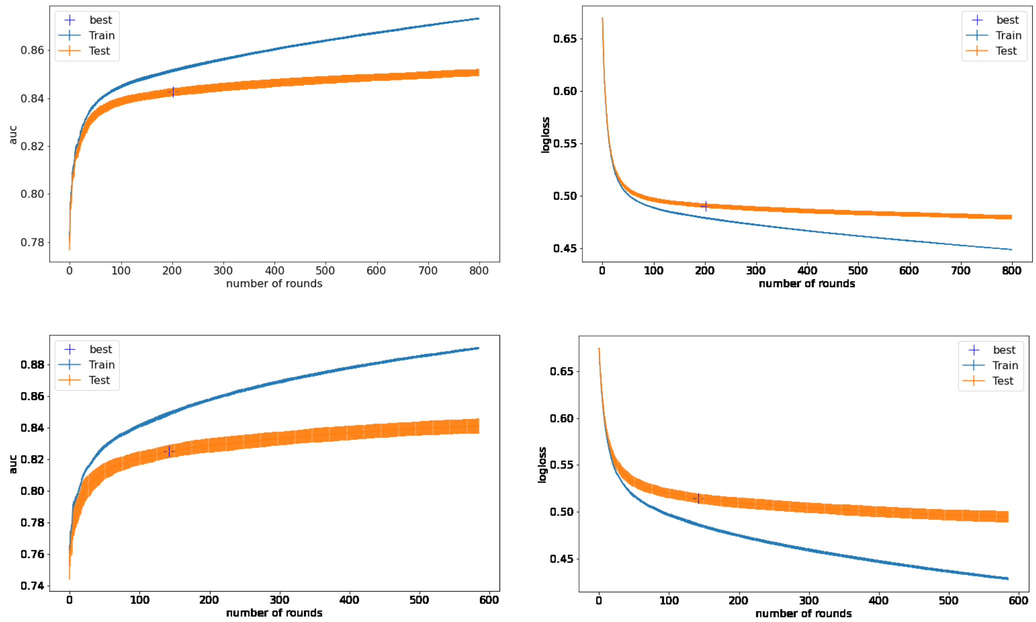


Figure 5.9: Evolution of the metrics during cross validation fro Resolved (top row) and Boosted (bottom row) categories. Left: the Area Under the ROC Curve. Right: the logloss. The "Best" point correspond to the optimum of the the scoring function.

Once the best parameters and number of rounds are obtained, the final training is performed on the whole training subset. During this training, a Kolmogorov-Smirnov test measuring the similarity between the BDT output on training and testing dataset for signal and backgrounds is performed. The model is kept if the KS test p-value is over 5% at the end of the training. If the p-value is decreasing a lot, then that is another hint of overtraining.

6.1.2 *BDT Results*

In Fig. 5.10 are presented the performances obtained with the final training for the BDT for both categories. To evaluate the performances we use the AUC, where higher values are indicating better discriminating power. The histogram showing the BDT output for the signal and backgrounds is also shown, as well as the KS test p-values. For the BDT output, the weights are the ones used for training, not the real ones, and serves only to show the different populations and not expected yield.

6.2 DEEP NEURAL NETWORKS

The second class of classifier tested in this analysis is a Multi-Layered Perceptron (MLP) or fully connected deep neural network (DNN). The framework selected is Keras [169]. The data set is split in the same ways as 5.6.1 with 80% of the samples used for training and 20% for testing the network. Once again, a model is trained for the Boosted category and another for Resolved.

6.2.1 *DNN architecture*

The architecture selected is a dense neural network, using binary cross-entropy as a loss function and Adam as the optimizer. Adam is a gradient descend optimizer that dynamically updates the learning rate on a per-parameter basis. The layers are built with Rectified LinearUnit (ReLU) activation function, l1, and l2 kernel regularization. A dropout function is added (forgetting a given proportion of the weights at every epoch) to avoid overtraining. The output layer is a single neuron activated by a sigmoid function to get an output in the range [0,1]. The size and number of hidden layers were optimized using Bayesian optimization, grid-search being ineffective for several parameters this big. The optimized values and range explored for each parameter are shown in Tab.5.11. The optimum is defined as the training maximizing the scoring function defined in Eq. 5.2.

Table 5.11: Optimized values for DNN hyperparameters.

Parameter	Optimization range	Optimum Resolved	Optimum Boosted
Number of hidden layer	1-5	2	2
Neurons per hidden layer	20-200	165	170
dropout rate	0-50%	20%	20%
starting learning rate	$1e^{-5}$ - $1e^{-1}$	$1e^{-4}$	$1e^{-4}$
l1 kernel regularization	0-$1e^{-2}$	0	0
l2 kernel regularization	0-$1e^{-2}$	$1e^{-3}$	$1e^{-3}$
Max number of epochs	Not optimized	400	400
Batch size	Not optimized	512	512
Early Stopping patience	Not optimized	10	10

6.2.2 *DNN Training*

To monitor the model training and to be sure to control overtraining, the loss and AUC are monitored at each epoch for both training and testing samples. The results are shown in Fig. 5.11.

Once the training is over, the mode's performances are assessed by looking at the DNN output, KS test, and ROC curves. The results are presented in Fig. 5.12.

6.3 BDT AND DNN COMPARISON

Once the DNN and BDT models have been fully trained, their performances are compared by the way of the Area Under the Receiver Operator Characteristic curve (AUC). The values obtained are shown in Tab. 5.12.

Category	BDT	DNN
2018 Resolved SR (test)	83.8%	84.2%
2018 Resolved SR (train)	85.0%	84.9%
2018 Boosted SR (test)	82.2%	81.8%
2018 Boosted SR (train)	84.8%	84.1%

Table 5.12: AUC values for BDT and DNN

Performances are similar for both multivariate methods, but more optimization has to be realized, particularly on the DNN side.

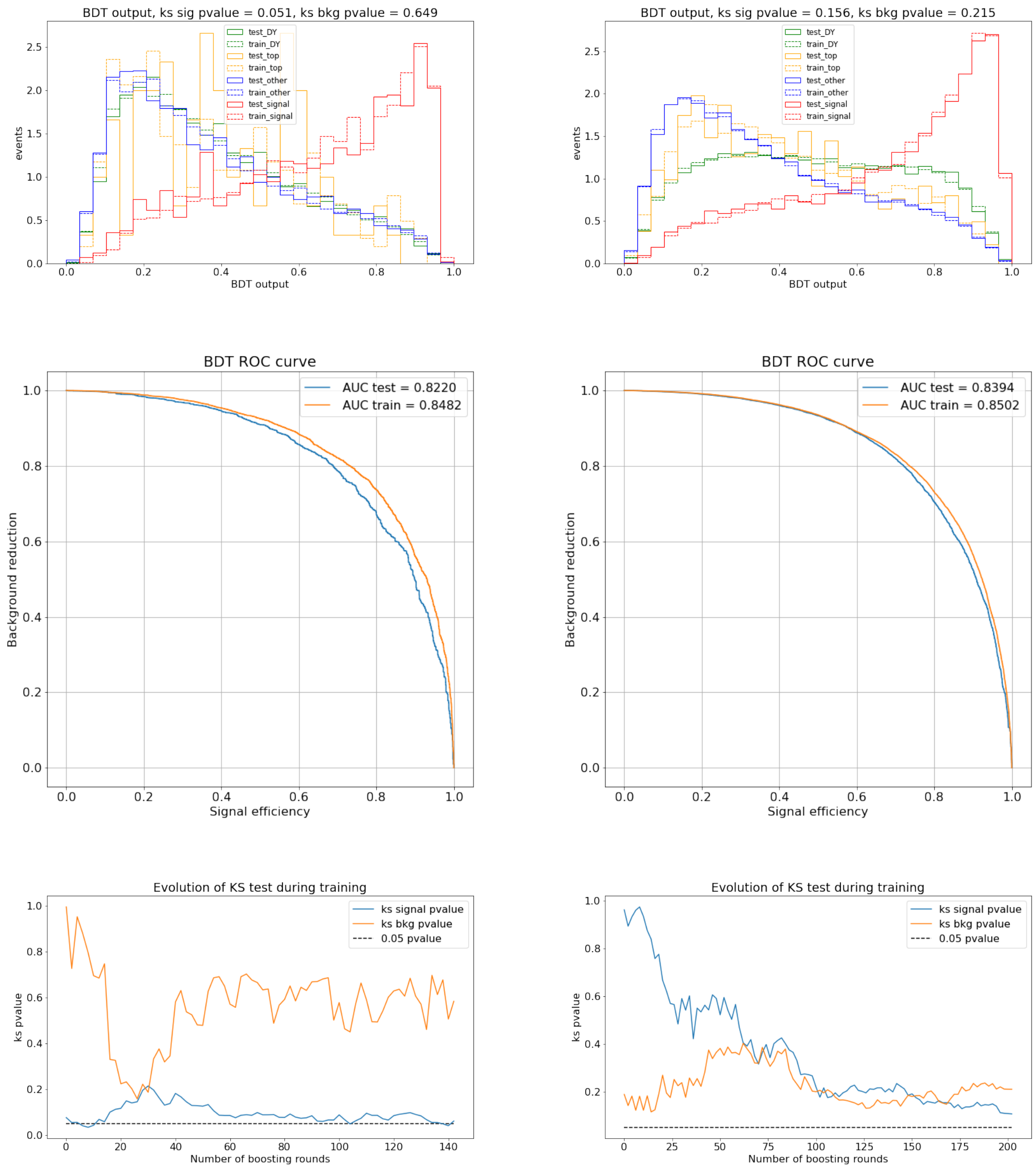


Figure 5.10: Models performances for Boosted (left) and Resolved (right) categories. Top: The BDT output for signal and backgrounds on the testing set (full lines) and training set (dashed). Middle: The ROC curves and AUC for training (orange) and testing (blue) dataset. Bottom: the evolution of the KS test p-value during training for signal (blue) and backgrounds (orange).

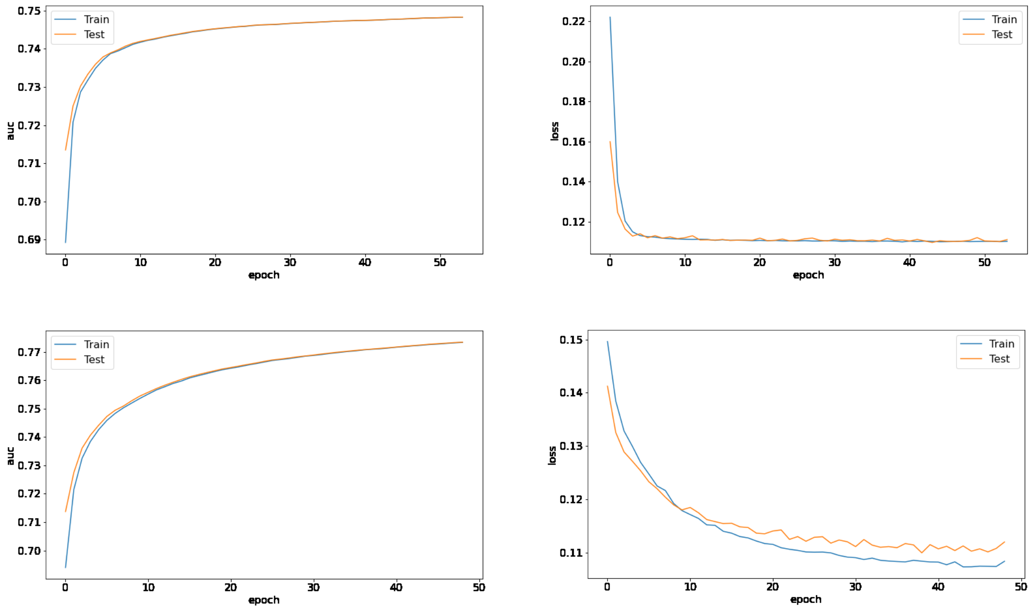


Figure 5.11: Training metrics for Resolved (top) and Boosted (bottom) categories. Left: The AUC. Right: the log loss. No divergence between training curves (blue) and testing curves (orange) hint to no overtraining.

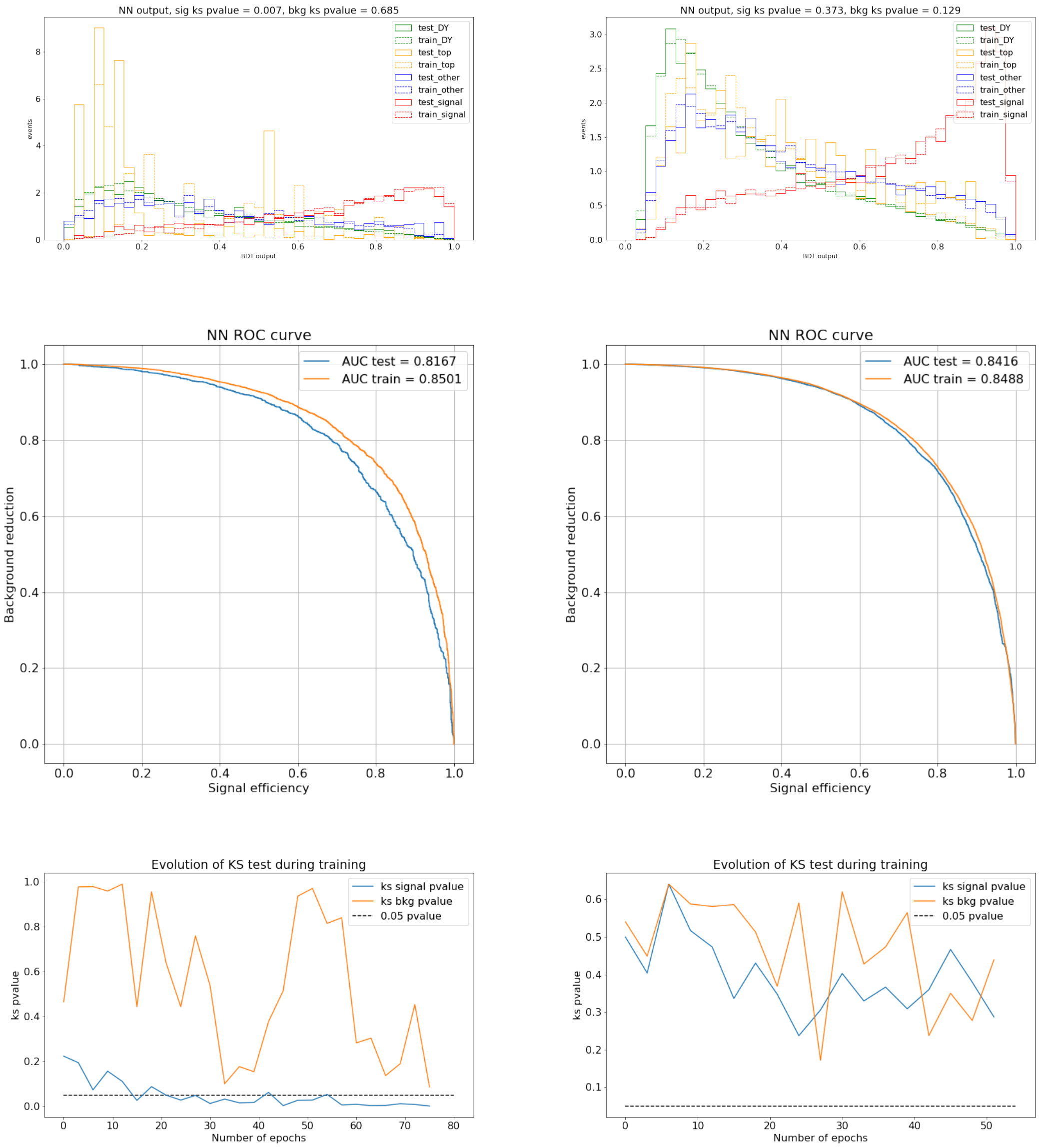


Figure 5.12: Models performances for Boosted (left) and Resolved (right) categories. Top: The DNN output for signal and backgrounds on testing set (full lines) and training set (dashed). Middle: The ROC curves and AUC for training (orange) and testing (blue) dataset. Bottom: the evolution of the KS test p-value during training for signal (blue) and backgrounds (orange).

7 BACKGROUND ESTIMATION

In this section, the agreement between data and MC is evaluated in two control regions for the two major background contributions (DY+jets and top processes) defined by the V-jet(s) mass, the flavor of the two leptons, and the requirements on the b tagging. Data are compared to background estimates year by year, showing a good agreement in shape and normalization. The definition of the two control regions and the signal region is reported in Figure 5.13. At the end of this section, we present the non-prompt background estimation procedure.

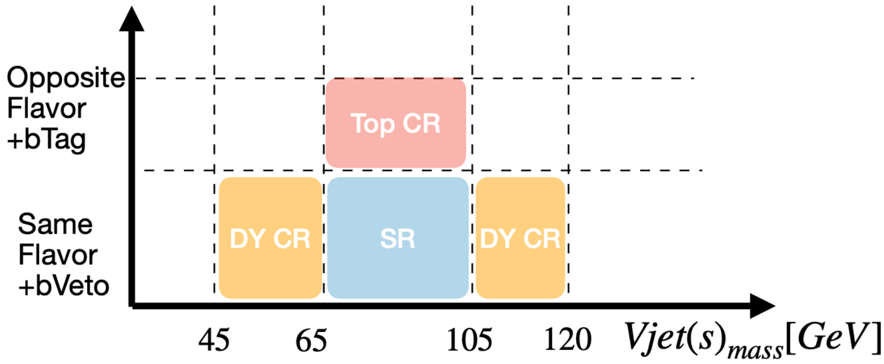


Figure 5.13: A pictorial representation of the variable used for the definition of the signal and background enriched regions.

7.1 DY+JETS BACKGROUND

The most significant background is due to irreducible DY+jets production. The phase space for checking the agreement of our MC predictions with the data is defined with the same cuts of the SR, but reverting the selection on the V-jet(s) mass, i.e. requiring $m_V < 65$ GeV or $m_V > 105$ GeV. Figures 5.14, 5.15 and 5.16 show the relevant variables for the analysis in data and MC spectrum for the 2016, 2017 and 2018 datasets.

7.2 TOP BACKGROUND

The expected contribution of $t\bar{t}$ and single-top production in the signal region is estimated with MC simulations.

All these processes are flavor-symmetric and have a branching ratio to a pair of leptons of a different flavor, $e\mu$, twice as large as the branching ratio to ee. That means the $e\mu$ mass spectrum provides an excellent probe to validate the Monte Carlo predictions for these backgrounds. Therefore, we define a top enriched control region by requiring pre-selected events (Table 5.6) opposite flavor in the two leptons, a loose b-tagging of the jets, and in the region $65 < m_V < 105$ GeV.

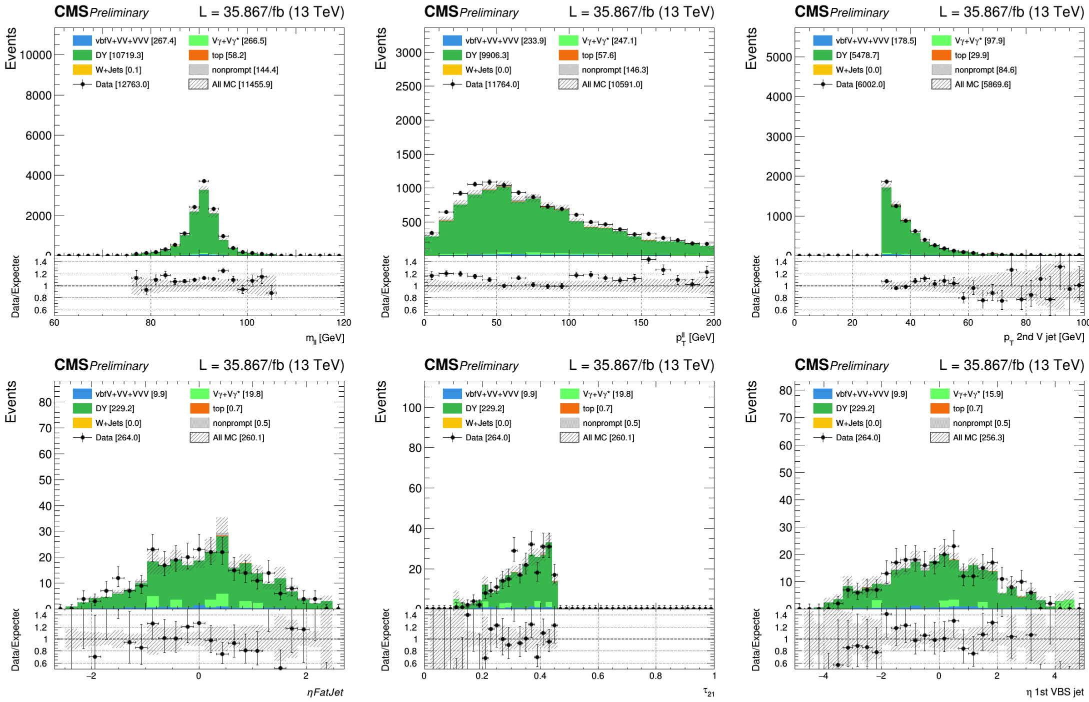


Figure 5.14: Comparison of data and MC estimate for the DY enriched control region for 2016 in the resolved (boosted) category on top (bottom) line for four chosen kinematical variables relevant in the analysis.

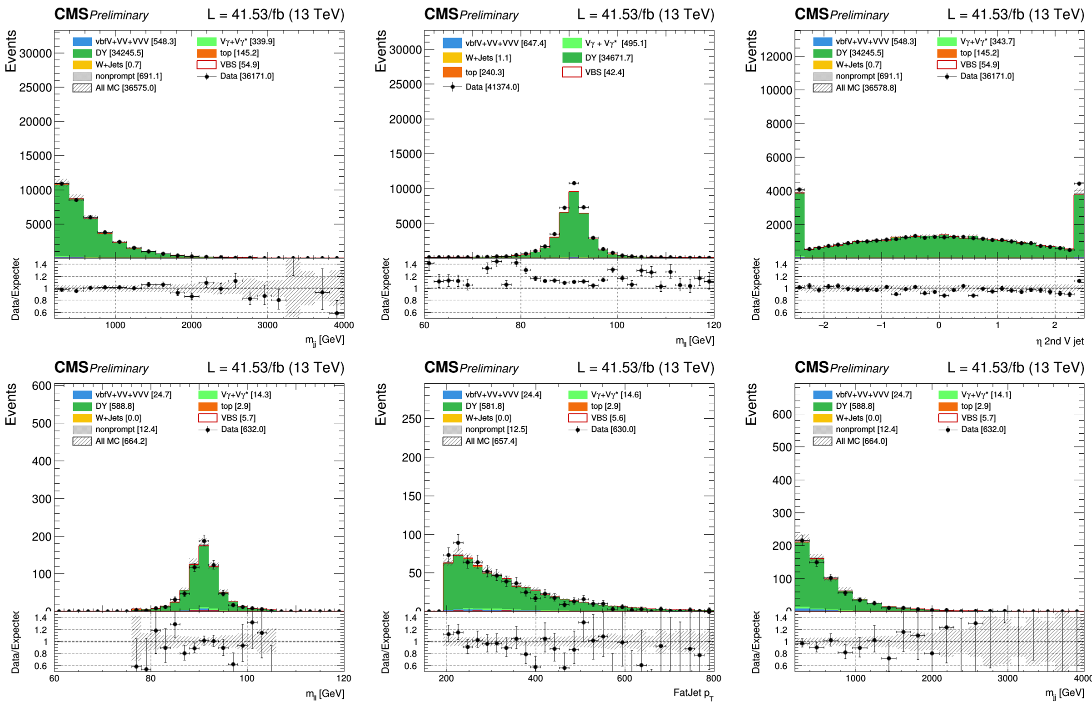


Figure 5.15: Comparison of data and MC estimate for the DY enriched control region for 2017 in the resolved (boosted) category on top (bottom) line for four chosen kinematical variables relevant in the analysis.

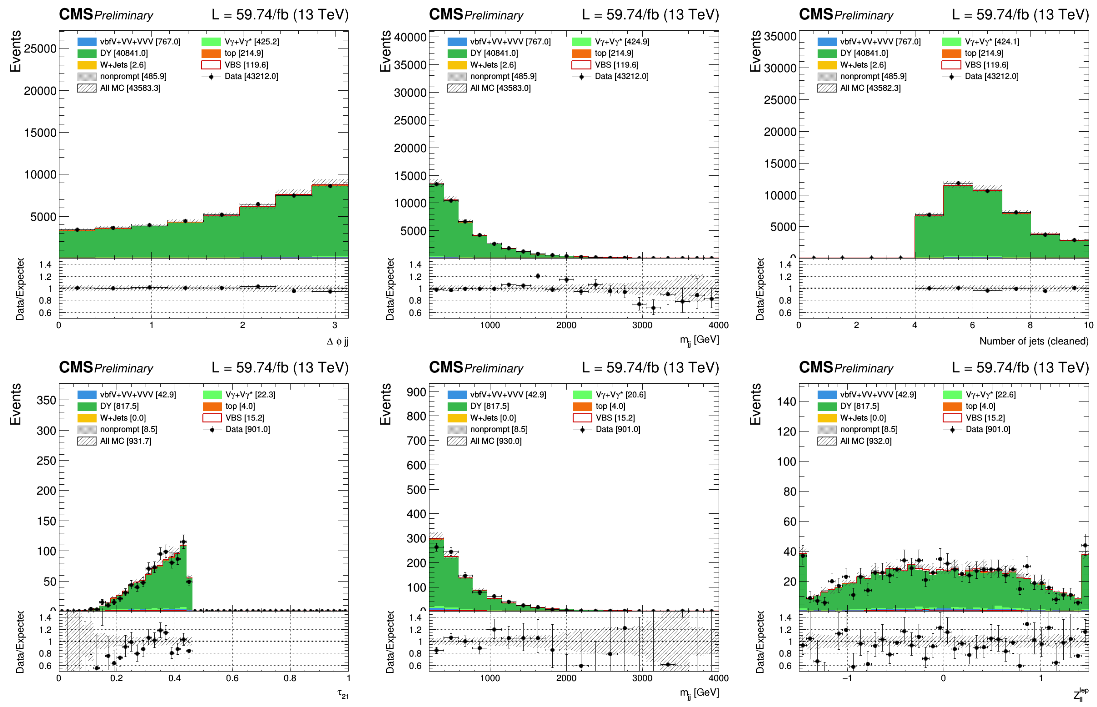


Figure 5.16: Comparison of data and MC estimate for the DY enriched control region for 2018 in the resolved (boosted) category on top (bottom) line for four chosen kinematical variables relevant in the analysis.

Figures 5.17, 5.18 and 5.19 shows the relevant variables for the analysis in data and MC spectrum for the 2016, 2017 and 2018 datasets. After having checked that the simulations describe the sample of $e\mu$ events well, the contribution of these backgrounds is estimated using the MC simulated sample with its cross-section, adopting the full set of cuts of the signal region, only changing the flavor of the leptons and b-tagging.

7.3 NON-PROMPT BACKGROUND

Non-prompt background derives from the incorrect identification of pure QCD interactions as prompt leptons. The contribution is almost negligible, as in the HCN search presented in the previous chapter. The data-driven estimation is based on the so-called "fakeable object" technique and provides a measurement of the yield and the kinematic distributions of the non-prompt background. The control sample is defined using a looser lepton definition, to get a higher misidentification rate. An extrapolation factor then relates background misidentified with this criteria, to background misidentified as passing the full particle selection of the signal region. Lepton fake (and prompt) rates are estimated using a data-driven method, as a function of the lepton (p_T, η) in a single lepton triggered sample (the details of the technique are documented in Ref. [151]).

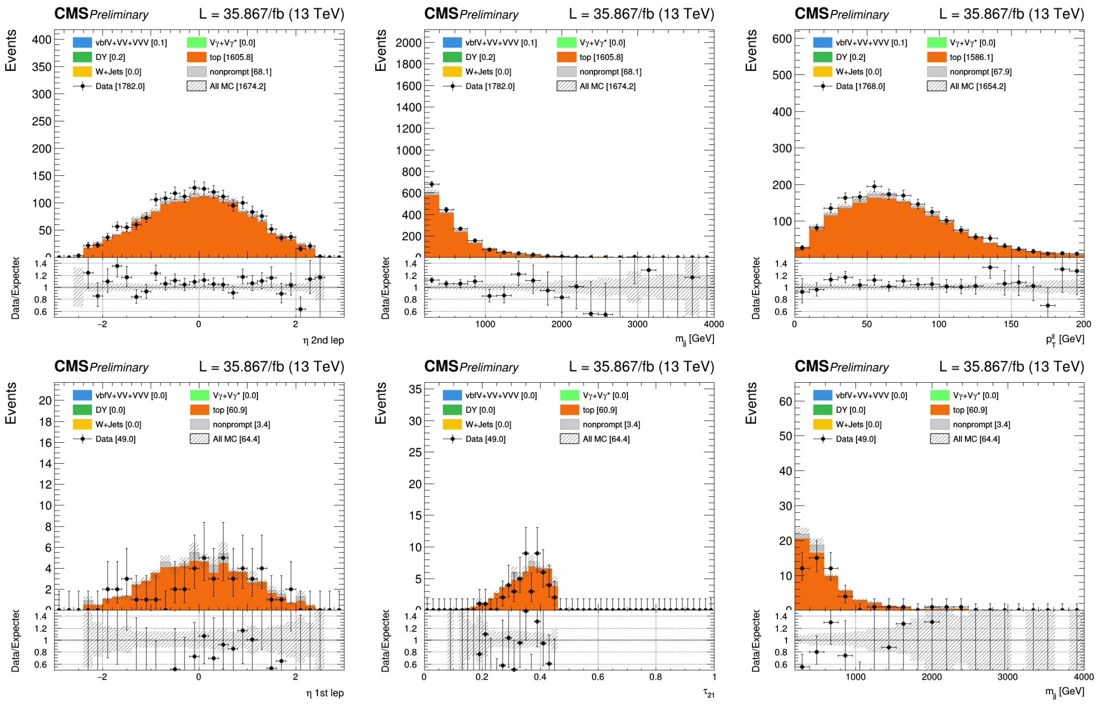


Figure 5.17: Comparison of data and MC estimate for the top enriched control region for 2016 in the resolved (boosted) category on top (bottom) line for four chosen kinematical variables relevant in the analysis.

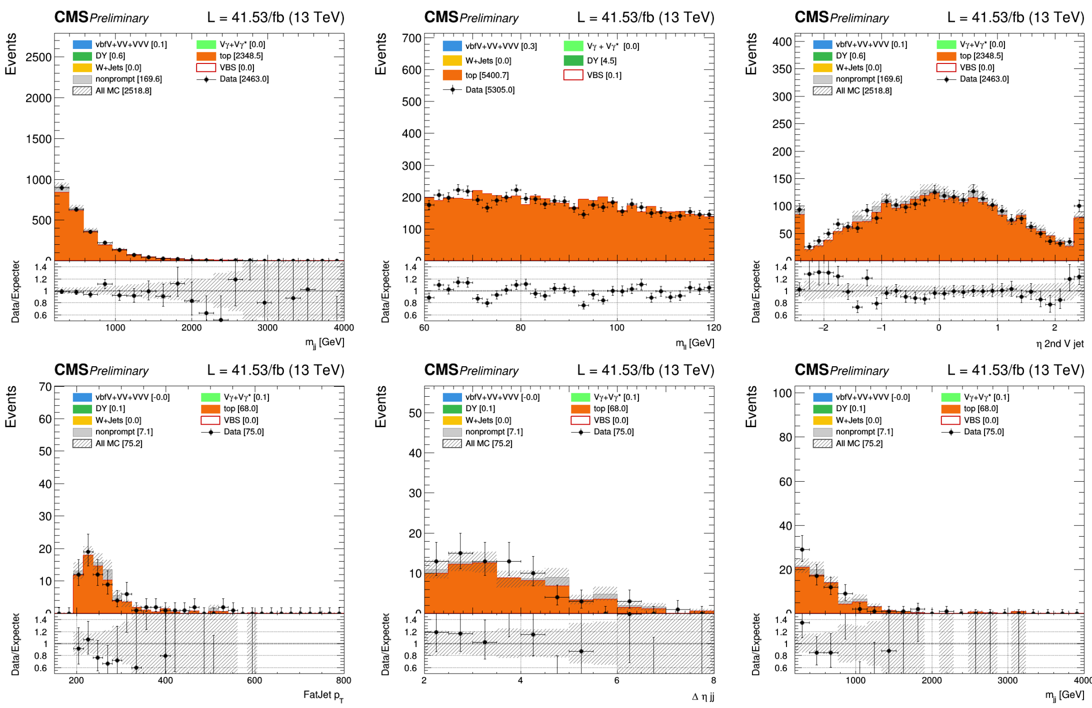


Figure 5.18: Comparison of data and MC estimate for the top enriched control region for 2017 in the resolved (boosted) category on top (bottom) line for four chosen kinematical variables relevant in the analysis.

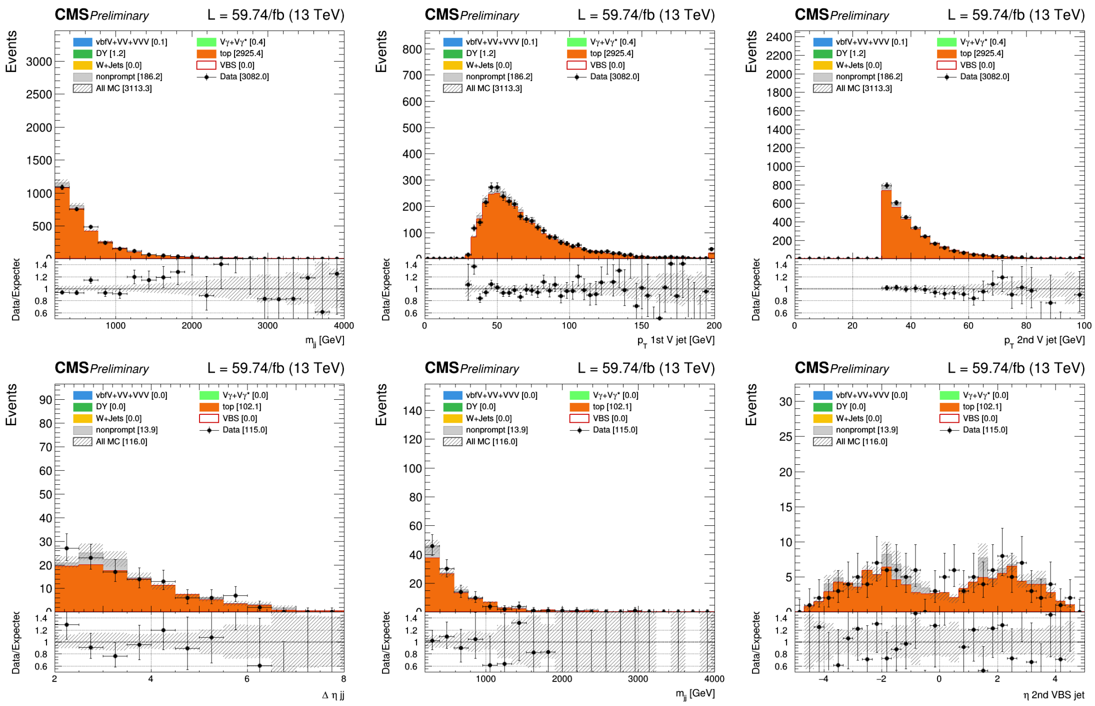


Figure 5.19: Comparison of data and MC estimate for the top enriched control region for 2018 in the resolved (boosted) category on top (bottom) line for four chosen kinematical variables relevant in the analysis.

8 TREATMENT OF SYSTEMATIC UNCERTAINTIES

In this section, we list the systematic uncertainties implemented in this search that are introduced as nuisance parameters in the fit to extract the signal strength. Each uncertainty may be modeled by one or several nuisance parameters, corresponding to specific samples. The systematic uncertainties that are uncorrelated across the years have separate nuisance parameters for each year, while correlated uncertainties share a nuisance parameter. We note that experimental uncertainties that are in common with other measurements follow a standardized naming scheme to facilitate future combinations.

We consider two types of systematic uncertainties: rate uncertainties (or "log normal"), which only affect the normalization of a process, and shape uncertainties. The shape uncertainties are modeled in the Combine Tool by providing variation templates corresponding to $\pm 1\sigma$ variations of the uncertainty. The corresponding nuisance parameter is then used to interpolate between the nominal and variation templates.

- Uncertainties specific to individual background processes.
  - **Non-prompt background:** Non-prompt background is modeled by applying weights (fake factors) to a superset of signal candidate events identified by leptons passing the loose selection criteria, as described in Section 5.7. Systematic uncertainties in fake factors arise from the limited size of the samples used to measure them and the difference in the flavor composition of the jets faking the leptons between the measurement

sample and the signal region. The maximum deviation of the fake factors from the nominal values are of order 5 to 10% for both sources. There is, however, a limitation to capturing the full effect of jet flavor composition difference with the fake factors. Therefore, a conservative 30% normalization uncertainty is additionally assigned to the fake background prediction, independently for each source of non-prompt background. Uncertainties on the fake factors are uncorrelated among the three data sets.

  – **top, Drell-Yan, and other minor background processes:** Many of the MC simulation samples receive specific weights (either event-by-event or overall) to better match the simulation prediction with the observation from the collision data. These weight factors carry systematic uncertainties, which are correlated among the three data sets.

- Uncertainties affecting all MC-based background and signal modeling. This category mostly consists of uncertainties in MC modeling of reality. All uncertainties under this category are uncorrelated among the three years unless otherwise stated.
  – **Integrated luminosity:** The uncertainty determined by the CMS luminosity monitoring is 2.5%, 2.3%, and 2.5% for 2016, 2017, and 2018 data sets, respectively. The uncertainties are partially correlated among the data sets considering the luminosity measurement schemes.
  – **Trigger efficiency:** Trigger efficiency uncertainties are of the order of less than 1%.
  – **Lepton reconstruction and identification efficiency:** The lepton reconstruction and identification efficiencies are measured with the tag and probe method in data. To correct for the difference in the lepton identification efficiencies between data and MC, data/MC scale factors dependent on p_T and η are applied to the MC simulation. The resulting uncertainty in the signal region is $\sim$1% for electrons and $\sim$2% for muon.
  – **Muon momentum and electron momentum scale:** These uncertainties arise due to different detector effects and are p_T and η-dependent. Uncertainties in both the scale and resolution individually amount to $\sim$0.6-1% for electrons and $\sim$0.2% for muons.
  – **Jet energy scale (JES) uncertainties:** The effect of the JES uncertainties are estimated by shifting the jet energy scale, and thereby p_{Tmiss}, by $\pm 1\sigma$ variations. The fully split set of JES uncertainties is used. The fit templates are modified both in shape and normalization as a result of this variation. JES uncertainties affect the rates in the signal region at the level of $\sim$10%. The JES uncertainties are partially correlated between years.
  – **b-tagging scale factor:** Uncertainties in the b-tagging scale factor are evaluated by shifting the per-jet scale factors. The components of these uncertainties related to theoretical modeling within the scale factor measurement are correlated among the data sets, while the remainder, arising from statistical uncertainties, are uncorrelated.

- **Pileup:** Uncertainties on the amount of pileup present in the event are assessed by varying the minimum bias cross section used to generate the pileup distribution by $\pm 1\sigma$. Events are reweighted to match this alternate distribution. Pileup uncertainties are uncorrelated between years.
- **Prefiring:** Uncertainty associated to the pre-firing corrections in 2016 and 2017.

- Theoretical uncertainties. All nuisance parameters associated to uncertainties in this category are correlated among three years unless stated otherwise.
  - **Parton distribution function (PDF):** PDF affects the distribution shapes as well as the overall cross sections of the processes. PDF uncertainties for the background processes are evaluated by reweighting the MC simulation events with different PDF sets based on parton-level kinematics, and then adding the resulting weights over the events in the signal region.
  - **Missing higher-order corrections (renormalization and factorization scale):** Missing higher-order terms in the perturbation series of the cross section calculations will also affect the shapes and the overall cross sections. For the background processes, MC simulation events are reweighted to alternative event weights where renormalization and factorization scales are varied up and down by factor 2, and the envelopes of the varied distributions are taken as one standard-deviation variations.
  - **Parton shower modeling:** Uncertainty in parton shower (PS) modeling primarily affects the jet multiplicity. The PS uncertainty is evaluated using the PS variation weights computed in Pythia 8 corresponding to the ISR and FSR variations. The uncertainty is evaluated separately in each njet bin.
  - **Underlying event modeling:** Uncertainty in UE is evaluated by shifting the respective UE tunes (CUETP8M1 for the 2016 analysis, CP5 for 2017 and 2018) of Pythia 8. This uncertainty is uncorrelated between 2016 and the two other years because the UE tunes are different. Currently, the UE uncertainty is parameterized as a flat value approximated from measuring the UE variation envelope.

A summary of the nuisance parameters corresponding the experimental uncertainties is reported in Table 5.13.

Since the DY+jets and top CRs are included with a simultaneous fit in the statistical analysis of the signal region, the final set of nuisance parameters included in this measurement are the "rateParams", used to extract a background normalization from a control region. The rateParam affects the background normalization in the signal region and control region simultaneously and is allowed to float freely in the fit. With a background-dominated, large statistics control region, the background normalization will then be constrained primarily by the control region and applied in the signal region. Each background normalized in a control region for each channel has a separate rateParam, indicating that these background normalizations are all independent. As the control regions are well-divided in kinematics, this is a reasonable assumption to make.

Uncertainty	Nuisance Parameter	Type	Years	Samples
Luminosity	lumi_13TeV_[year]	rate	All, uncorrelated	
	lumi_13TeV_XYFact	rate	All, correlated	
	lumi_13TeV_BBDefl	rate	2016-2017, correlated	All
	lumi_13TeV_DynBeta	rate	2016-2017, correlated	non-CR
	lumi_13TeV_Ghosts	rate	2016-2017, correlated	MC
	lumi_13TeV_LSCale	rate	2017-2018, correlated	
	lumi_13TeV_CurrCalib	rate	2017-2018, correlated	
Trigger	CMS_eff_hwwtrigger_[year]	shape	All, uncorrelated	All
Lepton eff.	CMS_eff_e_[year]	shape	All, uncorrelated	All
	CMS_eff_m_[year]	shape	All, uncorrelated	All
Lepton scale	CMS_scale_e_[year]	shape	All, uncorrelated	All
	CMS_scale_m_[year]	shape	All, uncorrelated	All
Prefiring	CMS_eff_prefiring_[year]	shape	2016-2017, uncorrelated	All
JES	CMS_scale_JESAbsolute	shape	All, correlated	All
	CMS_scale_JESAbsolute_[year]	shape	All, uncorrelated	All
	CMS_scale_JESBBEC1	shape	All, correlated	All
	CMS_scale_JESBBEC1_[year]	shape	All, uncorrelated	All
	CMS_scale_JESEC2	shape	All, correlated	All
	CMS_scale_JESEC2_[year]	shape	All, uncorrelated	All
	CMS_scale_JESFlavorQCD	shape	All, correlated	All
	CMS_scale_JESHF	shape	All, correlated	All
	CMS_scale_JESHF_[year]	shape	All, uncorrelated	All
	CMS_scale_JESRelativeBal	shape	All, correlated	All
	CMS_scale_JESRelativeSample_[year]	shape	All, uncorrelated	All
b tag	CMS_btag_jes	shape	All, correlated	All
	CMS_btag_lf	shape	All, correlated	All
	CMS_btag_hf	shape	All, correlated	All
	CMS_btag_hfstats1_[year]	shape	All, uncorrelated	All
	CMS_btag_hfstats2_[year]	shape	All, uncorrelated	All
	CMS_btag_lfstats1_[year]	shape	All, uncorrelated	All
	CMS_btag_lfstats2_[year]	shape	All, uncorrelated	All
	CMS_btag_cferr1	shape	All, correlated	All
	CMS_btag_cferr2	shape	All, correlated	All
Top background	singleTopToTTbar	shape	All, correlated	Top
Top p_T	topPtRew	shape	All, correlated	Top
Parton Shower	PS_ISR	rate	All, correlated	main bkg
	PS_FSR	rate	All, correlated	All
Underlying Event	UE_whss	rate	All, correlated	All

Table 5.13: Nuisance parameters used to model main experimental uncertainties.

9 STATISTICAL INTERPRETATION OF THE RESULTS

In this section we present the results of the statistical analysis, comparing the two methods shown in Sections 5.5.2, and 5.6. We recall that the analysis is blind in the signal region, and thus we can just perform an estimation of the statistical significance for the measurement of the ZVjj VBS EW process. The significance of a result is calculated using the ratio of profiled likelihoods, one in which the signal strength is set to 0 and the other in which it is free to float:

$$-2\ln\left[\mathcal{L}\left(\text{ data } \mid r=0,\hat{\theta}_0\right)/\mathcal{L}(\text{ data } \mid r=\hat{r},\hat{\theta})\right], \tag{5.3}$$

where the nuisance parameters are profiled separately for $r = \hat{r}$, and $r = 0$. For a fast estimate we work within the so-called *Asymptotic approximation* [170] for the distribution of this test-statistic, after having checked that our input distributions respect the fundamental requirements of Wilke's theorem. To take into account the contribution of the main background, we perform a simultaneous fit of the DY and top control regions.

We now show the results of the expected statistical significance for the year 2018 by comparing the *cut-based* approach presented in Section 5.5.2 to the DNN approach of Sec. 5.6. Within the *cut-based* approach, the significance is extracted by comparing several kinematical distributions. The highest performance in terms of statistical significance is obtained using the invariant mass of the two VBS-jets in the event, shown in Fig. 5.20. For the multivariate approach, once the training is satisfying, and statistical analysis is performed (for the moment we present the results only with the DNN), we evaluate the agreement for the distributions of the DNN output for MC and Data in control regions (Figure 5.21). The significance is then extracted in the same way by comparing the distributions of the output for SR and CR combined, for each topology. The summary of the results is presented in Table 5.14: the improvement given by the DNN techniques is evident. To conclude, Table 5.15 reports the estimate of the statistical significance with the DNN approach for the Full Run 2 statistics. It is worth noticing that some of the main systematics (including JER, JES, and QCD scale) are not included in this estimate.

Significance (σ)	Boosted	Resolved	Combined
DNN	1.11	2.12	2.38
Cut-based	0.85	1.25	1.51

Table 5.14: Expected significance on datasets for the "DNN" and the "cut-based" approach: we split into "boosted", "resolved" and the combination of the two categories. For the moment this results include only a subset of all the systematic uncertainty of the analysis: Luminosity, Trigger, Lepton efficiency and energy scale, Prefiring, b tag, UE, PS, PDF, and theoretical uncertainties on top momentum.

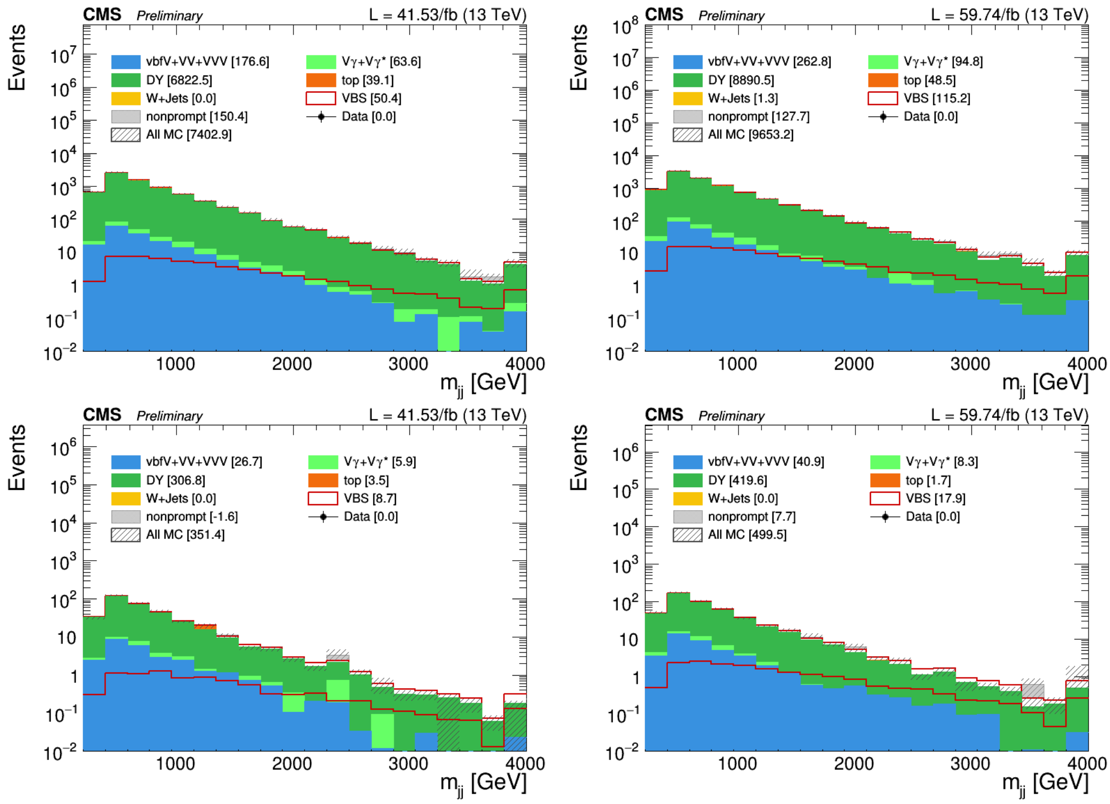


Figure 5.20: Distributions of invariant mass of the VBS-jets m_{jj} in the resolved (top line) and boosted (bottom line) signal region for 2017 on the left, and 2018 on the right. VBS ZVjj signal is both stacked and superimposed to underline the high-mass region sensitivity. Error bands account for both statistical and systematic uncertainties. The SR is still blind to data samples, and the plots refer to the SR, as presented in Sec. 5.5.2.

Significance (σ)	Boosted	Resolved	Combined
2016	0.4	0.67	0.77
2017	0.73	1.06	1.21
2018	1.11	2.12	2.38
Combined	1.39	2.45	2.77

Table 5.15: Expected significance on the Run 2 data taking, considering the three years separated and combined (last line): we split into "boosted", "resolved" and the combination of the two categories. For the moment this results include only a subset of all the systematic uncertainty of the analysis: Luminosity, Trigger, Lepton efficiency and energy scale, Prefiring, b tag, UE, PS, PDF, and theoretical uncertainties on top momentum.

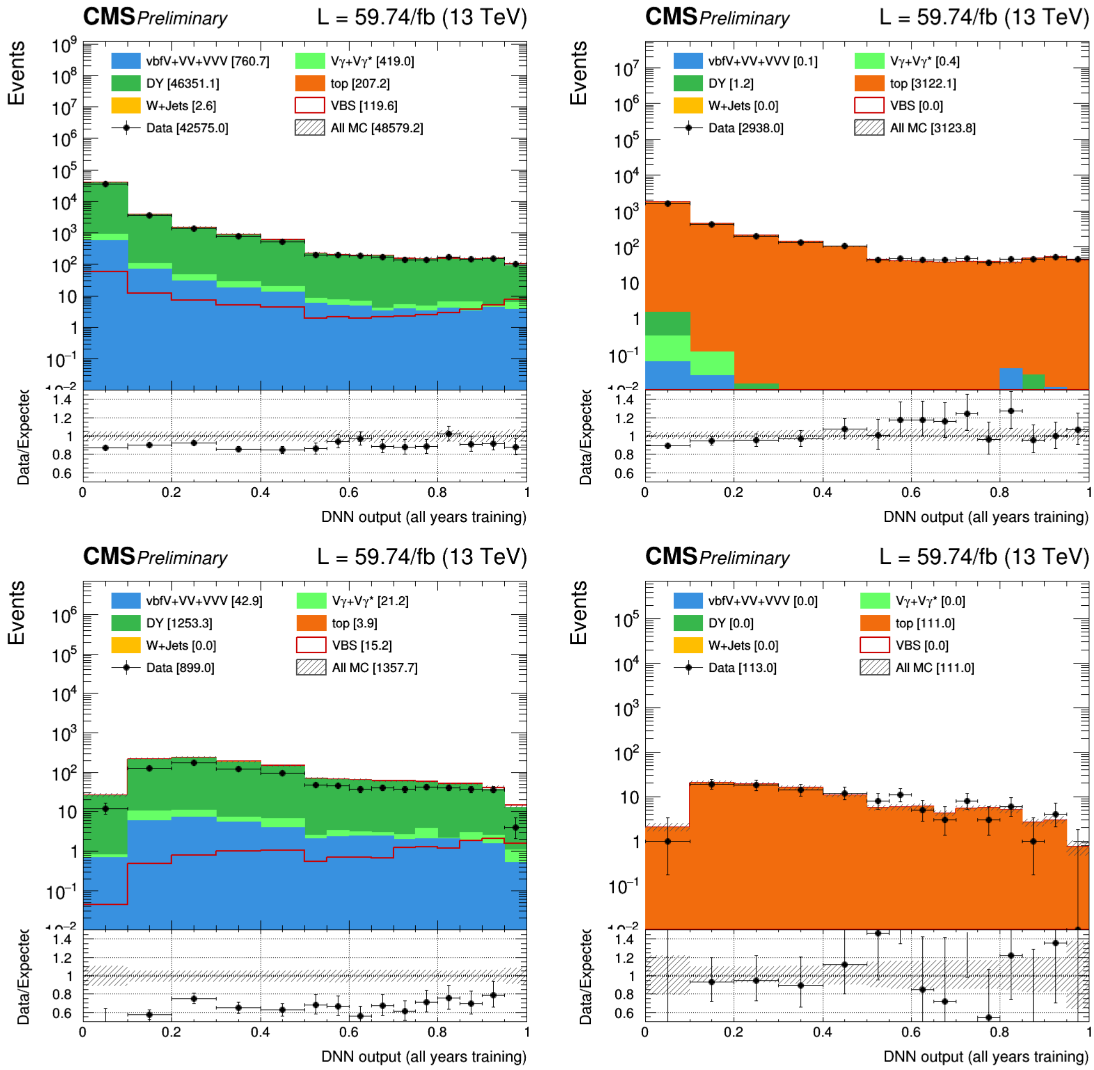


Figure 5.21: Comparison in log scale of data and MC estimate for DNN output variable in DY CR (left), top CR (right). First line: resolved category. Second line: boosted category. Except for the overall normalization, the agreement looks mostly flat across the range.

10 SUMMARY OF THE RESULTS

In conclusion, we have presented the very search for the VZ bosons $\mathcal{O}(\alpha^6)$ electroweak production, where V is the vector boson decaying hadronically (V=W, Z) and Z decays in charged leptons (e, μ), plus two jets in proton-proton collisions at 13 TeV. The data correspond to an integrated luminosity of 136 fb^{-1} collected by the CMS detector from 2016 to 2018. We have highlighted the potentiality of the introduction of multivariate techniques on the measurement and analyzed in detail two different approaches: a BDT and a DNN. The statistical significance of the full Run 2 LHC dataset is reported for the ZV EW $\mathcal{O}(\alpha^6)$ SM signal using a Deep Neural Network technique: an optimistic estimate points to a statistical significance of 2.77 σ.

Part III

POST LHC ERA: THE HL-LHC AND FUTURE COLLIDERS

6 FUTURE COLLIDERS STUDIES

In this chapter, we comment on the post-LHC era, and we test the discovery potential for new physics choosing as benchmark the search for composite Majorana neutrinos presented in Chapter 4. We start with a short review of the timeline of the LHC upgrade steps in Section 6.1, and we then present in Section 6.2 the main characteristics of novel CMS Phase 2 detector, that will operate in the High-Luminosity LHC. We then examine in detail a projection study performed within the CMS Collaboration using simulated samples within High-Luminosity LHC (HL-LHC) and High-Energy LHC (HE-LHC) colliders conditions. We conclude with a glance at the high-energy frontier of the Future Circular Collider (FCC) in Section 6.3. We present phenomenological studies at the hadron-collider option, the FCC-hh, for the composite Majorana neutrinos, and we point out how this 100 TeV machine will be able to access for the first time a phase space favored by the observed baryon asymmetry of the Universe.

1 LHC IN THE FUTURE: HL/HE-LHC AND FCC

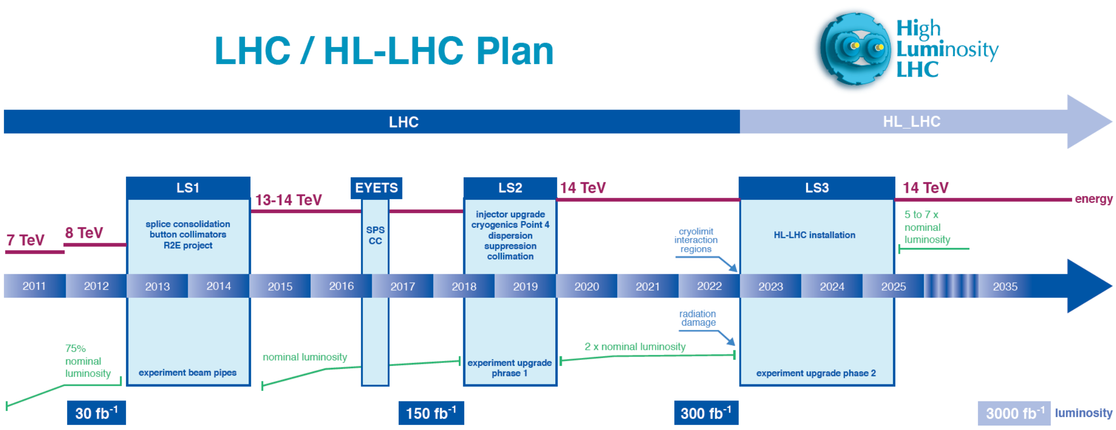


Figure 6.1: The timeline for the planned HL-LHC and LHC operations.

After the Long Shutdown 2 (LS2), the LHC is about to restart operations at increased center-of-mass energy, $\sqrt{s} = 14$ TeV, beginning the so-called Run 3. The general-purpose experiments will collect approximately double of the Run 2 statistics, trying to maximize the actual detector acceptance to new physics and precision measurements through hardware innovations and new searching techniques. After three years of data taking, and another time slot for installation procedures (LS3), the LHC will be giving-way to the HL-LHC phase, which should eventually deliver 3 ab^{-1} of integrated luminosity. All the LHC main experiments involved in Phase 2 will undergo a series of upgrades to be able to profit from such an extreme luminosity scenario. We will see in the next Section the description of the main feature of the CMS Phase 2 detector.

While the current priority lies in profiting from this experimental program, some effort should also be devoted to the design of future colliders, planning the investigation of the energy frontier on the timescale of several decades. This may well be premature: the next Run 3 and Phase 2 could radically change the situation should new particles be discovering, in which case the priority would be on characterizing their properties and nature. However, an assessment of future colliders' capabilities based on the current theoretical understanding and experimental status might still be a useful exercise. Proposed future machines come in two main categories: lepton colliders, like the ILC [171], CLIC [172], FCC-ee [173], and hadron, such as the HE-LHC [174] of the FCC-hh [175] colliders, which will search for new physics from complementary sides.

Lepton colliders are more suited to perform precision measurements, targeting indirect searches, while hadron colliders push the energy domain of the machine, aiming for the direct production of heavy new particles beyond the present reach.

2 THE CMS PHASE-2 DETECTOR

In this section, we present the main features of the CMS upgrade detector, that will become operational for the CERN High-Luminosity LHC (HL-LHC) phase. The HL-LHC is expected to deliver a peak instantaneous luminosity of up to 7.5×10^{34} $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [176], which is an increase in instantaneous luminosity of about four times with respect to the LHC Run 2 performance. With this increase in instantaneous luminosity, the number of overlapping proton-proton interactions per bunch crossing, or pileup (PU), are expected to increase from its mean value of about 40 at the LHC to a mean value of up to 200 at the HL-LHC. Similarly, the levels of radiation are expected to significantly increase in all regions of the detector, in particular in its forward regions.

The CMS detector [95] will be substantially upgraded in order to fully exploit the physics potential offered by the increase in luminosity, and to cope with the demanding operational conditions at the HL-LHC [177–181]. In particular, in order to sustain the increased PU rate and associated increase in flux of particles, the upgrade will provide the detector with: higher granularity to reduce the average channel occupancy, increased bandwidth to accommodate the higher data rates, and improved trigger capability to keep the trigger rate at an acceptable level without compromising physics potential. The upgrade will also provide an improved radiation hardness to withstand the increased radiation levels.

The upgrade of the first level hardware trigger (L1) will allow for an increase of L1 rate and latency to about 750 kHz and 12.5 μs, respectively. The upgraded L1 will also feature inputs from the silicon strip tracker, allowing for real-time track fitting and particle-flow reconstruction [182] of objects at the trigger level. The high-level software trigger (HLT) is expected to reduce the rate by about a factor of 100 to 7.5 kHz.

The entire pixel and strip tracker detectors will be replaced to increase the granularity, reduce the material budget in the tracking volume, improve the radiation hardness, and extend the geometrical coverage and provide efficient tracking up to pseudorapidities of about $|\eta| = 4$. In addition, the tracker will provide information on tracks above a configurable transverse momentum threshold to the L1 trigger, information presently only available at the HLT. It will also allow for including tracks with low momentum ($\approx$ 2GeV).

The muon system will be enhanced by upgrading the electronics of the cathode strip chambers (CSC), resistive plate chambers (RPC) and drift tubes (DT). New muon detectors based on improved RPC and gas electron multiplier (GEM) technologies will be installed to add redundancy, increase the geometrical coverage up to about $|\eta| = 2.8$, and improve the trigger and reconstruction performance in the forward region.

The barrel electromagnetic calorimeter (ECAL) will be operated at lower temperatures to mitigate noise in avalanche photodiodes (APDs) due to radiation damage. Its upgraded front-end electronics will be able to exploit the information from single crystals at the L1 trigger level, to accommodate trigger latency and bandwidth requirements, and to provide an increased sampling rate of 160 MHz and high-precision timing capabilities. The hadronic calorimeter (HCAL), consisting in the barrel region of brass absorber plates and plastic scintillator layers, will be read out by silicon photomultipliers (SiPMs).

The endcap electromagnetic and hadron calorimeters will be replaced with a new combined sampling calorimeter (HGCal) that will provide coverage in pseudorapidity from about $|\eta| = 1.5$ up to $|\eta| = 3$. The new calorimeter will be based on a lead tungsten followed by stainless steel absorber with silicon sensors as the active material in the front section, and it will feature plastic scintillator tiles readout by SiPMs towards its back section at large distances from the beam. It will provide highly-segmented spatial information in both transverse and longitudinal directions, as well as 160 MHz sampling allowing high precision timing capability for photons, which will allow for improved PU rejection and identification of electrons, photons, tau leptons, and jets.

Finally, the addition of a new precision timing detector for minimum ionizing particles (MTD) in both the barrel and endcap regions is envisaged to provide the capability for 4-dimensional reconstruction of interaction vertices that will significantly offset the CMS performance degradation due to high PU rates. The MTD is expected to achieve timing resolution of about 30 to 40 ps, and will provide coverage up to pseudorapidities of about $|\eta| = 3$.

A detailed overview of the CMS detector upgrade program is presented in Ref. [177–181]. The expected performance of the reconstruction algorithms and PU mitigation with the CMS detector is summarised in Ref. [183].

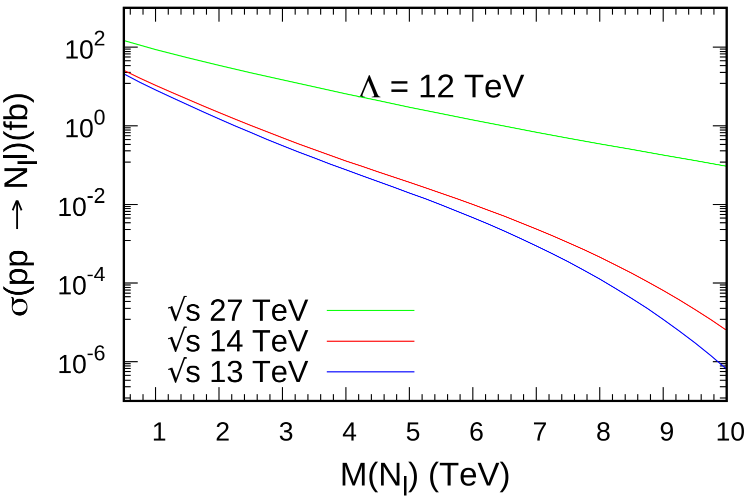


Figure 6.2: Comparison between production cross section of a heavy composite Majorana neutrino with a SM lepton for $\sqrt{s} = 13$ TeV, 14 TeV, and 27 TeV. The composite scale Λ is 12 TeV.

2.1 SEARCH FOR HEAVY COMPOSITE MAJORANA NEUTRINOS AT THE HIGH-LUMINOSITY AND THE HIGH-ENERGY LHC

In this part, the sensitivity of a search for Majorana neutrinos with the CMS Phase-2 detector in a final state with two leptons and at least one large-radius jet is investigated. Such new particles arise in theories beyond the standard model with compositeness. The study is based on searches previously performed with Run 2 CMS data in 2015 [42], where no evidence for a signal was found. We will show how the HL-LHC, with a center-of-mass energy of 14 TeV and an integrated luminosity of 3 ab^{-1}, will allow an extension of the sensitivity to cross-sections of the order of a few ab for heavy neutrino masses $M(N_\ell)$ ranging from 3 to 9 TeV for the $\ell\ell q\bar{q}'$ channel, where ℓ is an electron or a muon, and q is a quark. For the compositeness scale $\Lambda = M(N_\ell)$, the existence of a heavy Majorana neutrino could be excluded for masses up to 8 TeV at the 95% confidence level. The projection of the study to the High-Energy LHC (HE-LHC) scenario, with a center-of-mass energy of 27 TeV, is also presented here.

We have seen in Chapter 1.3, how many scenarios beyond the SM (BSM) suggest that the dynamics of new physics at high energies may be represented by an effective theory of higher-dimensional operators, with new degrees of freedom. We also presented in detail the compositeness of ordinary fermions as one possible BSM scenario that may lead to a solution of the hierarchy problem or the explanation of the proliferation of ordinary fermions [54, 57, 106, 184–186]. Here, we propose again the collider exploration for one such composite particle, a heavy Majorana neutrino N_ℓ. The interactions involving this resonance have already been presented in detail for the Full Run 2 analysis treated in Chapter 4. The only sizeable difference comes from the 1 TeV center-of-mass energy gain we will have at the HL-LHC, as shown in Figure 6.2.

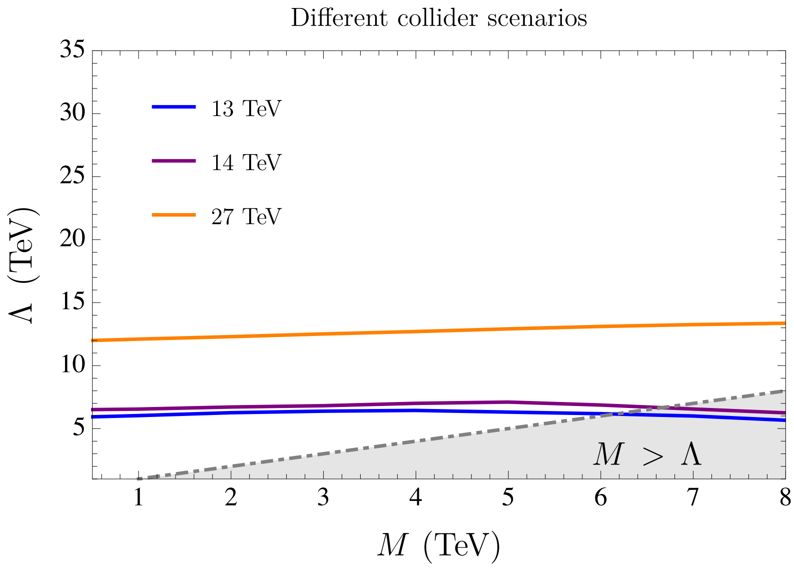


Figure 6.3: Comparison between the unitarity bounds for three different collider hypotheses: the LHC (blue line), the HL-LHC (purple line) and the HE-LHC (orange line). The 100% unitarity curves are considered, as explained in detail in Sec. 2.2.

The other details, such as the topology of the event or the ruling production mechanisms, remain unaltered. Hence, we will not repeat all the features of this model but only linger on the time scale of this CMS publication to emphasize the state of the art of this framework at the time of study we are presenting. This simulation study comes after the 2015 Run 2 publication for the search with heavy composite Majorana neutrino in di-lepton di-jet channel with 2015 data [42], but a year before the beginning of the Full Run 2 effort (Chapter 4), and more importantly, before the phenomenological studies on perturbative unitarity bounds in composite models 2.2. Therefore, the results will take the 2015 publication as a benchmark and do not directly include any consideration on the unitarity bounds. Nonetheless, the findings on unitarity violation will not spoil the validity of the HL/HE-LHC results. Indeed, if we compare the unitarity bounds for the three different collider scenarios in Figure 6.3, we can see that the bound remains stable from 13 TeV to 14 TeV. On the other hand, the enhancement in terms of statistics from LHC to HL-LHC allows a wider exploration of the (Λ, M) phase space, keeping the results safe from EFT ambiguous interpretations.

2.2 SIMULATED SAMPLES

We use Monte Carlo (MC) samples for the signal and the SM backgrounds. The MC samples for the signal are generated at Leading Order (LO) with CalcHEP (v3.6) [83] for $\sqrt{s}$ = 14 TeV proton-proton collisions, using the NNPDF3.0 LO parton distribution functions with the four-flavor scheme [84], for the Λ and mass values given in Table 6.1. Regarding the choice of background samples, the list of major background processes in Run 2 analysis has been considered. The dominant background process is top quark pair production ($\mathrm{t\bar{t}}$) which is estimated together with the single top quark (tW) contribution. These two sources of background are

$\Lambda\,[TeV]$	$M(N_\ell)\,[TeV]$			
6	-	-	6	-
9	-	3	6	9
12	0.5	3	6	9
15	0.5	3	6	-
18	0.5	3	-	-
21	0.5	-	-	-

Table 6.1: Parameters used in the HL-LHC analysis, $\ell = e, \mu$.

always considered together in this analysis and the combination is referred to as "TTtW". The Drell-Yan (DY) process gives rise to another source of background when two leptons are produced together with initial-state radiation that results in a jet. Other SM background sources such as W+jets and diboson production (WW, WZ, ZZ) are small ($\sim$ 5% of the total) and will be referred to as ”Other”. The background samples are generated with MadGraph5_aMC@NLO [187] using the CTEQ6L1 PDF set [188].

For all of the MC samples the hadronization of partons is simulated with PYTHIA 8 [189] and the expected response of the upgraded CMS detector with the fast-simulation package DELPHES [190]. The object reconstruction and identification efficiencies, as well as the detector response and resolution, are parameterized in Delphes using the detailed simulation of the upgraded CMS detector based on the Geant4 package [160,191]. The contribution from 200 additional pileup events has been included in the simulation as well.

The list of the samples for the signal points and for the backgrounds is reported in Table 6.2.

2.3 SIMULATION OBJECTS AND EVENT SELECTION

Final state objects are reconstructed by the particle flow (PF) algorithm [192] as implemented in Delphes. The PF algorithm combines information from all CMS subdetectors and reconstructs individual particles in the event such as electrons, muons, photons, neutral hadrons, and charged hadrons.

The event selection of the present analysis is based on exploiting some specific kinematic features of the leptons and the large-radius jet in the signal samples to minimize the contamination from the SM backgrounds.

The transverse momentum p_T of the leading lepton is required to be greater than 110 GeV, while the p_T of the subleading lepton must be greater than 40 GeV (Figure 6.4). All lepton candidates are required to be in the pseudorapidity range $|\eta| < 2.4$. Restricting to the high-mass region given by $M(\ell, \ell) > 300$ GeV, where $M(\ell, \ell)$ is the dilepton invariant mass, allows reducing the DY background and part of the TTtW background, without affecting the signal acceptance.

An overview of the main kinematical distributions for both the channels, for signal sample events passing the selection, is displayed in Fig.6.5. The large-radius

	Dataset	σ (pb)
	Signal samples	
	/ExtendedWeakIsospinModel_*jj_L6000_M6000_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	1.610^{-1}
	/ExtendedWeakIsospinModel_*jj_L9000_M3000_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	1.57
	/ExtendedWeakIsospinModel_*jj_L9000_M6000_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	3.1610^{-2}
	/ExtendedWeakIsospinModel_*jj_L9000_M9000_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	2.0510^{-4}
	/ExtendedWeakIsospinModel_*jj_L12000_M500_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	1.110
	/ExtendedWeakIsospinModel_*jj_L12000_M3000_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	4.9610^{-1}
	/ExtendedWeakIsospinModel_*jj_L12000_M0000_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	9.9910^{-3}
	/ExtendedWeakIsospinModel_*jj_L12000_M9000_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	9.2910^{-5}
	/ExtendedWeakIsospinModel_*jj_L15000_M500_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	4.46
	/ExtendedWeakIsospinModel_*jj_L15000_M3000_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	1.910^{-1}
	/ExtendedWeakIsospinModel_*jj_L15000_M6000_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	4.110^{-3}
	/ExtendedWeakIsospinModel_*jj_L18000_M500_nnpdf30lo_14TeV_CMS_22000PU_PhaseII.root	2.14
	/ExtendedWeakIsospinModel_*jj_L18000_M3000_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	7.710^{-2}
	/ExtendedWeakIsospinModel_*jj_L21000_M000_nnpdf30lo_14TeV_CMS_200PU_PhaseII.root	1.16
	Background samples	
"TTtW"	tt-4p-0-600-v1510_14TEV_200PU/	530.89
	tt-4p-600-1100-v1510_14TEV_20200PU/	42.5535
	tt-4p-1100-1700-v1510_14TEV_200PU/	4.48209
	tt-4p-1700-2500-v1510_14TEV_200PU/	0.52795
	tt-4p-2500-100000-v1510_14TEV_200PU/	0.05449
	tB-4p-0-500-v1510_14TEV_200PU/	63.889
	tB-4p-500-900-v1510_14TEV_200PU/	7.12172
	tB-4p-900-1500-v1510_14TEV_200PU/	0.98030
	tB-4p-1500-2200-v1510_14TEV_200PU/	0.08391
	tB-4p-2200-100000-v1510_14TEV_200PU/	0.00953
"DY"	LL-4p-0-100-v1510_14TEV_200PU/	1341.36
	LL-4p-100-200-v1510_14TEV_200PU/	156.295
	LL-4p-200-500-v1510_14TEV_200PU/	42.401
	LL-4p-500-900-v1510_14TEV_200PU/	2.84373
	LL-4p-900-1400-v1510_14TEV_200PU/	0.20914
	LL-4p-1400-100000-v1510_14TEV_200PU/	0.02891
"Other"	BB-4p-0-300-v1510_14TEV_200PU/	249
	BB-4p-300-700-v1510_14TEV_200PU/	35.2
	BB-4p-1300-2100-v1510_14TEV_200PU/	0.412
	BB-4p-2100-100000-v1510_14TEV_200PU/	0.0477
	Bjj-vbf-4p-0-700-v1510_14TEV_200PU/	86.456
	Bjj-vbf-4p-700-1400-v1510_14TEV_200PU/	4.3486
	Bjj-vbf-4p-1400-2300-v1510_14TEV_200PU/	0.32465
	Bjj-vbf-4p-2300-3400-v1510_14TEV_200PU/	0.03032

Table 6.2: List of samples with their path in the storage and their cross sections. Signal samples path is referred to the main folder "/YR_Delphes/Delphes_342pre12/ExtendedWeakIsospin_200PU". Background samples paths are referred to "/DelphesFromLHE_342pre05hadd_2017Aug/".

jets, i.e. clustered with a size parameter $R = 0.8$ ("AK8 jets"), are reconstructed using the anti-k_T algorithm [193], implemented in the FastJet package [194]. The large-radius jets are analyzed using the Pileup-per-particle-identification (PUPPI) mitigation algorithm [163]. This algorithm is designed to remove PU using event

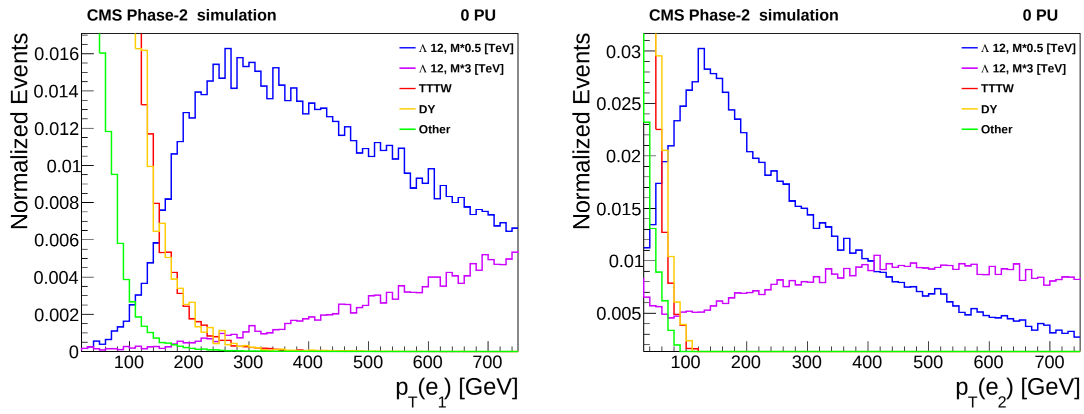


Figure 6.4: Details of the kinematical distribution for transverse momenta of the two electrons.

information both at the global and local level, identifying pileup at the particle level. The AK8 jets are required to have a minimum p_T of 200 GeV, to be within a pseudorapidity region with $|\eta| < 2.4$ and to be separated from leptons by a distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} > 0.8$. The process under consideration has a fairly central distribution of its final state particles. A study of the effect of increasing the pseudorapidity acceptance for the final state objects, made possible by the Phase-2 CMS detector upgrade, has shown that the increased background contribution would spoil the advantage given by the larger efficiency, lowering the overall sensitivity. Hence the more central selection, $|\eta| < 2.4$, has been kept for both leptons and large-radius jets (see Figure 6.6). . Requiring one or more large-radius jets is suitable regardless of whether N_ℓ decays through gauge or contact interactions. In fact, for gauge mediated decays of the heavy composite neutrino, the two quarks are expected to overlap and thus form a large-radius jet, while in the case of contact-mediated decays, the two quarks are well separated, but form two large-radius jets because of the overlap with final state radiation. The signal region is therefore defined by requiring two same-flavor isolated leptons (electrons or muons) and at least one large-radius jet (Figures 6.7, 6.8).

With this selection, the total efficiency for the signal is about 55% in the $eeq\bar{q}'$ channel and 65% in the $\mu\mu q\bar{q}'$ channel, for heavy neutrinos with masses greater than 3 TeV.

A shape-based analysis is performed investigating the invariant mass, $M(\ell\ell J)$, of the two leptons and the leading large-radius jet. As shown in Fig. 6.11 and commented in detail in Chapter 4, this variable provides a good discrimination between the signal and the SM backgrounds.

We finally discuss the choice of a charge-blind analysis instead of a selecting same-sign leptons. Naively, asking a same-sign strategy would improved the sensitivity to the presence of a Majorana neutrino. However, our studies show a small gain limited to the low heavy neutrino mass region (Figure 6.12).

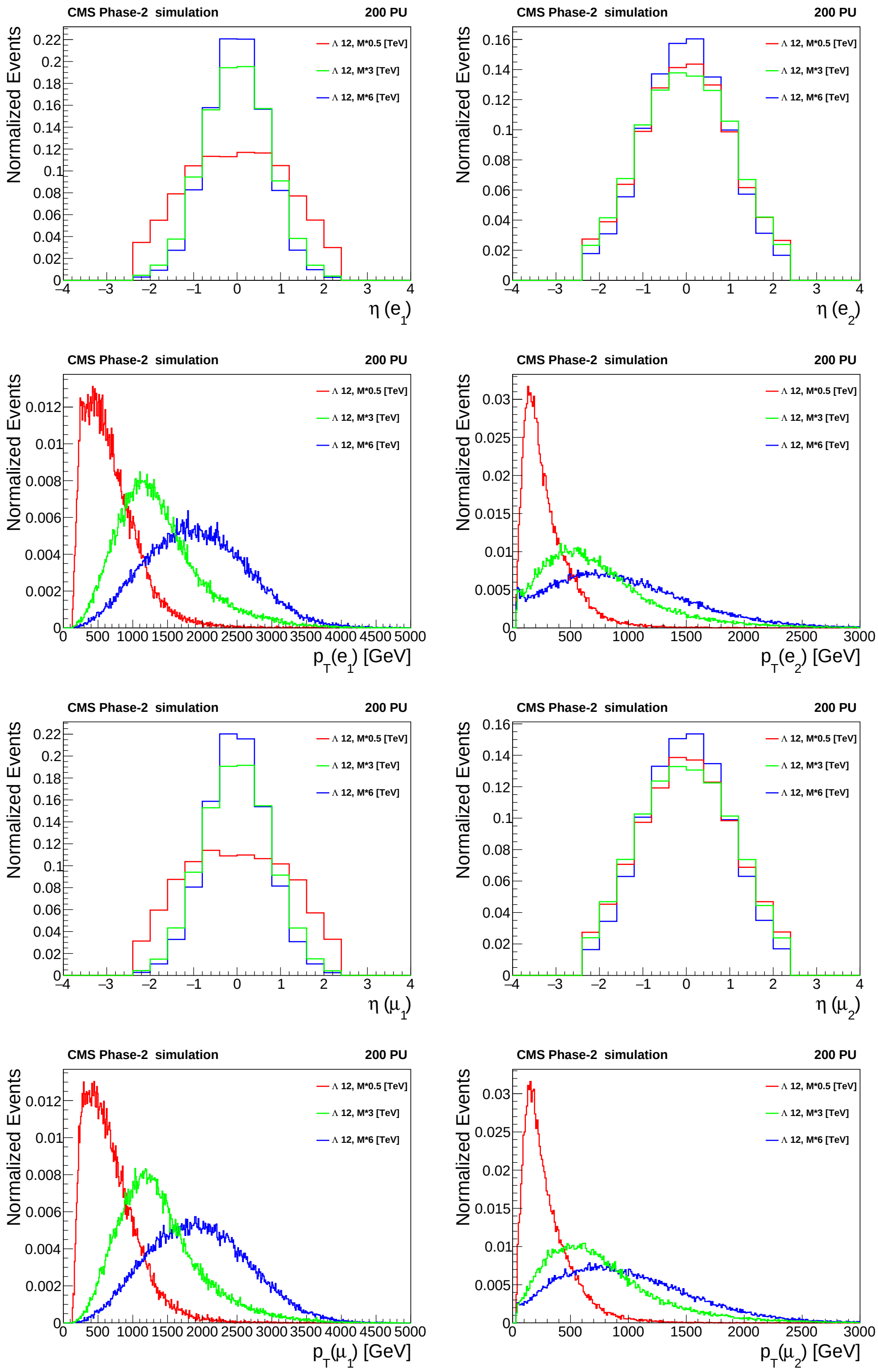


Figure 6.5: Main kinematical distributions for the reference composite scenario $\Lambda = 12$ TeV and $M(N_l) = 0.5$ TeV, $M(N_l) = 3$ TeV and $M(N_l) = 6$ TeV for the $eeq\bar{q}'$ channel (first two rows) and $\mu\mu q\bar{q}'$ channel (last two rows).

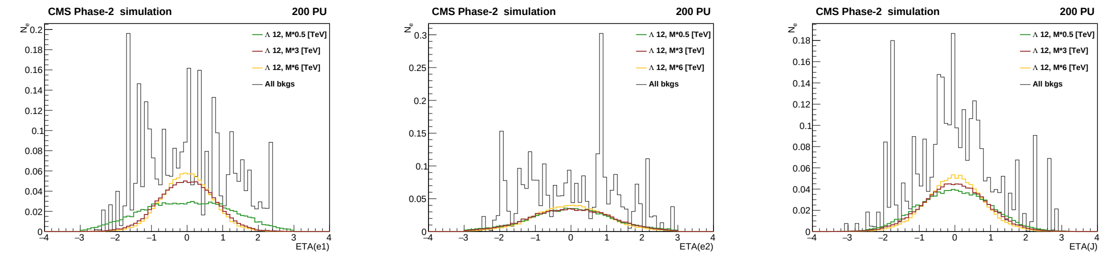


Figure 6.6: The pseudorapidity distributions of the leading electron (left), second leading electron (centre) and the leading large-radius jet (right) for the signal samples with $M(N_l)$ =0.5, 3 and 6 TeV ($\Lambda = 12$ TeV) and for all the background samples considered together (solid black lines). The signal objects are detected in the central region, hence justifying the narrow selection $|\eta(l, J)| \leq 2.4$. The performances of this choice on the limit extraction, compared to the acceptance offered by the CMS Phase 2 detector, are described in Ref. [195].

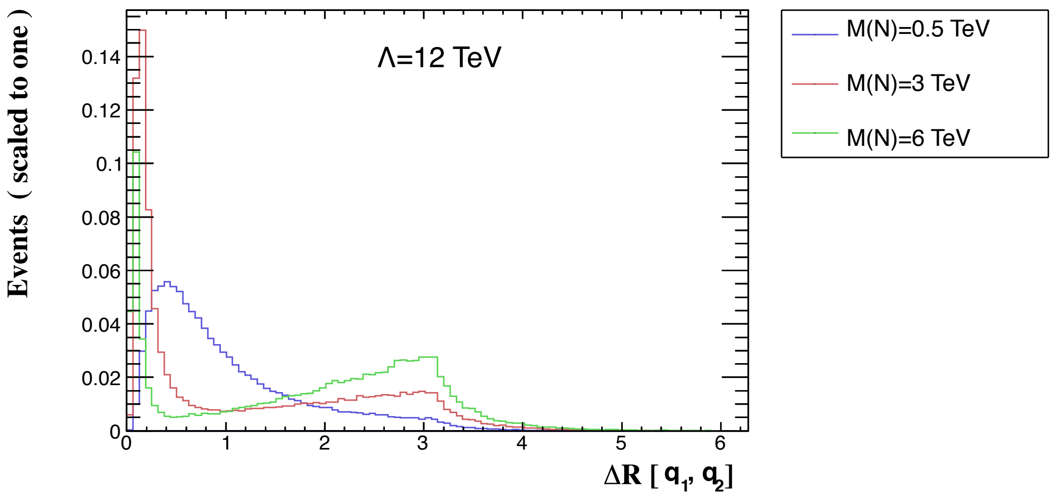


Figure 6.7: Generator level ΔR distribution between the two quarks q_1 and q_2 for signal samples with $\Lambda = 12$ TeV and $M(N_l) = 0.5$, 3 and 6 TeV. As the mass of the heavy neutrino approaches low values, the decay mechanism is ruled by the gauge interaction and the two quarks emerge close. The distribution is obtain directly from LHE sample, performed with MadAnalysis 5 [85].

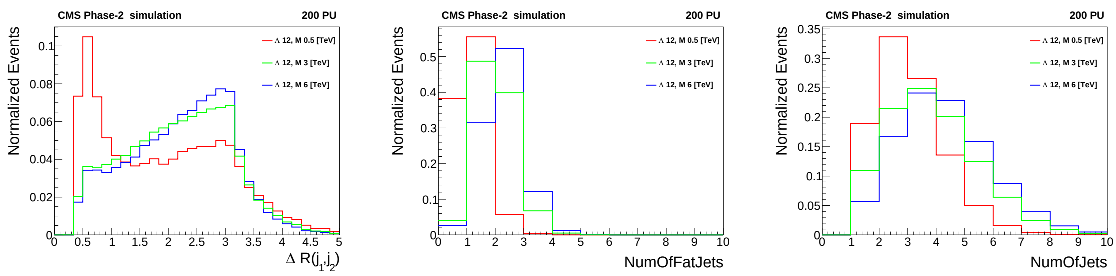


Figure 6.8: Left: distribution of the smallest ΔR between the two AK4 jets j_1 and j_2, decaying from the heavy composite Majorana neutrino, for signal samples with parameters $\Lambda = 12$ TeV and $M(N_l) = 0.5$ TeV, $M(N_l) = 3$ TeV and $M(N_l) = 6$ TeV. Centre: multiplicity distributions of the fat jets (AK8) given the same signal parameters. Right: multiplicity distributions of the AK4 jets.

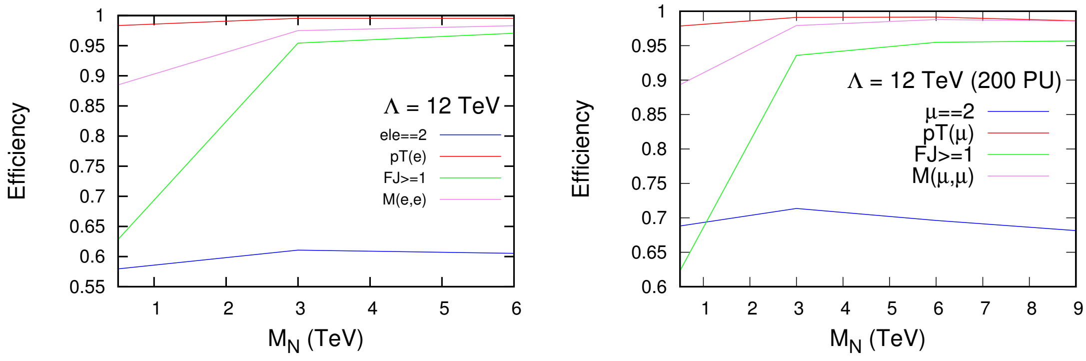


Figure 6.9: The relative efficiency of the various cuts used in the analysis for the signal region for the electron channel on the left, and for the muon channel on the right. The relative efficiency of the cuts is derived considering the order shown in the label.

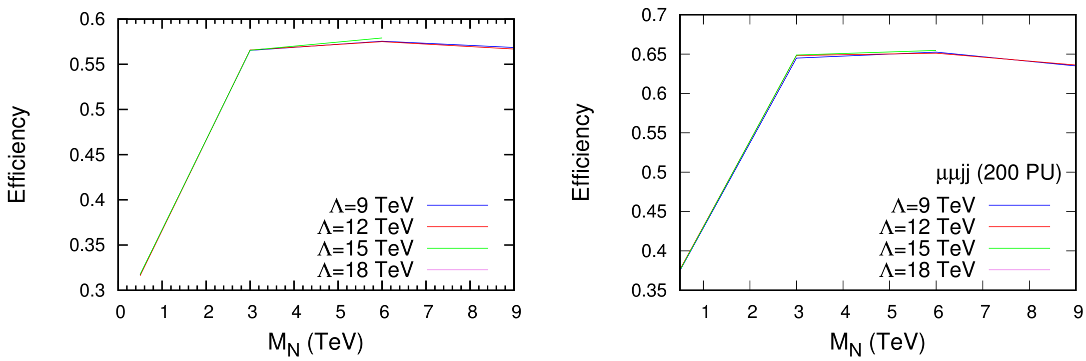


Figure 6.10: The cumulative efficiency in the signal region for the electron channel (left) and for the muon channel (right). The efficiency depends only on the neutrino mass, as expected.

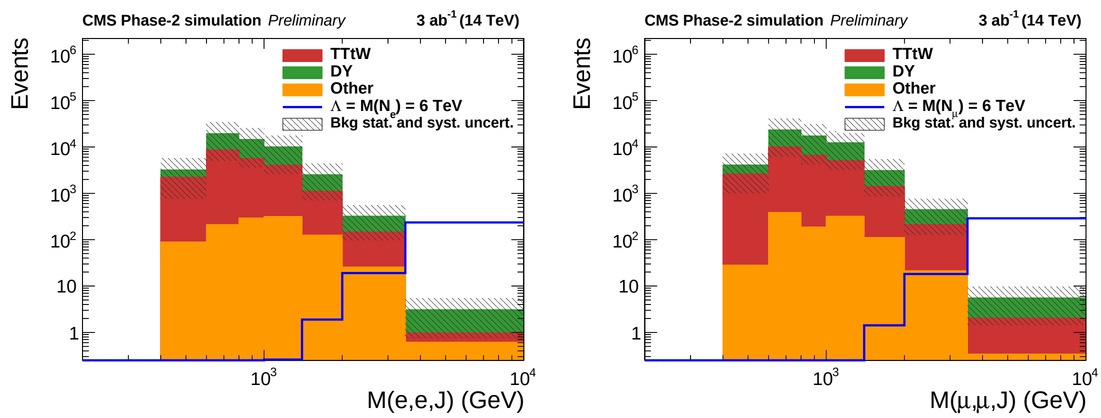


Figure 6.11: Distribution of the variable $M(\ell\ell J)$ of backgrounds (stacked plots) and expected signal (lines) in the signal region, considering the model parameters $\Lambda = M(N_\ell) = 6$ TeV, for the $eeq\bar{q}'$ channel (left) and for the $\mu\mu q\bar{q}'$ channel (right). The background statistical and systematic uncertainties have been combined.

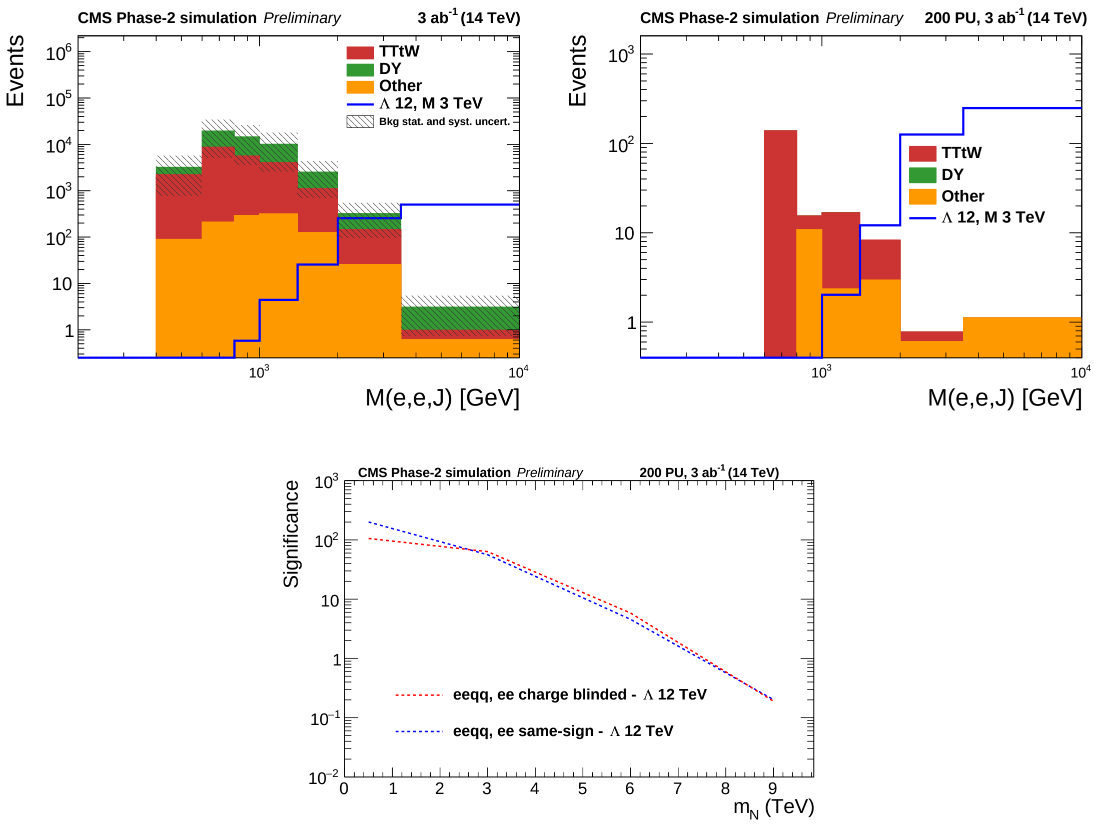


Figure 6.12: First line: the invariant mass distributions for the HL-LHC projection study for the charge-blind analysis on the left, and with the additional requirement of a same-sign analysis on the right. Second line: comparison between the expected statistical significance for the charge blinded analysis strategy adopted so far (red) and for the same-sign analysis strategy (blue).

2.4 TREATMENT OF SYSTEMATIC UNCERTAINTIES AT THE HL-LHC

In this section we try to asses the impact of systematic uncertainties on a future search for heavy composite Majorana neutrinos at the HL-LHC. We follow the general prescription defined in Ref. [196], extrapolating the expected effect for each systematic source starting from the reference Run 2 search published with 2015 Run2 data [42]. Table 6.6 presents the list of main systematic uncertainties for the analysis, in comparison with the Run 2 search.
All systematic effects apply the same method for signal and background samples and consider fixed values for all the mass spectrum. To be as consistent as possible with the UPSG recommendations, we have adopted this reasonable simplification, further justified considering that the effect of systematics on the analysis is not dominant, as can be seen from the results in the next section when we consider the special case of forcing to zero the systematic contributions.
One of the systematic contributions that requires a more detailed description is the one associated to the background estimate methods. In the 2015 Run 2 analysis this systematics contribution ("background") is evaluated using a control sample, which is limited by data statistics and scale factors that are limited by data and MC statistics[1].
The relation used for the systematic estimate is: $BKG_{SR} = BKG_{CR} \times w$, where BKG_{CR} is the background in a given control region, and w is a weight used to normalise BKG_{CR} to the signal region. Therefore, the systematic error reads: $\sqrt{(\delta B_{CR} w)^2 + (B_{CR} \delta w)^2}$, obtained bin by bin.

In the case of the Phase 2, analysis we have considered two options. In the first one (v1), we rescale only the uncertainties related to the data to the $3\,\mathrm{ab}^{-1}$ value. In this example this scaling would affect only the number of events in the BKG_{CR}.

In the second case (v2), we follow the prescription in [196] more tightly and bring all the uncertainties related to MC statistics to zero. In this case also the uncertainty on the other components of the formula would go to zero. In Table. 6.4 are reported the values of the systematic uncertainties obtained in the different bins of the invariant mass $M(\ell\ell J)$ and for the two main background components. The total uncertainty is obtained with a weighted sum of these contributions, see Table.6.5.

We would like to report here two possible scenarios for the "Background" systematic estimation at the HL-LHC:

- "HL sys v1": 5.9 % , scaling only the data related quantity BKG_{CR} to the luminosity;
- "HL sys v2": 0.3 %, scaling BKG_{CR} to the luminosity and considering the MC uncertainty to be zero.

The projections in the next section are shown considering mainly the most optimistic scenario ("HL sys v2").

1 For more details see the description regarding the Run 2 analysis in AN-2016-096, lines 351-357.

Source	(HL-LHC) Value	Obtained from	(Run 2) Value Sig (Bkg)
Luminosity	1	UPGAnalysisSystematics	2.7 (2.7)
Pile-up	2	UPGAnalysisSystematics	2 (4)
Electron ID	0.5	UPGAnalysisSystematics	2 (1)
Electron scale	0.5	UPGAnalysisSystematics	- (2)
Muon ID	0.5	UPGAnalysisSystematics	2 (3)
Muon scale	0.5	UPGAnalysisSystematics	- (2)
Jet Energy Scale	1	UPGAnalysisSystematics	0.4 (2)
Jet Energy Resolution	1	UPGAnalysisSystematics	1 (3)
Background	0.3 (6)	EXO-16-026, v2(v1)	- (15)
Drell-Yan theory	4	EXO-16-026, scaled by 0.5	- (8)

Table 6.3: List of systematic uncertainties used in the present analysis in percentage, compared to the systematic uncertainties adopted by the Run 2 [197] analysis for the *eeqq* channel in the last column.

	TTtW
bin content (Run 2)	$\{0, 0, 5.34763, 14.4046, 9.09282, 12.8957, 2.88059, 0, 0\}$
Run 2 sys	$\{0, 0, 24.727, 15.6224, 20.6516, 18.4183, 45.6614, 0, 0\}$
HL sys v1	$\{0, 0, 7.03097, 4.02197, 4.03945, 5.02422, 10.0764, 0, 0\}$
HL sys v2	$\{0, 0, 0.659216, 0.419777, 0.563178, 0.492744, 1.23843, 0, 0\}$
	DY
bin content (Run 2)	$\{0, 0, 0.796066, 8.13157, 8.07128, 8.28951, 3.39201, 0.751095, 0.074165\}$
Run 2 sys	$\{0, 0, 46.9726, 13.4199, 12.3129, 10.4141, 13.2238, 17.9083, 38.0802\}$
HL sys v1	$\{0, 0, 46.9309, 13.2734, 12.1531, 10.2247, 13.0751, 17.7988, 38.0288\}$
HL sys v2	$\{0, 0, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05\}$

Table 6.4: Details of the "Background" systematic uncertainty derivation. Two methods for rescaling systematic uncertainty are presented: "HL sys v1" is obtained by scaling BKG_{CR} to the luminosity; "HL sys v2" is obtained by scaling BKG_{CR} to the luminosity and considering the MC uncertainty to be zero. The weighted average according to the bin content is shown in Table 6.5.

	HL sys v1	HL sys v2
TTtW	5 %	0.5 %
DY	13 %	0.01 %
TTtW $\oplus$ DY	5.9 %	0.3 %

Table 6.5: The weighted average according to the bin content for the two scenarios: for the TTtW and DY backgrounds considered independently and togheter with the squared weighted sum (last row). We note that the weight due to the DY is about 0.4, whether the TTtW is about 0.6.

2.5 RESULTS

Figure 6.11 shows the distribution of the variable $M(\ell\ell J)$ for both the SM backgrounds and the signal for a particular benchmark choice of the model parameters, namely $\Lambda = M(N_\ell) = 6$ TeV. The expected discovery sensitivity of a heavy composite Majorana neutrino, produced in association with a lepton, and decaying into a same-flavor lepton and two jets, is shown in Figure 6.13. The CMS Phase-2 detector will be able to find evidence for a composite neutrino with mass below $M(N_\ell) = 7.6$ TeV. Expected exclusion limits on the mass of the heavy neutrino are also evaluated. An asymptotic CL_s criterion [170, 198] is used to set an upper limit at 95% confidence level on the cross section of the heavy composite Majorana neutrino produced in association with a lepton times its branching fraction to a same-flavor lepton and two quarks, $\sigma(pp \to \ell N_\ell) \times \mathcal{B}(N_\ell \to \ell q\bar{q}')$. The $M(\ell\ell J)$ distributions from MC simulations of signal and SM backgrounds are used as input in the limit computation together with the systematic uncertainties, summarized in Table 6.6, as discussed in the previous Section. The impact of systematic effects is negligible and, for high neutrino masses, the analysis is only statistically limited, as emerges from the comparison with limits extracted considering only statistical uncertainties in Fig. 6.14. The results are presented in Fig. 6.15 for the $eeq\bar{q}'$ channel and for the $\mu\mu q\bar{q}'$ channel for different values of the compositeness scale: $\Lambda = 12, 24, 35$ TeV and $\Lambda = M(N_\ell)$. Figure 6.16 displays the corresponding upper limits on the $(\Lambda, M(N_\ell))$ plane for both of the final states considered. The extrapolation is similar for the two channels. We see that the sensitivity to Λ is higher at low neutrino masses, reaching the exclusion of $\Lambda = 35$ TeV at $M(N_\ell) = 0.5$ TeV and decreases at higher N_ℓ masses. For the representative case $\Lambda = M(N_\ell)$, while in the analysis of Run 2 data it was possible to exclude heavy neutrino masses up to 4.60 (4.70) TeV in the $eeq\bar{q}'$ ($\mu\mu q\bar{q}'$) channel [42], under planned HL-LHC conditions this limit would be extended to 8 TeV.

Source	Value
Integrated luminosity	1%
Pileup	2%
Electron ID	0.5%
Electron scale	0.5%
Muon ID	0.5%
Muon scale	0.5%
Jet energy scale	1%
Jet energy resolution	1%
Background	0.3%
Drell-Yan (theory)	4%

Table 6.6: List of systematic uncertainties used in the present analysis.

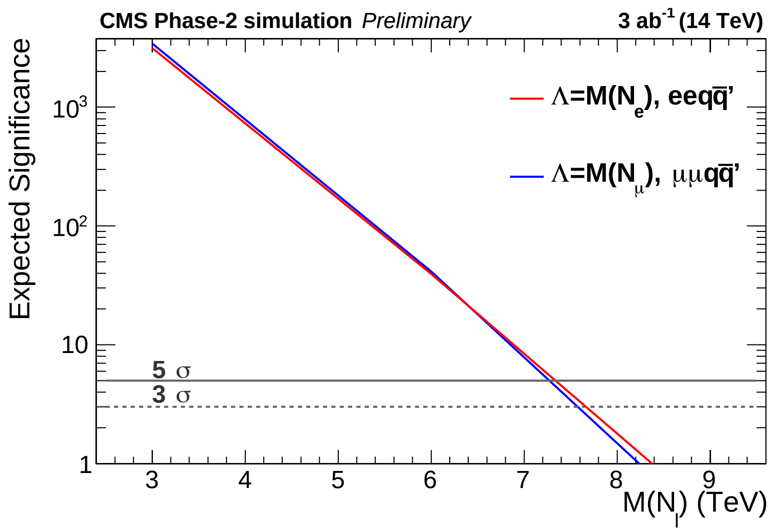


Figure 6.13: The expected statistical significance for both the $eeq\bar{q}'$ (red line) and $\mu\mu q\bar{q}'$ (blue line) channel for the case $\Lambda = M(N_\ell)$. The gray solid (dotted) line represents 5 (3) standard deviations, respectively.

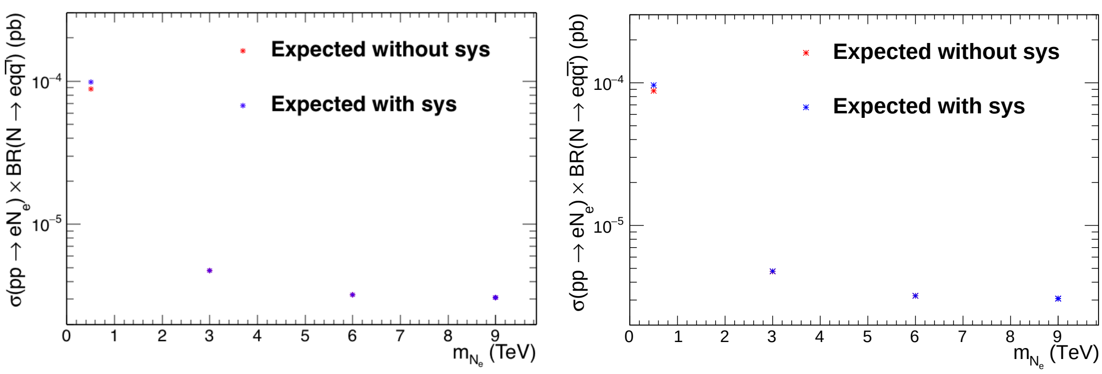
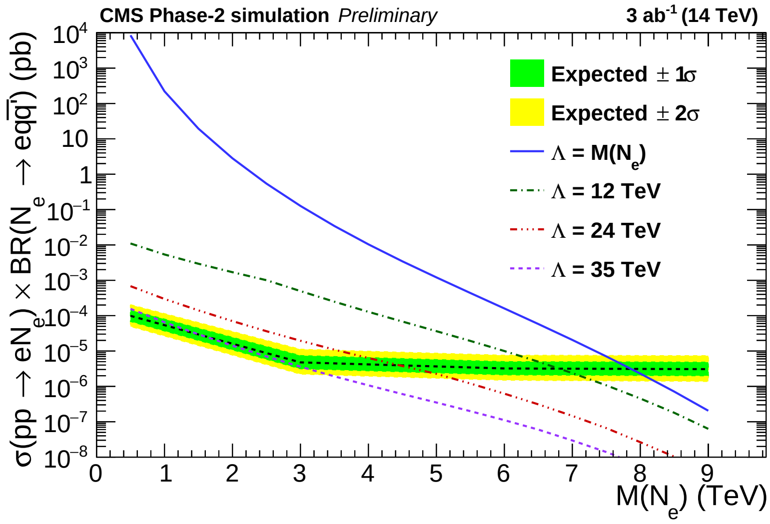


Figure 6.14: Expected central value of the limits on $\sigma \times \mathcal{BR}$ obtained for the $eeq\bar{q}'$ channel for $\Lambda = 12$ TeV for two different systematic scenarios: in red limits obtained without systematic uncertainties (i.e. only with statistical uncertainties), in blue with HL-LHC systematic uncertainties. On the left considering "background" contribution 6%, on the right 0.3%. The agreement is very good in both cases, especially in the high-mass region.



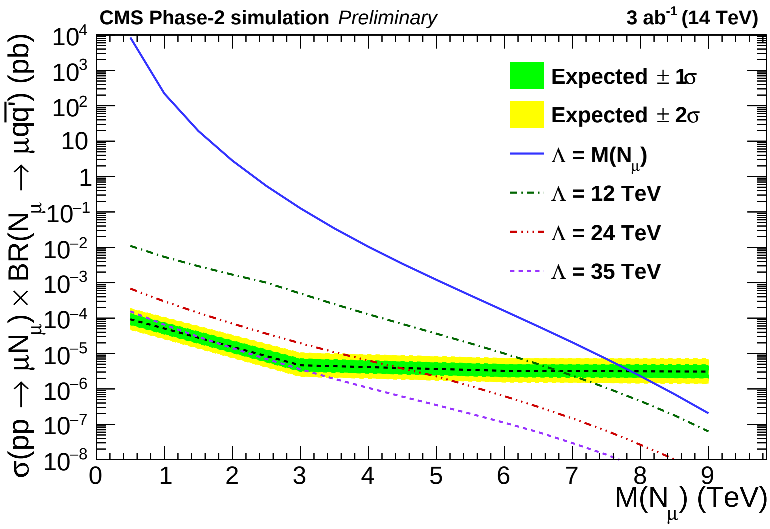


Figure 6.15: The expected 95% CL upper limits (black dotted lines) on $\sigma(pp \to \ell N_\ell) \times \mathcal{B}(N_\ell \to \ell q\bar{q}')$, obtained in the analysis of the $eeq\bar{q}'$ (top) and the $\mu\mu q\bar{q}'$ (bottom) final states, as a function of the mass of the heavy composite Majorana neutrino. The corresponding green and yellow bands represent the expected variation of the limit to one and two standard deviation(s). The solid blue curve indicates the theoretical prediction of $\Lambda = M(N_\ell)$. The textured curves give the theoretical predictions for Λ values ranging from 12 to 35 TeV.

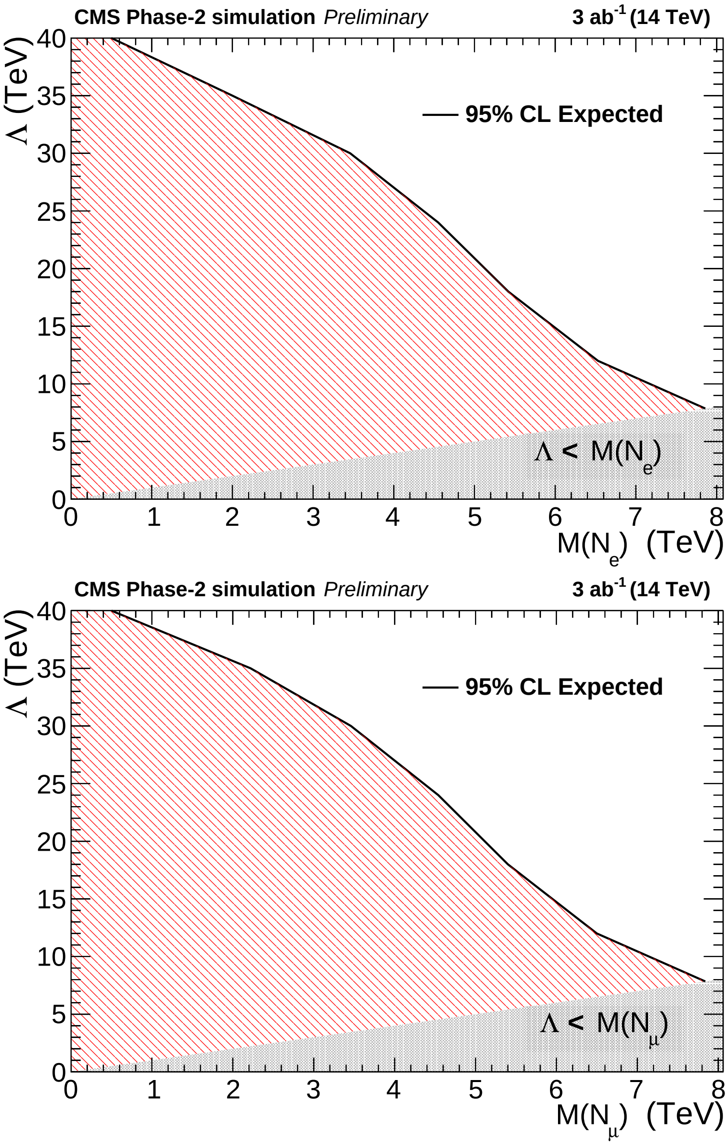


Figure 6.16: The expected 95% CL lower limits (black lines) on the compositeness scale Λ, obtained in the analysis of the $eeq\bar{q}'$ (top) and the $\mu\mu q\bar{q}'$ (bottom) final states, as a function of the mass of the heavy composite Majorana neutrino. The gray zone corresponds to the phase space $\Lambda < M(N_\ell)$ not allowed by the model.

2.6 PROSPECTS FOR THE HE-LHC

In this section we extend the study presented above to the HE-LHC, which is projected to reach a centre-of-mass energy of 27 TeV and an integrated luminosity of 15 ab^{-1}.

The MC simulation samples for the signal are generated at Leading Order (LO) with CalcHEP (v3.6) [83] for $\sqrt{s}$ = 27 TeV proton-proton collisions, using the NNPDF3.0 LO parton distribution functions with the four-flavor scheme [84]. The signal samples are generated for both the $eeq\bar{q}'$ and $\mu\mu q\bar{q}'$ channels for the following choice of parameters: $\Lambda = 10, 100$ TeV for $M(N_\ell) = 0.5, 5, 10$ TeV, $\Lambda = 25, 50$ TeV for $M(N_\ell) = 0.5, 5, 10, 12.5, 15$ TeV, and $\Lambda = M(N_\ell) \in [3, 18]$ TeV. The main background sources considered are the top quark pair production, Drell-Yan (DY) process, vector boson+jet and diboson production, and are provided by the Future Circular Collider group. For both signal and background samples the parton hadronization is treated with PYTHIA 8 and the response of the detector (assumed to perform the same as the CMS Phase-2 detector) with DELPHES.

In the spirit of a projection study, the event selection criteria remain identical to those of the HL study described in Section 6.2.3.

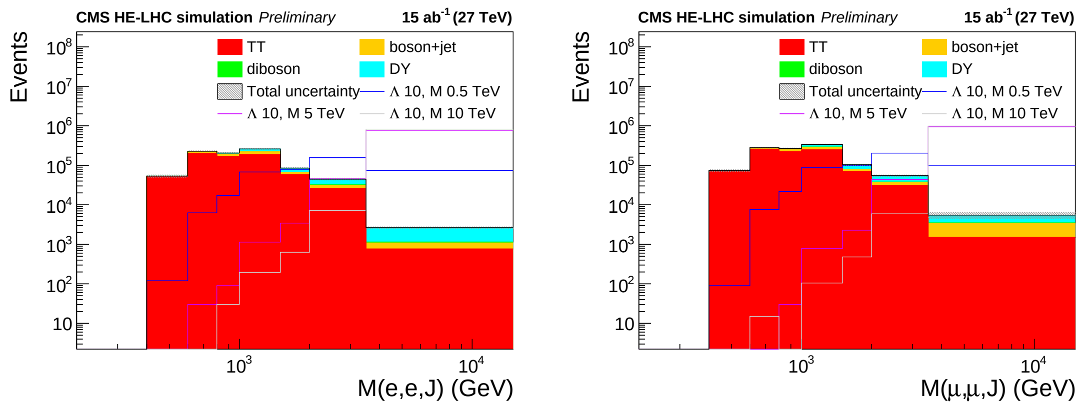


Figure 6.17: Distribution of the invariant mass of two leptons and leading large-radius jet of backgrounds and three signal samples for the $eeq\bar{q}'$ channel (left) and $\mu\mu q\bar{q}'$ channel (right) for the HE-LHC. The error bands are given by the combination of systematic and statistical uncertainties.

The expected discovery sensitivity and the exclusion limits are extracted using the $M(\ell\ell J)$ variable shown in Fig. 6.17. According to the prescriptions given in Ref. [196], the systematic uncertainties are those listed in Table 6.6. Figure 6.18 shows that with the HE-LHC we could find evidence for a composite Majorana neutrino with mass below $M(N_\ell)$ = 12 TeV, for $\Lambda = M(N_\ell)$. Upper limits on the production cross section times the branching fraction $\sigma(pp \to \ell N_\ell) \times \mathcal{B}(N_\ell \to \ell q\bar{q}')$ are shown in Fig. 6.19 for some benchmark choices of the parameters. The projection on the exclusion limits is also presented in Fig. 6.20 for the $(\Lambda, M(N_\ell))$ plane. We can conclude that, given the model condition $\Lambda = M(N_\ell)$, the HE-LHC could exclude a heavy composite Majorana neutrino with mass up to 12.5 TeV in both $eeq\bar{q}'$ and $\mu\mu q\bar{q}'$ channels.

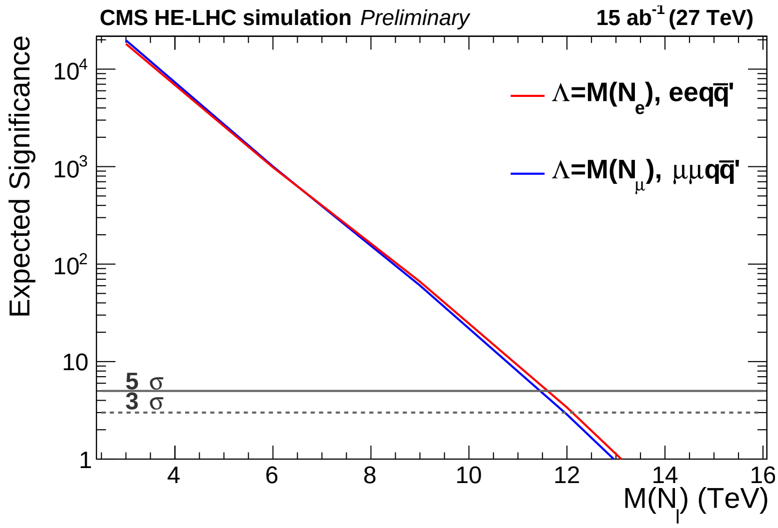


Figure 6.18: The expected statistical significance for the HE-LHC projection of the $eeq\bar{q}'$ (red line) and the $\mu\mu q\bar{q}'$ (blue line) channel for the case $\Lambda = M(N_\ell)$. The gray solid (dotted) line represents 5 (3) standard deviations, respectively.

2.7 SUMMARY OF THE RESULTS

Studies have been conducted of the expected performance at the HL-LHC and at the HE-LHC of a search for a composite Majorana neutrino. This search has been carried out within the CMS Collaboration considering a heavy composite Majorana neutrino produced in association with a lepton and decaying into a same-flavor lepton plus two quarks, with the requirement of two leptons and at least one large-radius jet in the signal region. The HL(HE)-LHC running conditions and Phase-2 detector allow a significant extension of the parameter space that can be probed.

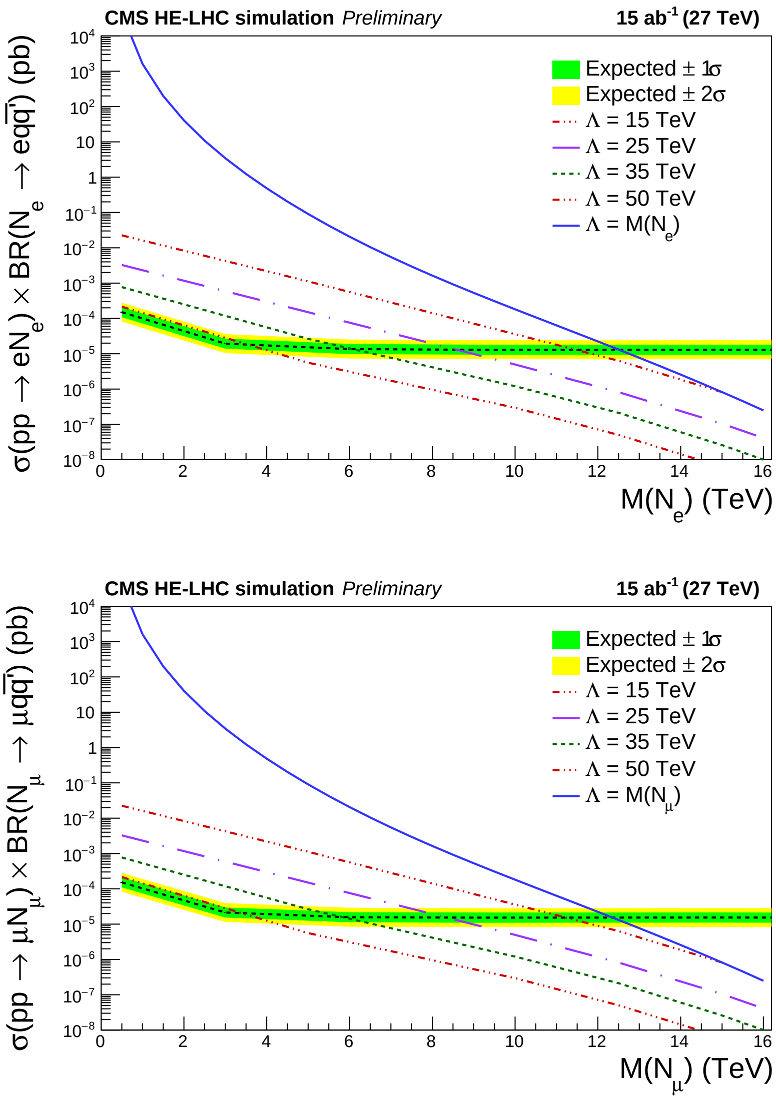


Figure 6.19: The expected 95% CL upper limits for the HE-LHC projection of the $eeq\bar{q}'$ channel (top) and the $\mu\mu q\bar{q}'$ channel (bottom). The cross section limits are higher in the HE case because of the much larger background expectation at 27 TeV.

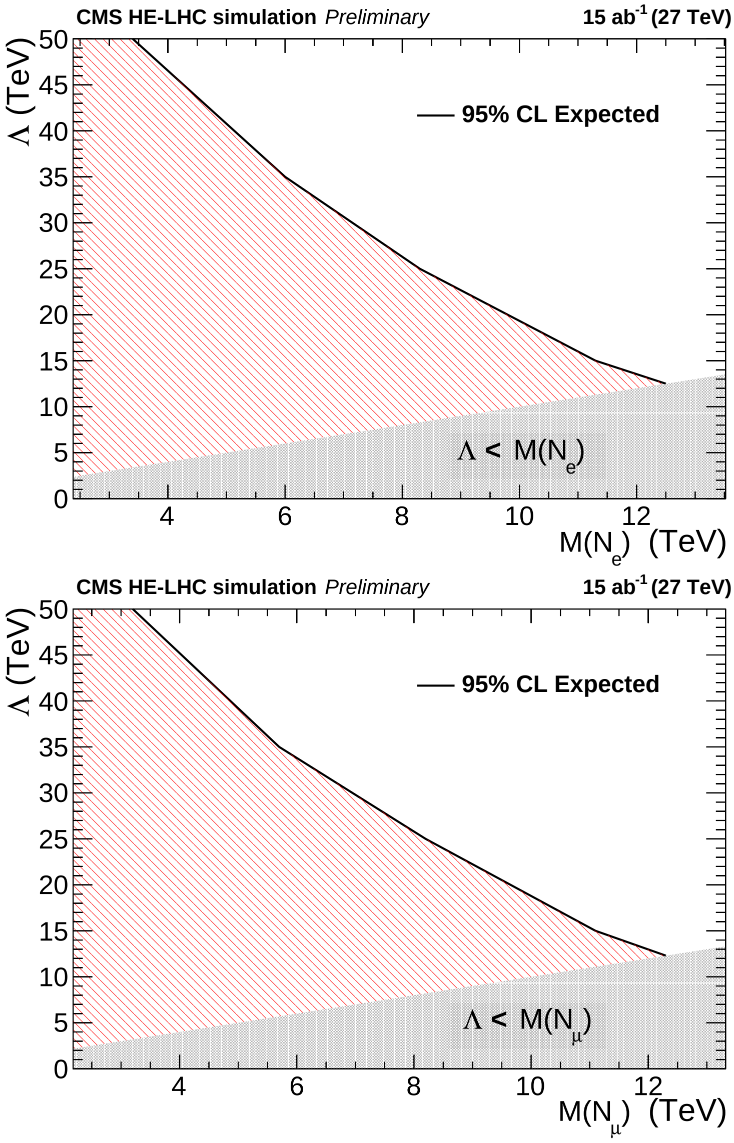


Figure 6.20: The expected 95% CL lower limits (black lines) on the compositeness scale Λ, obtained in the analysis of the $eeq\bar{q}'$ (top) and the $\mu\mu q\bar{q}'$ (bottom) final states, as a function of the mass of the heavy composite Majorana neutrino for the HE-LHC projection. The red shaded zone highlights the excluded parameter space. The gray zones are not allowed by the model.

3 TESTING LEPTOGENESIS FROM COMPOSITE HEAVY NEUTRINOS AT HE-LHC AND FCC-HH

3.1 MOTIVATION AND INTRODUCTION TO THE STUDY

In this section, we report an extract of a phenomenological study, where we have estimated the sensitivity of future hadron colliders to heavy Majorana neutrinos in composite models. In particular, we have found that this very high-mass and couplings exploration is crucial for accessing the parameters space favored by the so-called Leptogenesis mechanism. As mentioned in the first chapter, one of the mysteries of our universe is the observed density of baryons. Since there is good evidence that the universe is mostly made up of matter and no antimatter, the baryon density corresponds to the cosmological matter-antimatter asymmetry. This asymmetry is typically estimated by the ratio of the number density of baryons to photons in the universe, determined with the measurement of the angular distribution of the temperature fluctuations of the microwave background radiation [199]:

$$\eta_B = \frac{n_B - n_{\bar{B}}}{n_\gamma} \simeq \frac{n_B}{n_\gamma} = 6.19 \pm 0.15 \times 10^{-10}. \tag{6.1}$$

Within the SM dynamics, starting from a matter-antimatter symmetric initial state at high temperatures, as suggested by an inflationary phase of the very early universe, the baryon asymmetry should be much smaller. An attractive explanation of this anomalous ratio is the Leptogenesis mechanism proposed by Fukugita and Yanagida [200] for the case of a light neutrino.

During the past decade, a variety of models have been developed, which make use of the leptogenesis idea to explain the matter-antimatter asymmetry. Among these, the decays of a heavy Majorana neutrino could generate a lepton asymmetry which is partly converted to a baryon asymmetry via sphaleron processes. Indeed, composite neutrinos can be produced in the thermal bath of the early universe due to their interactions with the SM particles. The minimal requirements for a successful generation of the matter-antimatter asymmetry, namely the Sakharov conditions, have been investigated for both gauge and contact interactions of this model [70, 71]. The lepton number is violated by the presence of Majorana neutrinos in the first place. Then, the out-of-equilibrium dynamics are implemented in the usual way: inverse decays and scatterings, which tend to wash out the lepton asymmetry, become smaller than the Hubble rate and then inefficient. Therefore, the matter-antimatter asymmetry is effectively generated in the heavy neutrino decays and this usually happens when the temperature drops below the heavy neutrino mass.

We have shown that a successful generation of the matter-antimatter asymmetry can occur [71], and we have estimated the final baryon asymmetry in our model, by solving a network of rate equations that track the composite neutrino and lepton asymmetry number densities. We have studied decays and scattering processes entering the Boltzmann equations for the gauge- and contact-interactions implementation of the effective Lagrangians. Moreover, we have compared the parameter space of the model compatible with the observed baryon asymmetry with exclusion bounds from LHC, its forthcoming high-luminosity, and high-energy setting.

Since a full theoretical treatment of our results is beyond the scope of this thesis, we report only the details of the projection study in the next part, and we comment on the general outcome of the work.

3.2 FUTURE COLLIDERS PROJECTION

We consider the cases of two proposed high-energy colliders extending the current energy frontier by a factor 3 to 10 larger than the LHC: the HE-LHC [201], the possible energy upgrade of the LHC facility running at 27 TeV and with an integrated luminosity exceeding 10 ab^{-1} and presented in the previous section, and the FCC-hh [175, 202], a 100 TeV collider with an integrated luminosity of 15 ab^{-1}. While for the HE-LHC we can profit from a MC simulation study performed and validated by the CMS Collaboration, we can try to assess the FCC-hh coverage on this model with some extrapolation. Therefore, to understand the potentiality of the FCC-hh machine on a future heavy composite neutrino search, we use the 95% CL exclusion bounds on the cross-section for the di-electron di-jets process measured with 2015 Run 2 data [42] and we adopt the approach discussed in Refs. [203, 204].

The procedure, that we briefly described here, consists of a series of approximations that allow simplifying the cross-section calculation of the background processes and lies on the assumption that an upper limit on production cross section times branching ration $[\sigma \times BR](s, \mathcal{L}; M_\rho)$ entirely depends on this number of background events, specific of the signature considered in the search.

For a given collider, the number of background events at a resonance with mass M can be simply factorized as $N_{Bkg} = \mathcal{L} \cdot \sigma_B(M, s)$, and the cross section can be written as:

$$\sigma_B(M, s) \propto \sum_{i,j} \int_{M^2-\Delta\hat{s}}^{M^2+\Delta\hat{s}} d\hat{s} \frac{dL_{ij}}{d\hat{s}} \hat{\sigma}_{ij}(\hat{s}). \tag{6.2}$$

Here, the partonic cross section $\hat{\sigma}_{ij}(\hat{s})$ and the luminosity parton density

$$\frac{d\mathcal{L}_{ij}}{d\hat{s}}(\sqrt{\hat{s}}; \sqrt{s}) = \frac{1}{s} \int_{s/s}^{1} \frac{dy}{y} f_i(y; \hat{s}) f_j\left(\frac{\hat{s}}{ys}; \hat{s}\right) \tag{6.3}$$

are evaluated at a partonic centre of mass energy $\sqrt{\hat{s}}$ and refers to a pair of partons with indices i and j.

For a sufficiently narrow resonance ($M^2 \gg \Delta s^2$) supposed to be produced above the weak scale, the partonic cross sections of the background is expected to scale as $\hat{\sigma}_{ij} \propto \hat{s}$, and thus we can reasonably simplify the integral in $\hat{s}$ writing:

$$\sigma_B(M, s) \propto \frac{\Delta\hat{s}}{M^2} \sum_{i,j} \hat{s}\hat{s}_{ij} \frac{dL_{ij}}{d\hat{s}}(M, s). \tag{6.4}$$

Now we have all the ingredients to extrapolate the upper limits for a certain resonance mass point M^0, at some $\sqrt{s}^0$ collider energy and with $\mathcal{L}^0$ integrated luminosity, to a limit on an equivalent mass M' for a different collider with $\sqrt{s'}$ and $\mathcal{L}'$. We said that the limits are considered driven by it

$$N_{Bkg}\left(s, \mathcal{L}, M_\rho\right) = N_{Bkg}\left(s_0, \mathcal{L}_0, M_\rho^0\right) \tag{6.5}$$

$$L_0 \cdot \sum_{i,j} C_{ij} \frac{dL_{ij}}{d\hat{s}}\left(M_0, s_0\right) = L' \cdot \sum_{i,j} C_{ij} \frac{dL_{ij}}{d\hat{s}}\left(M', s'\right) \tag{6.6}$$

From this relation, we identify the relevant parton luminosity function at the 13 TeV LHC, read its value at the reference mass $\sqrt{\hat{s}} = M^0$ and rescale it by the ratio of the two luminosities $\mathcal{L}^0/\mathcal{L}'$.

We note that there is some arbitrariness in the starting point of the extrapolated exclusion curve since it depends on a re-scaling by the luminosity ratio $\mathcal{L}^0/\mathcal{L}'$. All the assumptions on which the method is based are rather strong, nonetheless, we have profited from the available HE-LHC limits for validation for our extrapolation to the FCC-hh. Moreover, for the present work, a rough estimate is sufficient to have a feeling of the order of magnitude of the scales Λ that can be explored in a future search at the FCC machine, whose specific characteristics are still unknown.

The results are shown in Fig. 6.21 where the upper limit expected by the 100 TeV FCC-hh collider is shown together with several signal cross sections (for the electron channel) generated at the Leading Order (LO) with CalcHEP with NNPDF3.1 [205] (assuming for the model the values commonly used by the experiments, i.e. $f = f' = 1$, $g^\star\eta = 4\pi$ and η factors equal to the unity). The upper limit on the cross-section has been then translated to a level curves plot in the 2D space (M, Λ) in Fig. 6.22, together with the limits reported in Section 6.2.6. For both the expected limits we report the saturation curve of unitarity bounds, that we can see barely touch the curves at their highest mass reach.

After the derivation of the expected exclusion limits at the forthcoming collider facilities, we can compare them with the cosmologically preferred parameter space for successful baryogenesis. In Figure 6.23, the expected exclusion limit are shown for the HL-LHC, HE-LHC and FCC collider options. To make a fair comparison, one should look at the orange level curves from the baryogenesis side since they correspond to the case $g_*^2 = 4\pi$, i.e. the choice commonly used in collider simulations.

Concluding, the studies presented indicate that the FCC-hh with 15 to 30 fb^{-1} could provide for the first time a clear coverage on the leptogenesis favored parameter space in the heavy composite Majorana neutrino scenario. Finally, we note that the exclusion curves produced by CMS for the HE-LHC start at a heavy neutrino mass of 3.5 TeV. However, extending the curve to the left we could imagine some first overlap with orange curves also at 27 TeV. Our results underline the importance of the energy frontier in the collider searches for the deep exploration of the Universe through the study of the dynamics of the EW phase transition.

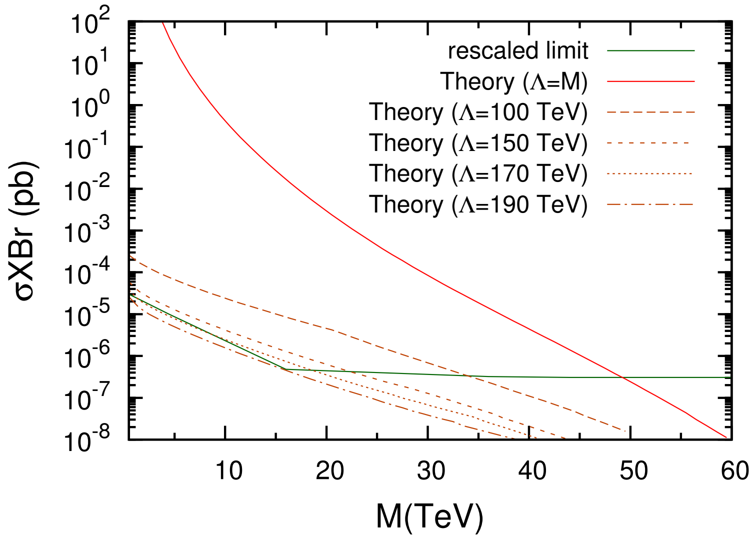


Figure 6.21: The expected 95% CL upper limits (green line) on $\sigma(pp \to eN_e) \times B(N_e \to eqq')$ (obtained with extrapolation method described in Ref. [203] for the FCC-hh) with the cross section of the process for different Λ values as a function of the mass of the heavy composite Majorana neutrino for the $eeq\bar{q}'$.

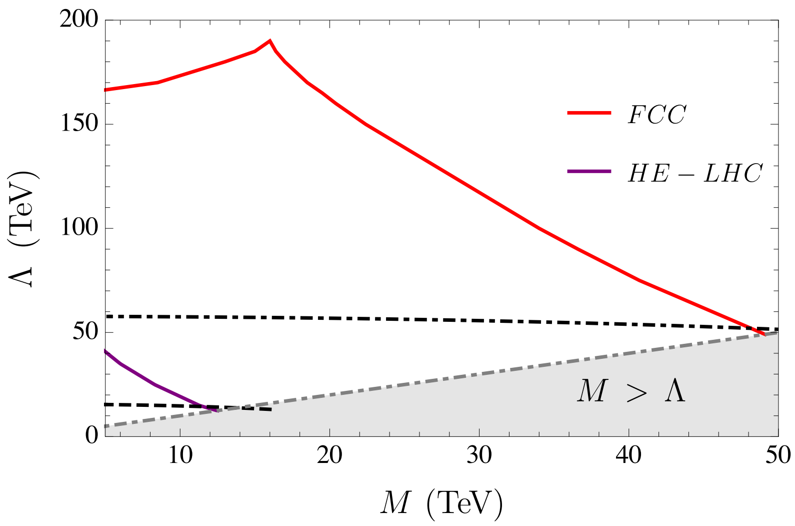


Figure 6.22: The expected exclusion limits on the (Λ, M) plane for different collider scenarios: the violet line marks the HE-LHC projection (x-axis range of the curve taken as in the official CMS publication Ref. [80]) and the ref line the FCC-hh extrapolation of the present LHC search. The black (dashed) line represents the saturation of the unitarity bound presented in Ref. 2 for the HE-LHC (FCC-hh) collider. The gray zone corresponds instead to the phase space $\Lambda < M(N_e)$ not allowed by the model.

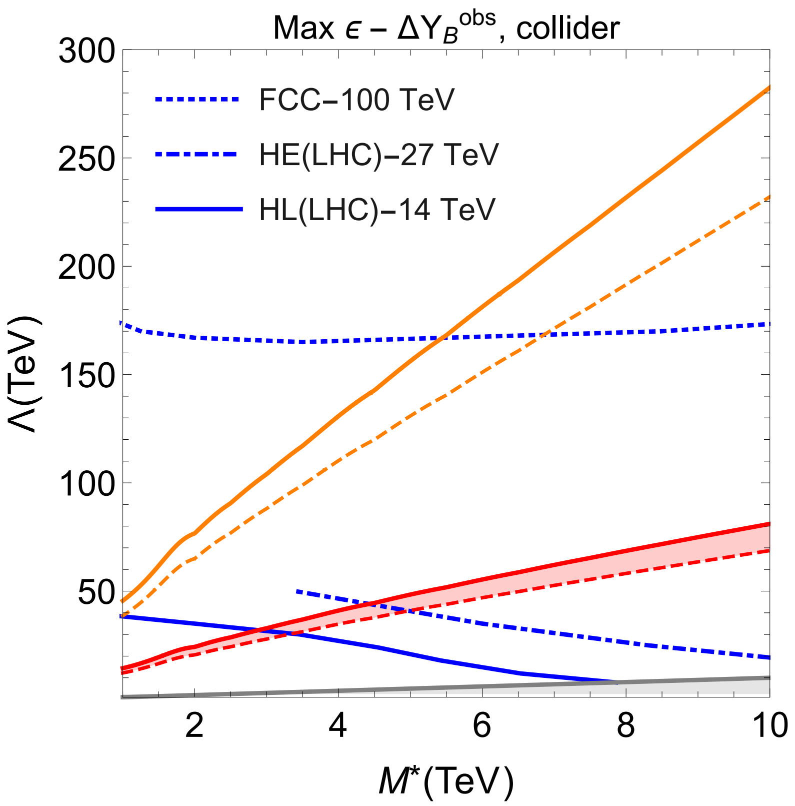


Figure 6.23: The orange and red bands represent the region where the observed baryon asymmetry is reproduced. The orange lines refer to the standard choice for the effective coupling $g_*^2 = 4\pi$, the dashed (solid) stand for $\sin(2\phi) = 1$ ($\sin(2\phi) = 0.5$).

Part IV

CONCLUSIONS

7

CONCLUSIONS

In the present Chapter, we try to summarize and show the outcomes of the main topics touched throughout this manuscript, whose content spanned from two full Run 2 data analyses and a search in the future colliders HL/HE-LHC performed within the CMS Collaboration in the final state with two leptons plus jets, and several phenomenological insights linked to the effective scenarios and unitarity violation.

The Standard Model of fundamental interactions, albeit being an incredibly successful theory, lacks explanations for some experimental and theoretical phenomena. Interestingly, many of these problems seem to be related to the electroweak symmetry breaking sector of the theory, whose dynamical generation is still unknown. Therefore, one of the main targets of the LHC is to probe the nature of the electroweak symmetry breaking. In Chapter 5, we focused on an ideal test bench for accessing the non-Abelian gauge structure of the electroweak interactions through quartic gauge couplings at hadron colliders: the Vector Boson Scattering process. Experimentally, VBS is a purely electroweak signature characterized by the exchange of weak gauge bosons in association with two forward jets with large rapidity separation and a large dijet mass. We focus on the final state with a pair of leptons plus jets, that is still lacking any evidence by both ATLAS and CMS. This semi-leptonic channel, ZVjj, considers a vector boson, V (V = W, Z), that decays in leptons and the other V, that decays in quarks. This part of the thesis sets up and perform a measurement with the full Run 2 data of the $\mathcal{O}(\alpha^6)$ electroweak ZVjj signal. The semi-leptonic final state profits from the fact that the $V \to q\bar{q}'$ branching fraction is much larger than the leptonic one, however, it suffers from larger background contamination from other processes, such as the irreducible $\mathcal{O}(\alpha^4\alpha_s^2)$ QCD background and the DY+jets processes. Two distinct categories are considered to enlarge the signal acceptance, depending on the topology of the V hadronic decay: the "boosted category" where we reconstruct a single object (an AK8 jet), and the "resolved category", when two jets (AK4) are distinct. The "V-jet mass", either meaning the boosted jet mass or the invariant mass of the two resolved AK4-jets, provides a discriminating variable for the analysis phase space and the background estimation with control regions.

A first rough estimate of the signal significance obtained with the simultaneous fit of the invariant mass of the forward jets and backgrounds in control regions shows that the ZVjj channel can reach a statistical significance of approximately 2.77 σ for the combination of the three years.

In this search, a DNN approach for signal extraction is also investigated, and compared to a more traditional BDT approach in Section 5.6. For both methods, the strategy is the same: a single model is trained for each category (Boosted or Resolved) to discriminate the signal against all backgrounds at the same time, pro-

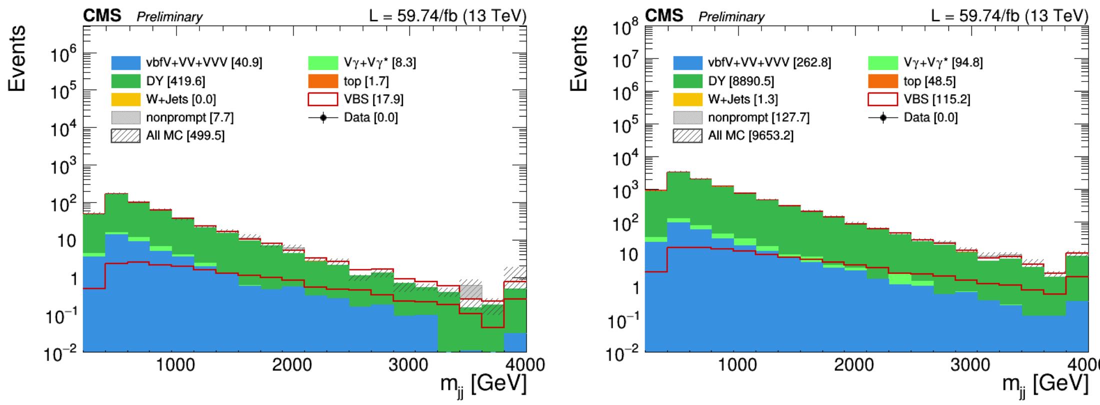


Figure 7.1: Yields of MC predictions in the 2018 signal region for the invariant mass of the two forward jets for the boosted category (left) and the resolved category (right). VBS EW process in red is both stacked and superimposed. Error bars account for both statistical and systematic uncertainty. The distributions are obtained after the "cutting" selection procedure presented in Sec. 5.5.2.

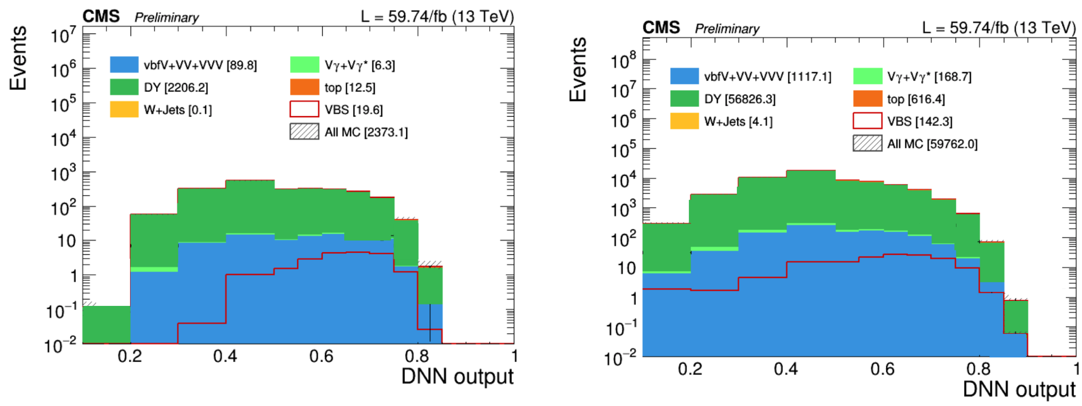


Figure 7.2: Yields of MC predictions in the 2018 signal region for the DNN output for the boosted category (left) and the resolved category (right). VBS EW process in red is both stacked and superimposed.

viding a binary output as a floating-point number in the $[0, 1]$ range (Figure 7.2 for the DNN case, with 2018 statistics).

The performances of the multivariate methods are studied in detail, and here we can mention that a first rough estimate on the statistical significance for 2018 statistics, shows an improvement of more than 30% with respect to the "cut-based" approach presented in Chapter 5. The promising findings in this work, together with the most recent expected results from the semileptonic WVjj VBS search performed in CMS [206], clearly show that the combination of WV and ZV channels could allow the very first observation of the VBS semileptonic process.

Besides the pure EW measurement, it is well known that the study of the VBS process gives primary access to the non-Abelian gauge structure of the electroweak interactions through quartic gauge couplings. Therefore, VBS analyses are sensitive to new physics at energy scales beyond the collider's direct reach through an EFT description. The second step of the VBS semileptonic analysis, which is currently on-

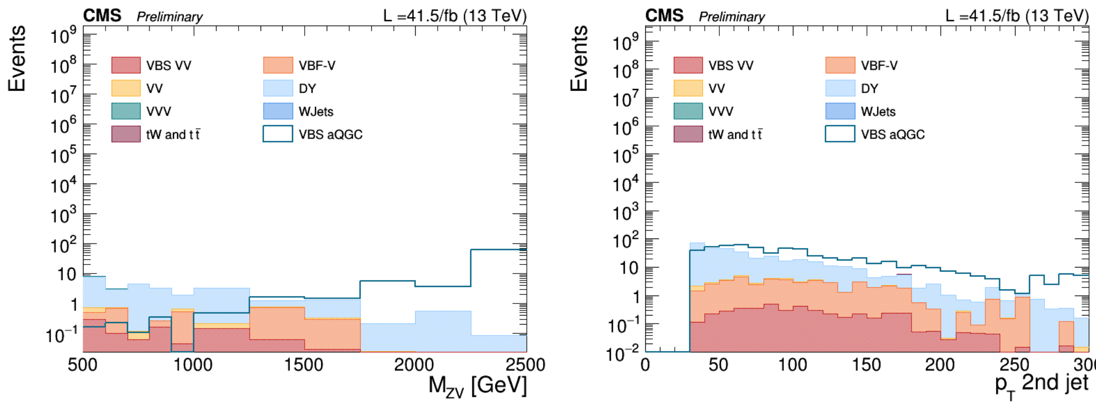


Figure 7.3: Yields of MC predictions in the EFT signal region for two relevant distributions: the invariant mass of the di-boson system (left) and the transverse momentum of the sub-leading AK4 jets (right) for 2017 Run 2 statistics.

going while we write, consists of the measurement of the aQGCs. Such an approach consists of adding higher-dimensional operators to the SM Lagrangian:

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \sum_i \frac{g_i \mathcal{O}_i}{\Lambda^{\Delta_i - 4}}, \tag{7.1}$$

where g_i are the so-called Wilson coefficients, Λ is the energy scale of the correction, and $\Delta_i \leq 4$ the dimensions of the operators, to parametrize small deviations from the original SM prediction. Within this simpler EFT scenario, commonly adopted by experimental collaboration, we could think of limiting our investigation by just looking at dimension-8 operators. For our process the most relevant ones are of the transverse type (in the *Eboli et. Al.* basis [149]):

$$\begin{aligned}
\mathcal{O}_{T_5} &= \mathrm{Tr}\left[W_{\mu\nu}W^{\mu\nu}\right] \times B_{\alpha\beta}B^{\alpha\beta} \\
\mathcal{O}_{T_6} &= \mathrm{Tr}\left[W_{\alpha\nu}W^{\mu\beta}\right] \times B_{\mu\beta}B^{\alpha\nu} \\
\mathcal{O}_{T_7} &= \mathrm{Tr}\left[W_{\alpha\nu}W^{\mu\beta}\right] \times B_{\beta\nu}B^{\nu\alpha} \\
\mathcal{O}_{T_8} &= B_{\mu\nu}B^{\mu\nu}B_{\alpha\beta}B^{\alpha\beta} \\
\mathcal{O}_{T_9} &= B_{\alpha\mu}B^{\mu\beta}B_{\beta\nu}B^{\nu\alpha}.
\end{aligned} \tag{7.2}$$

In this context, the VBS process, modified by anomalous quartic gauge couplings (aQGCs), results in an enhancement of the production cross-section at large masses of the di-boson system and high-transverse momenta, as shown in Fig. 7.3 for the 2017 luminosity.

The current most stringent limits on the Wilson coefficients have been obtained combining the results of the WV and ZV semileptonic channels with a data sample corresponding to an integrated luminosity of 35.9 fb^{-1} collected with the CMS detector in proton-proton collisions at $\sqrt{s} = 13$ TeV [144]. With the full Run 2 data sample, we could narrow the confidence limits on the Wilson coefficients up to a factor of two. This can be achieved, on one hand, with the innovations to the current analysis presented in this proposal, like the introduction of additional selection on new topological variables (such as the Zeppenfeld variable [148]). On the other hand, the increased statistic that is shown to reduce the systematic uncertainties, such as the one regarding the normalization of the dominant background source (Z+jets), which was approximately 16%.

From a more phenomenological standpoint, it is important to note that the bounds obtained in the semileptonic final state, while being the most stringent, are limited to the interpretation in terms of dimension-8 operators [149]. However, the implication of perturbative unitarity bounds on EFT couplings has shown the relevance of their inclusion in a more adequate interpretation of the experimental results. Therefore, an important part of the project, which is still ongoing, concerns the extension of the interpretation of aQGCs measurement in VBS ZV semileptonic in terms of both dimension-8 and unitarized dimension-8 operators.

In preparation for these future steps of the VBS semileptonic analysis, in Chapter 2 we have extensively discussed the Effective Field Theory "bottom-up" approach to new physics, trying to emphasize a particularly delicate feature of this powerful tool that heavily affects the interpretability of hadronic collider measurements: the unitarity violation. Indeed, on general grounds, an effective theory is valid up to energy/momentum scales smaller than the large energy scale that sets the operator expansion. Since collider experiments are involving more and more energetic particle collisions, the use and the applicability of effective operators can be questioned. We have taken the composite models, widely used in experimental research, as a limiting case of the effective theory approach. We have highlighted the limitations and provided a possible strategy through a joint phenomenological/experimental effort in the calculation of unitarity limits. We presented the partial wave unitarity bound in the parameter space of dimension-5 and dimension-6 effective operators that arise in a compositeness scenario. These are routinely used in experimental searches at the LHC to constraint contact and gauge interactions between ordinary Standard Model fermions and excited (composite) states of mass M. After deducing the unitarity bound for the production process of a composite neutrino, we implement such a bound and compare it with the recent experimental exclusion curves for Run 2 (Figure 7.4).

Comparing the unitarity bounds, as represented by the solid violet lines in Figure 7.4, with the usual prescription of $M \leq \Lambda$ we can frame the interplay between the two. For small values of the heavy neutrino masses (M), the unitarity bound is more restrictive than $M \leq \Lambda$ and it shrinks the available parameter space quite considerably (at least for $f = 100\%$ and 95%). On the other hand, for higher values of the masses, the two becomes competitive. The relative importance depends on both the collider nominal energy and event fraction f. These phenomenological findings will have a significant impact on ongoing and future experimental searches for excited states with effective interactions. At the very least, the unitarity bounds play the role of a collider-driven theoretical tool for interpreting the experimental results of the considered composite models, more rigorous than the simple relation $M \leq \Lambda$.

In Chapter 4, we have tested the same effective composite scenario in a CMS search for new physics, based on a previous 2015 2.3 fb^{-1} Run 2 publication. The search has been carried out considering a Heavy Composite Majorana Neutrino produced

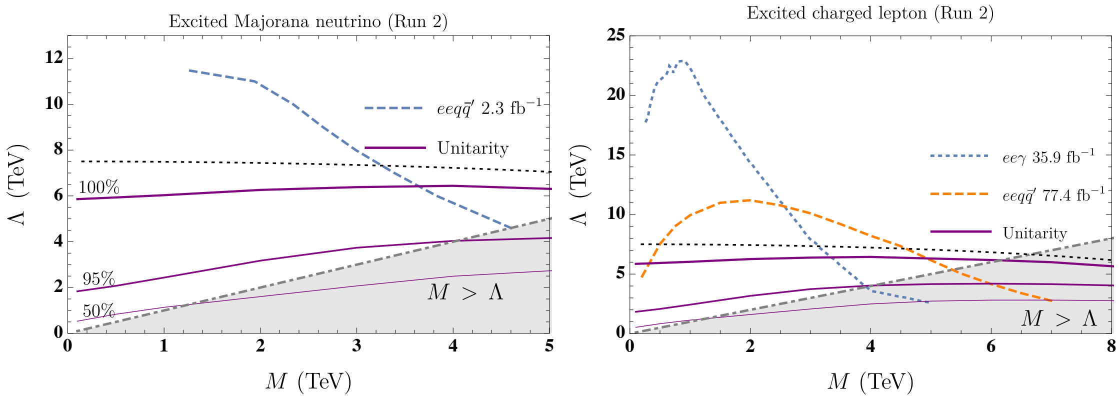


Figure 7.4: The unitarity bound in the (M, Λ) plane compared with the Run 2 exclusion at 95% CL from [42] on the left (and from [67, 68] on the right), dashed line (blue), for the $eeq\bar{q}'$ final state signature. The solid (violet) lines with decreasing thickness represent the unitarity bound respectively for 100%, 95%, and 50% event fraction satisfying Eq. 2.31. The dot-dashed (gray) line stands for the $M = \Lambda$ condition. Here and in the following figures both Λ and M start at 100 GeV, and the dotted (black) curve corresponds to the theoretical unitarity bound (Eq. 2.31 with $\hat{s} = s$).

in association with a lepton and decaying into a same-flavor lepton plus two quarks and requiring two leptons and at least one fat-jet in the signal region. We have used the 2016, 2017, 2018 dataset corresponding to an integrated luminosity of 136 fb^{-1} collected by the CMS detector in pp collisions at $\sqrt{s} = 13$ TeV. Besides being a pure extension of a previously performed search, for the very first time, the optimization of the analysis phase space was guided by phenomenological constraints on the validity of the effective interactions in composite models. The constraints on the parameter space given by the unitarity conditions on the Lagrangian of the model presented in Sec. 4.1.1 suggested reconsidering the optimization of the signal region. The target was not only to reach higher Heavy Composite Neutrino (HCN) masses but also to extend the sensitivity towards higher values of Λ (and small neutrino masses), being that region the safest to avoid unphysical predictions from the EFT interactions of the model. We can see the effect of our optimized selection in a comparison with the previous search analysis strategy in the explored parameter space for a 2016 luminosity projection (Figure 7.5).

The observations are in fair agreement with the SM predictions and will be examined further as part of the future development of this data analysis. A CL_s technique is then used to extract the expected upper limits on the product of the cross-section and branching fraction of the signal process, using the invariant mass of the two leptons and the leading fat jet, $M(\ell, \ell, J)$. The expected limits are summarized in the 2-dimensional plane $(M(N), \Lambda)$ for a comparison with the unitarity restrictions on the parameters space. Results are shown in Fig. 7.6 for both the channels. We can see that the accessible compositeness scale domain is almost double than that in the 2015 search (order of 20 TeV for the full Run 2 versus the order of 10 TeV for 2015), and the mass reach is increased by a factor of 0.5. Moreover, and most importantly, the explored region is safe from unitarity violation in almost all the mass range.

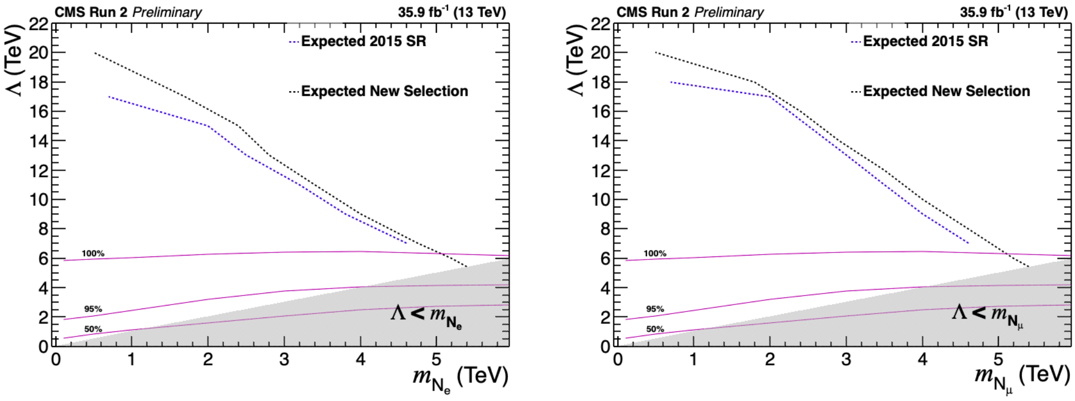
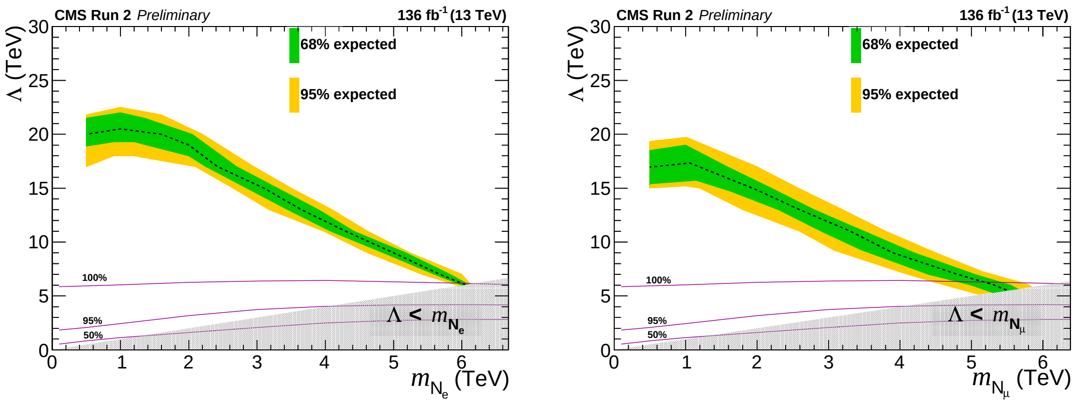


Figure 7.5: Comparison on the explored parameter space with statistical extrapolation to 2016 statistics between the present optimization and the 2015 published results analysis strategy [42]. Expectations are compared with the 2016 luminosity.



z

Figure 7.6: Exclusion limit for the *eeqq* (left) and $\mu\mu qq$ (right) channel for the combination of the three years of data taking, 2016, 2017, and 2018, in the 2D plane. The solid (violet) lines represent the fraction of MC events that satisfy the unitarity condition in the EFT approximation [110] with the various percentages considered.

The Third Part of the thesis (Chapter 6) is dedicated to the future of the LHC, and its successors: the HL-LHC, the HE-LHC, and the FCC.

The sensitivity of a search for heavy Majorana neutrinos with the CMS Phase-2 detector in a final state with two leptons and at least one large-radius jet is investigated. The study is based on the previous searches with Run 2 CMS data. The High-Luminosity LHC (HL-LHC) with a center-of-mass energy of 14 TeV and an integrated luminosity of 3 ab^{-1} will allow an extension of the sensitivity to cross-sections of the order of a few ab for heavy neutrino masses $M(N_\ell)$ ranging from 3 to 9 TeV for the $\ell\ell q\bar{q}'$ channel, where ℓ is an electron or a muon and q is a quark. For the compositeness scale $\Lambda = M(N_\ell)$, the existence of a heavy Majorana neutrino could be excluded for masses up to 8 TeV at the 95% confidence level. The projection of the study to the High-Energy LHC (HE-LHC) scenario, with a center-of-mass energy of 27 TeV, is also presented here. Results are summarized in Figure 7.7.

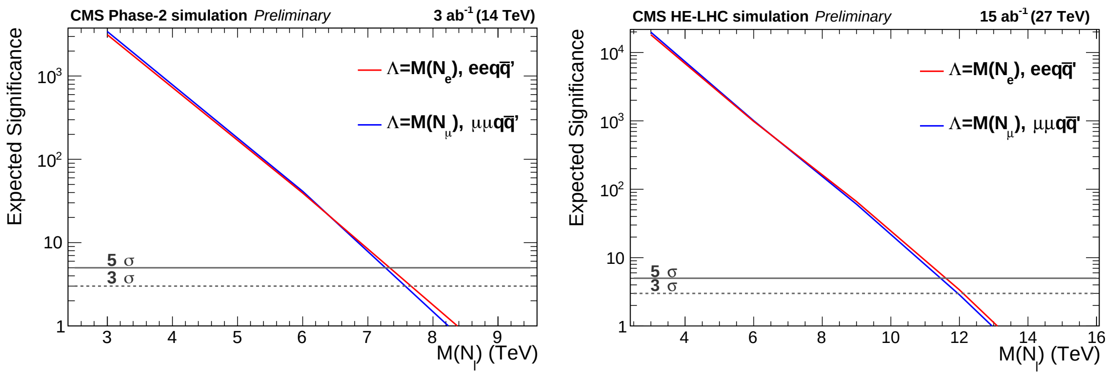


Figure 7.7: The expected statistical significance for the HL-LHC analysis (HE-LHC projection) on the left (right) for both the $eeq\bar{q}'$ (red line) and $\mu\mu q\bar{q}'$ (blue line) channel for the case $\Lambda = M(N_\ell)$. The gray solid (dotted) line represents 5 (3) standard deviations, respectively.

Finally, we have concluded projecting the search to the highest energy likely reachable by the end of the century, with the FCC-hh hadron collider at 100 TeV. The studies presented indicate that the FCC-hh with 15 to 30 fb^{-1} could provide for the first time a clear cover on the leptogenesis favored parameter space in the heavy composite Majorana neutrino scenario, pointing out the importance of the Energy Frontier for the deep exploration of the Universe through the study of the dynamics of the EW phase transition (Fig. 7.8).

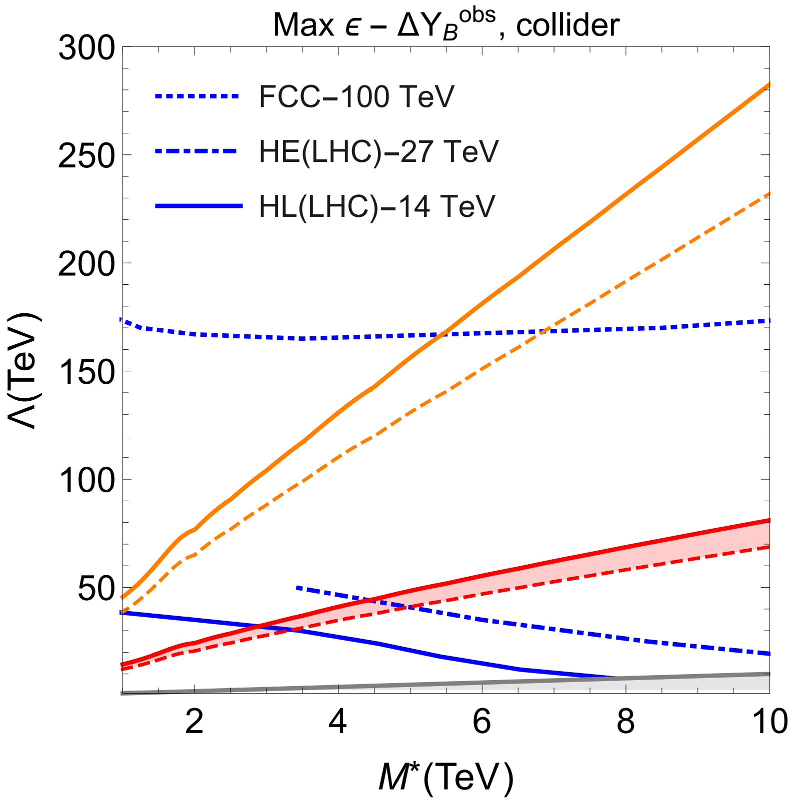


Figure 7.8: The orange and red bands represent the region where the observed baryon asymmetry is reproduced. The orange lines refer to the standard choice for the effective coupling $g_*^2 = 4\pi$, the dashed (solid) stand for $\sin(2\phi) = 1$ ($\sin(2\phi) = 0.5$).

BIBLIOGRAPHY

[1] M. J. G. Veltman. *Diagrammatica: The Path to Feynman rules*, volume 4. Cambridge University Press, 5 2012.

[2] Standard model lagrangian (density). http://nuclear.ucdavis.edu/~tgutierr/files/stmL1.html.

[3] Carlo Giunti and Kim Chung Wook. *Fundamentals of Neutrino Physics and Astrophysics*. Oxford Univ., Oxford, 2007.

[4] P. Langacker. Grand unification. *Scholarpedia*, 7(10):11419, 2012. revision #137135.

[5] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[6] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[7] Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012.

[8] Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, May 2015.

[9] Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012.

[10] M. Tanabashi et al. Review of Particle Physics. *Phys. Rev.*, D98(3):030001, 2018.

[11] A. Gando et al. Limit on neutrinoless $\beta\beta$ decay of ^{136}Xe from the first phase of kamland-zen and comparison with the positive claim in ^{76}Ge. *Phys. Rev. Lett.*, 110:062502, Feb 2013.

[12] S. M. Bilenky and S. T. Petcov. Massive neutrinos and neutrino oscillations. *Rev. Mod. Phys.*, 59:671–754, Jul 1987.

[13] S. Bilenky. *Introduction to the Physics of Massive and Mixed Neutrinos*. Lecture Notes in Physics. Springer Berlin Heidelberg, 2010.

[14] Csaba Csáki, Salvator Lombardo, and Ofri Telem. *TASI Lectures on Non-supersymmetric BSM Models*, pages 501–570. WSP, 2018.

[15] Th. Kaluza. Zum Unitätsproblem der Physik. *Int. J. Mod. Phys. D*, 27(14):1870001, 2018.

[16] A. J. Buras, John R. Ellis, M. K. Gaillard, and Dimitri V. Nanopoulos. Aspects of the Grand Unification of Strong, Weak and Electromagnetic Interactions. *Nucl. Phys. B*, 135:66–92, 1978.

[17] R. N. Mohapatra. *Unification and supersymmetry: The Frontiers of Quark-Lepton Physics*. Springer, Berlin, 1986.

[18] Ignatios Antoniadis, Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. New dimensions at a millimeter to a Fermi and superstrings at a TeV. *Phys. Lett. B*, 436:257–263, 1998.

[19] Lisa Randall and Raman Sundrum. An Alternative to compactification. *Phys. Rev. Lett.*, 83:4690–4693, 1999.

[20] Gian F. Giudice, Riccardo Rattazzi, and James D. Wells. Quantum gravity and extra dimensions at high-energy colliders. *Nucl. Phys. B*, 544:3–38, 1999.

[21] H. Georgi. Effective field theory. *Ann. Rev. Nucl. Part. Sci.*, 43:209–252, 1993.

[22] J.F. Donoghue, E. Golowich, and Barry R. Holstein. *Dynamics of the standard model*, volume 2. CUP, 2014.

[23] Thomas Appelquist and J. Carazzone. Infrared singularities and massive fields. *Phys. Rev. D*, 11:2856–2861, May 1975.

[24] Simone Biondini. *Effective field theories for heavy Majorana neutrinos in a thermal bath*. PhD thesis, Munich, Tech. U., 2016.

[25] W. Buchmuller and D. Wyler. Effective Lagrangian Analysis of New Interactions and Flavor Conservation. *Nucl. Phys. B*, 268:621–653, 1986.

[26] André David and Giampiero Passarino. Use and reuse of SMEFT. 8 2020.

[27] Pawel Kozów, Luca Merlo, Stefan Pokorski, and Michal Szleper. Same-sign WW Scattering in the HEFT: Discoverability vs. EFT Validity. *JHEP*, 07:021, 2019.

[28] Ilaria Brivio and Michael Trott. The Standard Model as an Effective Field Theory. *Phys. Rept.*, 793:1–98, 2019.

[29] Roberto Contino and Geraldine Servant. Discovering the top partners at the LHC using same-sign dilepton final states. *JHEP*, 0806:026, 2008.

[30] E. Eichten, Kenneth D. Lane, and Michael E. Peskin. New Tests for Quark and Lepton Substructure. *Phys. Rev. Lett.*, 50:811–814, 1983. [,369(1983)].

[31] M.E. Peskin. International Symposium on Lepton Photon Interactions at High Energies, 1985.

[32] N. Cabibbo, L. Maiani, and Y. Srivastava. Anomalous Z Decays: Excited Leptons? *Phys. Lett.*, 139B:459–463, 1984.

[33] U. Baur, M. Spira, and P. M. Zerwas. Excited Quark and Lepton Production at Hadron Colliders. *Phys. Rev.*, D42:815–824, 1990.

[34] H. Terazawa, K. Akama, and Y. Chikashige. Unified Model of the Nambu-Jona-Lasinio Type for All Elementary Particle Forces. *Phys. Rev.*, D15:480, 1977.

[35] Hidezumi Terazawa. Subquark Model of Leptons and Quarks. *Phys.Rev.*, D22:184, 1980.

[36] Hidezumi Terazawa and Masaki Yasue. Composite Higgs Boson in the Unified Subquark Model of All Fundamental Particles and Forces. 2014.

[37] K. Nakamura et al. Review of particle physics. *J.Phys.*, G37:075021, 2010.

[38] Albert M Sirunyan et al. Search for contact interactions and large extra dimensions in the dilepton mass spectra from proton-proton collisions at $\sqrt{s} =$ 13 TeV. *JHEP*, 04:114, 2019.

[39] Morad Aaboud et al. Search for new high-mass phenomena in the dilepton final state using 36 fb^1 of proton-proton collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 10:182, 2017.

[40] K.A. Olive et al. Review of Particle Physics. *Chin.Phys.*, C38:090001, 2014.

[41] Eiichi Takasugi. Composite neutrinos and double beta decay. *Prog. Theor. Phys.*, 94:1097–1104, 1995.

[42] A. M. Sirunyan et al. Search for a heavy composite Majorana neutrino in the final state with two leptons and two quarks at $\sqrt{s} = 13$ TeV. *Phys. Lett.*, B775:315–337, 2017.

[43] Giulia Pancheri and Y.N. Srivastava. Introduction to the physics of the total cross-section at LHC: A Review of Data and Models. *Eur. Phys. J. C*, 77(3):150, 2017.

[44] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3 2014.

[45] Luca Di Luzio, Jernej F. Kamenik, and Marco Nardecchia. Implications of perturbative unitarity for scalar di-boson resonance searches at lhc. *The European Physical Journal C*, 77(1):30, Jan 2017.

[46] Benjamin W. Lee, C. Quigg, and H. B. Thacker. The Strength of Weak Interactions at Very High-Energies and the Higgs Boson Mass. *Phys. Rev. Lett.*, 38:883–885, 1977.

[47] Benjamin W. Lee, C. Quigg, and H. B. Thacker. Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass. *Phys. Rev.*, D16:1519, 1977.

[48] Benjamin W. Lee, C. Quigg, and H. B. Thacker. Weak interactions at very high energies: The role of the higgs-boson mass. *Phys. Rev. D*, 16:1519–1531, Sep 1977.

[49] Michael Atkins and Xavier Calmet. Unitarity bounds on low scale quantum gravity. *Eur. Phys. J.*, C70:381–388, 2010.

[50] Stefania De Curtis, Stefano Moretti, Kei Yagyu, and Emine Yildirim. Perturbative unitarity bounds in composite two-Higgs doublet models. *Phys. Rev.*, D94(5):055017, 2016.

[51] Luca Di Luzio, Jernej F. Kamenik, and Marco Nardecchia. Perturbative unitarity bounds on di-boson scalar resonances. *EPJ Web Conf.*, 164:07026, 2017.

[52] Motoi Endo and Yasuhiro Yamamoto. Unitarity Bounds on Dark Matter Effective Interactions at LHC. *JHEP*, 06:126, 2014.

[53] Tyler Corbett, O. J. P. Éboli, and M. C. Gonzalez-Garcia. Unitarity Constraints on Dimension-six Operators II: Including Fermionic Operators. *Phys. Rev.*, D96(3):035006, 2017.

[54] Jogesh C. Pati, Abdus Salam, and J. A. Strathdee. Are Quarks Composite? *Phys. Lett.*, 59B:265–268, 1975.

[55] Jogesh C. Pati and Abdus Salam. Supersymmetry at the preonic or prepreonic level and composite supergravity. *Nucl. Phys.*, B214:109–135, 1983.

[56] Haim Harari. Composite Models for Quarks and Leptons. *Phys. Rept.*, 104:159, 1984. [,334(1982)].

[57] O. W. Greenberg and C. A. Nelson. Composite Models of Leptons. *Phys. Rev.*, D10:2567, 1974.

[58] P. A. M. Dirac. The evolution of the physicist's picture of nature. *Scientific American*, 208(5):45–53, 1963.

[59] Estia Eichten and Kenneth D. Lane. Dynamical Breaking of Weak Interaction Symmetries. *Phys. Lett.*, 90B:125–130, 1980.

[60] U. Baur, I. Hinchliffe, and D. Zeppenfeld. Excited Quark Production at Hadron Colliders. *Int. J. Mod. Phys.*, A2:1285, 1987.

[61] Search for excited leptons in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Rev. D*, 85:072003, 2012.

[62] Search for excited electrons and muons in $\sqrt{s} = 8$ TeV proton-proton collisions with the ATLAS detector. *New J. Phys.*, 15:093011, 2013.

[63] S. Chatrchyan et al. Search for excited leptons in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 720:309, 2013.

[64] Vardan Khachatryan et al. Search for excited leptons in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 03:125, 2016.

[65] Vardan Khachatryan et al. Search for excited leptons in the $\ell\ell\gamma$ final state at $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-EXO-16-009, CERN, Geneva, 2016.

[66] Search for excited leptons in the $\ell\ell\gamma$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-EXO-18-004, CERN, Geneva, 2018.

[67] Albert M Sirunyan et al. Search for excited leptons in $\ell\ell\gamma$ final states in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 2019.

[68] Search for excited leptons decaying via contact interaction to two leptons and two jets. Technical Report CMS-PAS-EXO-18-013, CERN, Geneva, 2019.

[69] R. Leonardi, L. Alunni, F. Romeo, L. Fanò, and O. Panella. Hunting for heavy composite Majorana neutrinos at the LHC. *Eur. Phys. J.*, C76(11):593, 2016.

[70] Dmitry Zhuridov. Baryogenesis from leptomesons. *Phys. Rev.*, D94(3):035007, 2016.

[71] Simone Biondini and Orlando Panella. Leptogenesis and composite heavy neutrinos with gauge mediated interactions. *Eur. Phys. J.*, C77(9):644, 2017.

[72] S. Biondini, O. Panella, G. Pancheri, Y. N. Srivastava, and L. Fanò. Phenomenology of excited doubly charged heavy leptons at LHC. *Phys. Rev.*, D85:095018, 2012.

[73] Swapan Majhi. QCD corrections to excited lepton (pair) production at the LHC. *Phys. Rev.*, D88(7):074028, 2013.

[74] R. Leonardi, O. Panella, and L. Fanò. Doubly charged heavy leptons at LHC via contact interactions. *Phys. Rev.*, D90(3):035001, 2014.

[75] S. Biondini and O. Panella. Exotic leptons at future linear colliders. *Phys. Rev.*, D92(1):015023, 2015.

[76] O. Panella, R. Leonardi, G. Pancheri, Y. N. Srivastava, M. Narain, and U. Heintz. Production of exotic composite quarks at the LHC. *Phys. Rev.*, D96(7):075034, 2017.

[77] Matteo Presilla, Roberto Leonardi, and Orlando Panella. Like-sign dileptons with mirror type composite neutrinos at the HL-LHC. 2018.

[78] Abdullatif Caliskan and Seyit Okan Kara. Single production of the excited electrons in the future FCC-based lepton–hadron colliders. *Int. J. Mod. Phys.*, A33(24):1850141, 2018.

[79] M. Jacob and G. C. Wick. On the general theory of collisions for particles with spin. *Annals Phys.*, 7:404–428, 1959. [Annals Phys.281,774(2000)].

[80] Search for heavy composite Majorana neutrinos at the High-Luminosity and the High-Energy LHC. Technical Report CMS-PAS-FTR-18-006, CERN, Geneva, 2018.

[81] X. Cid Vidal et al. Beyond the Standard Model Physics at the HL-LHC and HE-LHC. 2018.

[82] ATLAS and CMS Collaborations. Report on the Physics at the HL-LHC and Perspectives for the HE-LHC. Technical Report arXiv:1902.10229, CERN, Geneva, 2019.

[83] Alexander Belyaev, Neil D. Christensen, and Alexander Pukhov. CalcHEP 3.4 for collider physics within and beyond the Standard Model. *Comput. Phys. Commun.*, 184:1729–1769, 2013.

[84] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.

[85] Eric Conte, Benjamin Fuks, and Guillaume Serret. MadAnalysis 5, A User-Friendly Framework for Collider Phenomenology. *Comput. Phys. Commun.*, 184:222–256, 2013.

[86] Maximilien Brice. Aerial View of the CERN taken in 2008. Jul 2008.

[87] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[88] A. Augusto Alves, Jr. et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[89] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[90] Tai Sakuma and Thomas McCauley. Detector and Event Visualization with SketchUp at the CMS Experiment. *J. Phys. Conf. Ser.*, 513:022032, 2014.

[91] Overview of cms physics goals and detector. https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSExperiment.

[92] Joel Nathan Butler. Highlights and Perspectives from the CMS Experiment. In *5th Large Hadron Collider Physics Conference*, 9 2017.

[93] Danyer Perez Adan. Search for Light Bosons in Exotic Decays of the 125 GeV Higgs Boson, Jun 2020. Presented 05 05 2020.

[94] R. Leonardi. Phenomenology and experimental search of excited composite fermions at Run II of the LHC, Apr 2015.

[95] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[96] The CMS tracker: addendum to the Technical Design Report. 2000.

[97] The CMS electromagnetic calorimeter project: Technical Design Report. 1997.

[98] CMS: The hadron calorimeter technical design report. 6 1997.

[99] G. Acquistapace et al. CMS, the magnet project: Technical design report. 5 1997.

[100] The CMS muon project: Technical Design Report. 1997.

[101] O. Panella, C. Carimalo, and Y. N. Srivastava. Production of like sign dileptons in pp collisions through composite majorana neutrinos. *Phys. Rev.*, D62:015013, 2000.

[102] O. Panella, C. Carimalo, Y. N. Srivastava, and A. Widom. Neutrinoless double beta decay with composite neutrinos. *Phys.Rev.*, D56:5766–5775, 1997.

[103] O. Panella and Y. N. Srivastava. Bounds on compositeness from neutrinoless double beta decay. *Phys.Rev.*, D52:5308–5313, 1995.

[104] R. Leonardi , L. Alunni, F. Romeo, L. Fano', and O. Panella. Hunting for heavy composite Majorana neutrinos at the LHC. *Submitted on EPJC*, 2015.

[105] H. Terazawa. t quark mass predicted from a sum rule for lepton and quark masses. *Phys. Rev.*, D22:2921–2921.

[106] E. Eichten, Kenneth D. Lane, and Michael E. Peskin. New Tests for Quark and Lepton Substructure. *Phys. Rev. Lett.*, 50:811–814, 1983.

[107] N. Cabibbo, L. Maiani, and Y. Srivastava. Anomalous Z Decays: Excited Leptons? *Phys. Lett.*, B139:459, 1984.

[108] U. Baur, M. Spira, and P. M. Zerwas. Excited Quark and Lepton Production at Hadron Colliders. *Phys. Rev.*, D42:815–824, 1990.

[109] A.M. Sirunyan et al. Search for a heavy composite Majorana neutrino in the final state with two leptons and two quarks at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 775:315–337, 2017.

[110] S. Biondini, R. Leonardi, O. Panella, and M. Presilla. Perturbative unitarity bounds for effective composite models. *Phys. Lett.*, B795:644–649, 2019. [Erratum: Phys. Lett.B799,134990(2019)].

[111] How to work with files for good luminosity sections in json format. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideGoodLumiSectionsJSONFile.

[112] Albert M Sirunyan et al. Search for an excited lepton that decays via a contact interaction to a lepton and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV. 2020.

[113] Search for high mass resonances in dielectron final state. Technical Report CMS-PAS-EXO-18-006, CERN, Geneva, 2018.

[114] Lumi group. Reference guidelines. https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData.

[115] CMS. Miniaod. Technical report. https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMiniAOD2015#High_level_physics_objects.

[116] Egamma POG. Heep electron identification in run 2. Technical report. https://twiki.cern.ch/twiki/bin/view/CMS/HEEPElectronIdentificationRun2.

[117] Egamma POG. Egamma run 2 recommendation. Technical report. https://twiki.cern.ch/twiki/bin/view/CMS/EgammaRunIIRecommendations#Electron_Scale_Factors.

[118] Electron POG. Cut based electron identification. https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2.

[119] Muon POG. Muon identification and isolation. https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2.

[120] Muon POG. High-pt muon references for run 2. https://twiki.cern.ch/twiki/bin/view/CMS/HighPtMuonReferenceRun2.

[121] Muon POG. Rochester correction. https://twiki.cern.ch/twiki/bin/view/CMS/RochcorMuon.

[122] Jet POG. Jet identification for run 2 - 2016. https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016,.

[123] Jet POG. Jet identification for run 2 - 2017. https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2017,.

[124] Jet POG. Jet identification for run 2 - 2018. https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2018,.

[125] Jet POG. Recommended jet energy corrections and uncertainties for data and mc. https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC.

[126] Jet POG. Jet energy resolution. https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution.

[127] J.M. Lindert et al. Precise predictions for $V+$ jets dark matter backgrounds. *Eur. Phys. J. C*, 77(12):829, 2017.

[128] CMS Collaboration. Run 2 monojet / mono-V analysis. CMS AN-2020/042.

[129] $Z' \to ee$ group. Search for high mass di-electron resonances with full 2016 data (AN-16-404).

[130] $Z' \to \mu\mu$ group. Search for high mass resonances decaying to muon pairs in pp collisions at $\sqrt{s} = 13$ TeV with full 2016 data set of 37 fb^{-1} and combination with 2015 result (AN-16-391).

[131] CMS Collaboration. Cms luminosity measurements for the 2016 data taking period. CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

[132] CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-LUM-17-004, CERN, Geneva, 2018.

[133] CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-LUM-18-002, CERN, Geneva, 2019.

[134] CMS Collaboration. Documentation of the RooStats-based statistics tools for Higgs PAG. https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHiggsAnalysisCombinedLimit.

[135] B. Clerbaux, W. Fang, R. Goldouzian. Heep selection efficiency scale factor study for cms with the full 2016 data sample. CMS AN 2017/077, 2017.

[136] EGamma POG. Energy scale and smearing. https://twiki.cern.ch/twiki/bin/view/CMS/EgammaMiniAODV2.

[137] C.G. Willis, R. Brock, D. Hayden, T.-J. Hou, J. Isaacson, C. Schmidt, and C.-P. Yuan. New method for reducing parton distribution function uncertainties in the high-mass drell-yan spectrum. *Physical Review D*, 99(5), Mar 2019.

[138] Jon Butterworth et al. PDF4LHC recommendations for LHC Run 2. *J. Phys. G*, 43:023001, 2016.

[139] Alexander L. Read. Presentation of search results: The $\mathrm{CL_s}$ technique. *J. Phys. G*, 28:2693, 2002.

[140] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435, 1999.

[141] Daniel R. Green, Patrick Meade, and Marc-Andre Pleier. Multiboson interactions at the LHC. *Rev. Mod. Phys.*, 89(3):035008, 2017.

[142] CMS Collaboration. Observation of electroweak production of same-sign W boson pairs in the two jet and two same-sign lepton final state in proton-proton collisions at 13 TeV. Technical Report CMS-PAS-SMP-17-004, CERN, Geneva, 2017.

[143] ATLAS Collaboration. Observation of electroweak production of a same-sign W boson pair in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2018-030, CERN, Geneva, Jul 2018.

[144] CMS Collaboration. Search for anomalous electroweak production of WW/WZ/ZZ boson pairs in association with two jets in proton-proton collisions at 13 TeV. Technical Report CMS-PAS-SMP-18-006, CERN, Geneva, 2018.

[145] C. F. Anders et al. Vector boson scattering: Recent experimental and theory developments. *Rev. Phys.*, 3:44–63, 2018.

[146] ATLAS Collaboration. Search for anomalous electroweak production of WW/WZ in association with a high-mass dijet system in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev.*, D95(3):032001, 2017.

[147] Albert M Sirunyan et al. Measurement of vector boson scattering and constraints on anomalous quartic couplings from events with four leptons and two jets in proton–proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett.*, B774:682–705, 2017.

[148] David L. Rainwater, R. Szalapski, and D. Zeppenfeld. Probing color singlet exchange in Z + two jet events at the CERN LHC. *Phys. Rev.*, D54:6680–6689, 1996.

[149] O. J. P. Eboli, M. C. Gonzalez-Garcia, and J. K. Mizukoshi. p p $\to$ j j e+- mu+- nu nu and j j e+- mu-+ nu nu at O(alpha(em)**6) and O(alpha(em)**4 alpha(s)**2) for the study of the quartic electroweak gauge boson vertex at CERN LHC. *Phys. Rev.*, D74:073005, 2006.

[150] Raquel Gomez-Ambrosio. Studies of Dimension-Six EFT effects in Vector Boson Scattering. *Eur. Phys. J.*, C79(5):389, 2019.

[151] Common analysis object definitions and trigger efficiencies for the $H \to WW$ analysis with full Run-II data. CMS Analysis Note CMS-AN-2019/125, 2019.

[152] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching nlo qcd computations with parton shower simulations: the powheg method. *Journal of High Energy Physics*, 2007(11):070–070, Nov 2007.

[153] J. Alwall, S. Höche, F. Krauss, N. Lavesson, L. Lönnblad, F. Maltoni, M.L. Mangano, M. Moretti, C.G. Papadopoulos, F. Piccinini, and et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *The European Physical Journal C*, 53(3):473–500, Dec 2007.

[154] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015.

[155] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, José I. Latorre, and et al. Parton distributions for the lhc run ii. *Journal of High Energy Physics*, 2015(4), Apr 2015.

[156] The NNPDF Collaboration, Richard D. Ball, Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Patrick Groth-Merrild, Alberto Guffanti, Nathan P. Hartland, Zahari Kassabov, José I. Latorre, Emanuele R. Nocera, Juan Rojo, Luca Rottoli, Emma Slade, and Maria Ubiali. Parton distributions from high-precision collider data, 2017.

[157] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[158] V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, and et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *The European Physical Journal C*, 76(3), Mar 2016.

[159] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, and et al. Extraction and validation of a new set of cms pythia8 tunes from underlying-event measurements. *The European Physical Journal C*, 80(1), Jan 2020.

[160] S. Agostinelli et al. GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[161] Jet pu id: efficiency and scale factors with z+jets. https://indico.cern.ch/event/840819/contributions/3527893/attachments/1892944/3122168/puid.pdf.

[162] Rochester correction. https://twiki.cern.ch/twiki/bin/viewauth/CMS/RochcorMuon.

[163] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup Per Particle Identification. *JHEP*, 10:059, 2014.

[164] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft Drop. *JHEP*, 05:146, 2014.

[165] Jesse Thaler and Ken Van Tilburg. Identifying Boosted Objects with N-subjettiness. *JHEP*, 03:015, 2011.

[166] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13(05):P05011, 2018.

[167] Jet identification in semi-leptonic wv vector boson scattering at the lhc. http://govoni.web.cern.ch/govoni/tesi/docs/Giacomo_Boldrini_Tesi.pdf.

[168] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. 3 2016.

[169] François Chollet et al. Keras. https://github.com/fchollet/keras, 2015.

[170] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[171] The International Linear Collider Technical Design Report - Volume 1: Executive Summary. 6 2013.

[172] Lucie Linssen, Akiya Miyamoto, Marcel Stanitzki, and Harry Weerts. Physics and detectors at clic: Clic conceptual design report, 2012.

[173] Michael Benedikt, Alain Blondel, Olivier Brunner, Mar Capeans Garrido, Francesco Cerutti, Johannes Gutleber, Patrick Janot, Jose Miguel Jimenez, Volker Mertens, Attilio Milanese, Katsunobu Oide, John Andrew Osborne, Thomas Otto, Yannis Papaphilippou, John Poole, Laurent Jean Tavian, and Frank Zimmermann. FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2. Future Circular Collider. Technical Report CERN-ACC-2018-0057. 2, CERN, Geneva, Dec 2018.

[174] Frank Zimmermann, Michael Benedikt, Mar Capeans Garrido, Francesco Cerutti, Brennan Goddard, Johannes Gutleber, Jose Miguel Jimenez, Michelangelo Mangano, Volker Mertens, John Andrew Osborne, Thomas Otto, John Poole, Werner Riegler, Laurent Jean Tavian, and Davide Tommasini. HE-LHC: The High-Energy Large Hadron Collider: Future Circular Collider Conceptual Design Report Volume 4. Future Circular Collider. Technical Report CERN-ACC-2018-0059, CERN, Geneva, Dec 2018.

[175] A. Abada et al. FCC-hh: The Hadron Collider: Future Circular Collider Conceptual Design Report Volume 3. *Eur. Phys. J. ST*, 228(4):755–1107, 2019.

[176] G Apollinari, I Béjar Alonso, O Brüning, M Lamont, and L Rossi. High-Luminosity Large Hadron Collider (HL-LHC) : Preliminary Design Report. 2015.

[177] Technical Proposal for the Phase-II Upgrade of the CMS Detector. CMS Technical Proposal CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02, 2015.

[178] K. Klein. The Phase-2 Upgrade of the CMS Tracker. 2017.

[179] CMS Collaboration. The Phase-2 Upgrade of the CMS Barrel Calorimeters Technical Design Report. CMS Technical Design Report CERN-LHCC-2017-011. CMS-TDR-015, 2017.

[180] CMS Collaboration. The Phase-2 Upgrade of the CMS Endcap Calorimeter. CMS Technical Design Report CERN-LHCC-2017-023. CMS-TDR-019, 2017.

[181] CMS Collaboration. The Phase-2 Upgrade of the CMS Muon Detectors. Technical Report CERN-LHCC-2017-012. CMS-TDR-016, CERN, Geneva, Sep 2017.

[182] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the cms detector. *JINST*, 12:P10003, 2017.

[183] CMS Collaboration. Expected performance of the physics objects with the upgraded CMS detector at the HL-LHC. Technical Report CMS-NOTE-2018-006. CERN-CMS-NOTE-2018-006, 2018.

[184] O. W. Greenberg and Joseph Sucher. A quantum structure dynamic model of quarks, leptons, weak vector bosons, and Higgs mesons. *Phys. Lett. B*, 99:339, 1981.

[185] M. Abolins, B. Blumenfeld, E. Eichten, H. Kagan, K. Lane, J. Leveille, D. Pellett, M. Peskin, and J. Wiss. Testing the Compositeness of Quarks and Leptons. In *Elementary Particles and Future Facilities (Snowmass 1982)*, page 274, 1982. eConf C8206282.

[186] H. Harari. Composite models for quarks and leptons. *Phys. Rept.*, 104:159, 1984.

[187] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[188] Jonathan Pumplin, Daniel Robert Stump, Joey Huston, Hung-Liang Lai, Pavel Nadolsky, and Wu-Ki Tung. New generation of parton distributions with uncertainties from global qcd analysis. *Journal of High Energy Physics*, 2002(07):012, 2002.

[189] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[190] J. de Favereau et al. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 1402:057, 2014.

[191] John Allison et al. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53:270, 2006.

[192] A.M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.

[193] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[194] M. Cacciari, G. P. Salam, and G. Soyez. Fastjet user manual. pages CERN–PH–TH (2011) 297.

[195] Search for heavy composite Majorana neutrinos inthe two leptons plus two jets final state with the Phase2 CMS detector at the HL-LHC. CMS Analysis Note CMS-AN-2018/112, 2018.

[196] Recommendations for systematic uncertainties hl-lhc. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/HLHELHCCommonSystematics`.

[197] Search for heavy composite Majorana neutrinos produced in association with a lepton and decaying into a same-flavour lepton plus two quarks at $\sqrt{s}$ = 13 TeV with the CMS detector. Technical Report CMS-PAS-EXO-16-026, CERN, Geneva, 2016.

[198] A L Read. Presentation of search results: the cl_s technique. *J. Phys. G: Nucl. Part. Phys.*, 28:2693, 2002.

[199] W. Buchmuller. Leptogenesis. *Scholarpedia*, 9(3):11471, 2014. revision #144189.

[200] M. Fukugita and T. Yanagida. Baryogenesis Without Grand Unification. *Phys. Lett.*, B174:45–47, 1986.

[201] A. Abada et al. He-lhc: The high-energy large hadron collider: Future circular collider conceptual design report volume 4. *Eur. Phys. J. ST*, 228(5):1109–1382, 2019.

[202] Michelangelo Mangano, Patrizia Azzi, Michael Benedikt, Alain Blondel, Daniel Andreas Britzger, Andrea Dainese, Mogens Dam, Jorge de Blas, David Enterria, Oliver Fischer, Christophe Grojean, Johannes Gutleber, Claire Gwenlan, Clement Helsens, Patrick Janot, Max Klein, Uta Klein, Matthew Philip Mccullough, Stephane Monteil, John Poole, Michael Ramsey-Musolf, Christian Schwanenberger, Michele Selvaggi, Frank Zimmermann, and Tevong You. FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1. Future Circular Collider. Technical Report CERN-ACC-2018-0056. 6, CERN, Geneva, Dec 2018.

[203] Andrea Thamm, Riccardo Torre, and Andrea Wulzer. Future tests of Higgs compositeness: direct vs indirect. *JHEP*, 07:100, 2015.

[204] B. C. Allanach, Ben Gripaios, and Tevong You. The case for future hadron colliders from $B \to K^{(*)}\mu^+\mu^-$ decays. *JHEP*, 03:021, 2018.

[205] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J.*, C77(10):663, 2017.

[206] CMS Collaboration. (AN-20-139) Search for the EW production of a VW pair plus two jets in the semi-leptonic lvjj channel with full Run-II data.