Search for $Z'$ bosons decaying into tau pairs in pp collisions at $\sqrt{s} = 13$ TeV with the CMS detector

Thesis presented by
Carlos Felipe González Hernández

for the degree of
DOCTOR EN CIENCIAS, FÍSICA.

Supervised by:
Juan Carlos Sanabria Arenas PhD.
Co-Supervised by:
Alfredo Gurrola PhD.

Bogotá, Colombia.
2018
A Susana.

“La belleza del rostro es frágil, es una flor pasajera, pero la belleza del alma es firme y segura.”,

Moliere

La belleza de tu alma siempre estará presente en nuestras vidas.

Y a Paula.

“Distance is nothing when love has wings”.
New heavy neutral gauge bosons, generically referred as $Z'$ bosons, are predicted in several theoretical scenarios beyond the standard model (BSM), such as Grand Unification Theories (GUT), Supersymmetry (SUSY) models and Superstring models. In these kinds of scenarios, the breaking of extended symmetry usually results in an extra $U(1)'$ symmetry group. This additional symmetry would give rise to the existence of a $Z'$ boson with a mass that should be in the TeV scale and therefore should be produced in pp collisions at the Large Hadron Collider (LHC). There are some scenarios that predict a generational coupling dependence, where the $Z'$ boson would decay preferentially to the third generation of fermions, for instance the topcolor-assisted technicolor (TAT) models. These scenarios are a motivation to search for $Z'$ bosons decaying into tau pairs, which was the main goal of this PhD dissertation. If a $Z'$-like resonance were found to decay also in the other fermion-pair final states ($ee$ or $\mu\mu$), the search for $Z' \to \tau\tau$ would be also very interesting, since it would reveal the nature of the couplings.

Since a $Z'$ boson might be produced as a result of the proton-proton collisions at the LHC, it might be observed by the CMS and ATLAS experiments as a massive resonance in the invariant mass distribution of its decay products, which, in case of the $Z' \to \tau\tau$ channel, are two oppositely-charged high $p_T$ taus. The search for $Z'$ decaying into tau-pairs involves four experimental signatures since the $\tau$ lepton can decay leptonically ($\tau_e$, $\tau_\mu$) or hadronically ($\tau_h$): $\tau_e\tau_\mu$, $\tau_e\tau_h$, $\tau_\mu\tau_h$ and $\tau_h\tau_h$. The $\tau_e\tau_\mu$ and $\tau_\mu\tau_h$ channels have a significant sensitivity due to the high reconstruction efficiency of light leptons in CMS and a relatively low QCD background contribution. However, the dihadronic tau channel has the best sensitivity since it has the highest expected signal yield, but it has a high background contribution coming from QCD multijet production.

During the Run II of the LHC, the evidence of a $Z'$ boson has been excluded and, the CMS and ATLAS experiments have constrained its existence in a wide range of mass; in the particular case of the $Z' \to \tau\tau$ search, ATLAS has excluded its existence for masses below 2.42 TeV using the data collected during 2015 and 2016, while CMS has excluded it for masses below 2.1 TeV, using the data collected during 2015. In this dissertation, the search for $Z'$ bosons in the dihadronic tau final state performed using data collected by CMS during 2016, is presented. This data corresponds to pp collisions at centre-of-mass energy of 13 TeV, with an integrated luminosity of 35.9 fb$^{-1}$. As a result of this analysis, expected exclusion limits have been established for the mass of the $Z'_{\text{SSM}}$ and $Z'_{\text{TAT}}$ bosons.
Acknowledgements

Just a quick view of the last years of my life is enough to realize that a lot of people made possible this PhD’s adventure. First of all, I want to thank my beloved wife. Pau, you have been always there, standing by me in every single moment of this crazy adventure. Your unconditional love and support made this possible. This achievement is also yours. Pau, we did it!. I would like also to thank my parents and my brothers, who have always encouraged my curiosity to know how nature works; I am very grateful for their invaluable teachings and their endless support to pursue my dreams. Undoubtedly, their lives have inspired me to be a better person every day.

A very special thanks to Professor Juan Carlos Sanabria since his teachings have laid the foundation for my knowledge in the high energy physics field. His clear understanding of physics and his particular way to explain it let me solve millions of questions. Thanks for his corrections to the document, for his patience, but, above all, I thank him for each conversation we had and the innumerable advice he gave me, not only as an advisor but also as a friend. Juan Carlos, thanks for encouraging my professional and personal growth.

I am very grateful to the BSM3G Group, in particular with the Professors Teruki, Alfredo, and Andres. Thanks for giving me the opportunity to work with such an outstanding team. It has been a real pleasure to work with this group. They, besides having a very clear insight of the field, are wonderful human beings. I appreciate a lot all their collaboration, which was fundamental for both, my learning process as well as the development of this research. I would like to give a very special recognition to Andres for all his advice, his patience and, mainly, for his friendship. I really appreciate all his help and his talks. I also want to thank the High Energy Group of Universidad de los Andes for giving me the chance and the proper environment needed to develop this research. Special thanks to Professor Carlos Avila, who was always available to help and support this project; I would like also to thank him for giving me the first chance to participate in this particle physics adventure.

I really appreciate a lot the support of all my friends. Jose and Adelaida, thanks for making the years at CERN unforgettable, for their constant and sincere friendship, thanks for all the drinks, travels and moments we share. Thanks to Jose for all your help to move this project forward, thanks for all the discussions and suggestions which were very important to speed up the process, thanks for your support which was always there. Thanks to “el Costeño”, Walter, “el diablito”, Sergio y Robert. We started this crazy adventure together, and throughout all kind of experiences in the academy, we lived great moments. Thanks, Walter, Costa and Sergio for the support in difficult times. Of course I would like to thank to the Tunja’s family: Pocho, Diana, Edwin, Julie, Rodrigo, Laura, Omar, Brigith, Jorge, Adriana, Arturo, Juancho, and Diego. Thanks for always believing in me and for their unconditional friendship. Thanks for the crazy moments which made the PhD adventure was easier.

Acknowledgments to the professors and administrative staff of the Physics Department for the collaboration. This work would not have been possible without the financial support of the Faculty of Sciences of the Universidad de los Andes and COLCIENCIAS.
Agradecimientos

Solo un rápido vistazo de los últimos años de mi vida es suficiente darme cuenta de que mucha gente hizo posible la aventura de este doctorado. Antes que nada, quiero agradecer a mi querida esposa. Pau, siempre has estado allí, apoyándome en cada momento de esta alocada aventura. Tu amor incondicional y tu apoyo hicieron esto posible. Este logro también es tuyo. Amore, lo hicimos!. También me gustaría agradecer a mis padres y a mis hermanos, quienes siempre han alentado mi curiosidad por saber cómo funciona la naturaleza. Estoy muy agradecido por sus invaluables enseñanzas y su apoyo infinito para que perseguiera mis sueños. Sin lugar a dudas, sus vidas me han inspirado a ser una mejor persona todos los días.

Un agradecimiento muy especial al Profesor Juan Carlos Sanabria ya que sus enseñanzas han sentado las bases de mi conocimiento en el campo de la física de altas energías. Su clara comprensión de la física y su forma particular de explicarla me permitieron resolver millones de preguntas. Gracias por sus correcciones al documento, por su paciencia, pero, sobre todo, le agradezco por cada conversación que tuvimos y por los innumerables consejos que me brindó, no solo como asesor sino también como amigo, que siempre alentaron mi crecimiento profesional y personal.

Estoy muy agradecido con el grupo BSM3G, en particular con los profesores Teruki, Alfredo y Andrés. Gracias por darme la oportunidad de trabajar con un equipo tan destacado. Ha sido un verdadero placer trabajar con este grupo quienes, además de tener una visión muy clara del campo, son seres humanos maravillosos. Aprecio mucho su colaboración, que fue fundamental tanto para mi proceso de aprendizaje como para el desarrollo de esta investigación. Me gustaría dar un reconocimiento muy especial a Andrés por todos sus consejos, su paciencia y, principalmente, por su amistad. Realmente aprecio toda su ayuda y sus charlas. También quiero agradecerle al Grupo de Altas Energías de la Universidad de los Andes por brindarme el entorno adecuado para desarrollar esta investigación. Un agradecimiento especial al Profesor Carlos Avila, quien siempre estuvo disponible para ayudar y apoyar este proyecto. También me gustaría agradecerle por darme la primera oportunidad de participar en esta aventura de la física de partículas.

Realmente aprecio mucho el apoyo de todos mis amigos. José y Adelaida, gracias por hacer que los años en el CERN sean inolvidables, por su constante y sincera amistad, gracias por todas las bebidas, viajes y momentos que compartimos. Gracias a José por toda su ayuda para hacer avanzar este proyecto, gracias por todas las charlas y sugerencias que fueron muy importantes para el desarrollo de proceso, gracias por su apoyo y sus consejos que siempre estuvieron presentes. Gracias a “el Costeño”, Walter, “el diablito”, Sergio y Robert, comenzamos esta loca aventura juntos, y durante todo tipo de experiencias en la academia, vivimos grandes momentos. Gracias, Walter, Costa y Sergio por el apoyo en los tiempos difíciles. Por supuesto, me gustaría agradecer a la familia de Tunja: Pocho, Diana, Edwin, Julie, Rodrigo, Laura, Omar, Brigit, Jorge, Adriana, Arturo, Juancho y Diego. Gracias por creer siempre en mí y por su amistad incondicional. Gracias por los momentos locos que hicieron que la aventura de doctorado fuera más fácil.

Finalmente quiero agradecer a los profesores, directivos y al personal administrativo del Departamento de Física por toda su colaboración. Este trabajo no hubiera sido posible sin el apoyo financiero de la Facultad de Ciencias de la Universidad de los Andes. También quiero agradecer el apoyo financiero de COLCIENCIAS.
## Contents

**Introduction**

1. **The Physics of The $Z'$ Boson**
   1.1 $Z'$ Couplings ........................................... 5
   1.2 A Heavy Gauge Boson $Z'$ in BSM Theories ................. 8
   1.3 Searches for $Z'$ Gauge Bosons
      1.3.1 Indirect Searches .................................... 9
      1.3.2 Direct Searches .................................... 10
   1.4 The Physics of the tau-lepton ............................ 15

2. **The CMS experiment**
   2.1 LHC Accelerator ........................................ 19
      2.1.1 LHC Proton Accelerator Chain ........................ 19
      2.1.2 LHC Operational Parameters ........................ 20
   2.2 The CMS Detector ....................................... 23
   2.3 Superconducting Solenoid ................................ 24
   2.4 The Tracker System ..................................... 26
   2.5 The Calorimeter System
      2.5.1 The Electromagnetic Calorimeter .................... 29
      2.5.2 The Hadronic Calorimeter ........................... 31
   2.6 The Muon System ......................................... 33
   2.7 The Trigger and Data Acquisition Systems
      2.7.1 The Trigger System .................................... 35
      2.7.2 Data Acquisition System ............................. 36
      2.7.3 Data Processing ..................................... 36

3. **Physics Object Reconstruction** .......................... 39
   3.1 Particle Flow ............................................. 39
      3.1.1 Track and Vertex Reconstruction ...................... 40
      3.1.2 Clustering ........................................... 41
      3.1.3 Link algorithm ...................................... 44
      3.1.4 PF Candidates ....................................... 44
   3.2 Muon Reconstruction and Identification .................. 45
   3.3 Electron Reconstruction and Identification ............. 46
   3.4 Jet Reconstruction ....................................... 48
   3.5 b-jet Identification ..................................... 52
   3.6 Missing Transverse Energy ................................ 53
   3.7 Tau Reconstruction and Identification .................... 55
      3.7.1 Tau Reconstruction Algorithm ....................... 56
      3.7.2 Tau Identification ................................... 60

4. **Experimental Signature of $Z' \rightarrow \tau_h \tau_h$** ....... 67
   4.1 Signature .................................................. 67
4.2 Expected Background .................................................. 68
4.3 Mass Reconstruction ..................................................... 72
4.4 Strategy ................................................................. 72

5 Search For Z′ Bosons With 2016 Data .............................. 75
5.1 Trigger Selection .......................................................... 75
5.1.1 Trigger Efficiency ..................................................... 76
5.2 Data and Simulated samples ........................................... 78
5.2.1 Data Samples .......................................................... 78
5.2.2 Simulated Samples ..................................................... 78
5.3 Event Selection .......................................................... 80
5.3.1 τh Identification ......................................................... 81
5.3.2 Topological Selections ............................................... 83
5.3.3 Summary ............................................................... 86
5.3.4 Validation Plots for Signal Selections .......................... 87
5.4 Corrections for Simulated Events ................................. 90
5.5 Background Estimation ................................................ 92
5.5.1 Background Estimation for QCD ................................. 92
5.5.2 Background Estimation for Drell-Yan events .................. 97
5.5.3 Validation ............................................................... 99
5.6 Systematic Uncertainties ............................................... 100
5.7 Summary ............................................................... 104

6 Analysis and Conclusions .............................................. 105
6.1 Analysis ................................................................. 105
6.1.1 Statistical Method ..................................................... 105
6.1.2 Exclusion Limit Calculation ....................................... 107
6.2 Analysis of Results ..................................................... 109
6.2.1 Comparison with Other Channels ............................... 109
6.2.2 Comparison with previous CMS Searches ................. 110
6.2.3 Comparison with Results from ATLAS ....................... 112
6.3 Conclusions ........................................................... 113

A Trigger Studies ......................................................... 115

B Tau Identification Studies ............................................. 117
B.1 Number of prongs Study ............................................... 117
B.2 MVA-based Isolation Discriminator Study ..................... 119
B.3 MVA-based against Electron Discriminator Study .......... 119
B.4 Cutoff-based against Muon Discriminator Study ............. 119
B.5 Decay Mode Finding (DMF) Discriminator Study .......... 120

C Event Selection .......................................................... 123

Bibliography ............................................................... 125
**Introduction**

Our understanding of the behavior of elementary particles and their interactions is currently described by the Standard Model (SM). Nevertheless, there are still open questions in particle physics that are not addressed by this model. For instance: the SM does not provide an explanation for the matter-antimatter asymmetry; it does not explain the origin of the neutrino’s masses, as well as the neutrino oscillations; besides, the SM does not explain the nature of dark matter and dark energy; it does not include a quantum version of gravity; among others. For these reasons, many theories, known as Beyond Standard Model (BSM) theories [1], have been proposed. Most of them extend the SM symmetry group in order to address the open questions.

One of the simplest ways of extending the SM symmetry is through an additional $U(1)'$ group. This symmetry would give rise to the existence of a new neutral gauge boson, generically referred as $Z'$. These kind of models were initially motivated by the SM electroweak symmetry breaking, since it leaves the $U(1)_{EM}$ group as a remnant. In a similar way, breaking an extended SM symmetry could resort in an extra $U(1)'$ symmetry group. The additional $U(1)'$ group and its associated $Z'$ appear in several scenarios of Grand Unification Theories (GUT), Super Symmetry (SUSY) and Super String models. In the GUT case, at least one $Z'$ boson is predicted for scenarios where the gauge group is greater than SU(5), for instance SO(10) or $E_6$. There are some SUSY and Super String versions of GUT where the SM electroweak and the $U(1)'$ symmetry groups generally break at the soft SUSY breaking scale [2]; in these scenarios the $Z'$ boson would have a mass in the TeV range and, consequently, its experimental observation would stand for a clear evidence of Physics Beyond SM, which could come from scenarios of Super Symmetry or Super String models.

The discovery of a massive $Z'$ boson would have profound implications due to the nature of the $U(1)'$ symmetry, since it would require an extended Higgs sector and, in the case of SUSY scenarios, it would also require an extended neutralino sector [2]. This would have consequences for particle physics and for cosmology related with the dark matter problem. For example some SUSY models that predict a $Z'$, also predict extra Higgs, a right-handed neutrino and their superpartners; in these scenarios the right-handed sneutrino is a candidate for dark matter due to its interactions with the $Z'$ boson [3]. Besides, the $Z'$ discovery would also have implications for the nature of the neutrino mass, for the nature of the SUSY hidden sector and for the possible mediator(s) of SUSY breaking. On the other hand, there are many BSM models which attempt to solve the problem of quadratic divergencies for the Higgs mass through an extra $U(1)'$ symmetry, for example models with Little Higgs [2]. The Little Higgs models predict new Higgs in the TeV energy scale as well as one or more $Z'$ bosons.

The $Z'$ searches are motivated by those BSM scenarios which predict $Z'$ bosons with masses at TeV scale and therefore, if they exist, they would be produced at the Large Hadron Collider (LHC) at CERN [2]. In the $Z'$ searches the theoretical model generally used as a benchmark is the Sequential Standard Model (SSM). This is the simplest model since, the $Z'_{SSM}$ boson obeys the universality of the SM couplings: it has the identical couplings than the SM $Z^0$ gauge boson.
Nevertheless, there are also many BSM scenarios that predict a generational coupling dependence, where the $Z'$ boson would decay preferentially to the third generation of fermions \cite{4}. For instance, the topcolor-assisted technicolor (TAT) models, which attempt to give an explanation to the high mass of the top quark, predict a heavy gauge boson $Z'_{\text{TAT}}$ with enhanced couplings to the third generation \cite{5}. Consequently, since in these scenarios the coupling of the $Z'$ boson to the tau-lepton is much higher than the couplings to the other leptons if such $Z'$ boson would exist it would be observed in the $Z' \rightarrow \tau\tau$ channel. These models are an especial compelling motivation to search for $Z'$ bosons decaying into tau pairs, which is the main goal of this PhD dissertation. Additionally, in the case of models which predict the universality of couplings, or in which the difference between the coupling strength of the $Z'$ to the third and the second/first lepton generations is small, the new gauge boson would be observed first in the other fermion-pair final states (such as $\mu\mu$) and, therefore, the search for $Z' \rightarrow \tau\tau$ would be very important since it would confirm or it would discard the hypothesis of the universality of the couplings.

Because of the $Z'$ bosons are predicted at TeV scales and they would couple to the SM fermions, these hypothetical particles would be produced in the proton-proton collisions at the LHC. The $Z'$ signature would be reflected as a heavy resonance in the invariant mass distribution of its decay products: opposite-charged fermions (high momentum leptons or jets coming from $q\bar{q}$). The $Z'$ search is model-independent since it is based on the observation of an excess in the high spectrum of the effective visible mass distribution and, therefore, the search for a massive resonance can be sensitive to final states foreseen by other theories and the results can be reinterpreted according to a particular model. During the last two decades, the evidence of a $Z'$ boson has been excluded by several searches. The CDF \cite{6–9} and D0 \cite{10–12} experiments at the Tevatron, and the CMS \cite{13–24} and ATLAS \cite{25–34} experiments at the LHC, have constrained the existence of $Z'$ bosons in a wide range of mass. The tightest upper limits on the $Z'$ mass have been set by the CMS and ATLAS experiments in the dilepton final state ($Z' \rightarrow \ell\ell$, where $\ell = e, \mu$), where the $Z'_{\text{SSM}}$ has been excluded below 4.5 TeV in both searches. These searches were performed with data collected by CMS during 2016 \cite{13}, and by ATLAS during 2015 and 2016 \cite{25}. The channels where the $Z'$ decays into a light lepton pair ($ee$ and $\mu\mu$ final states) have been explored widely due to their high mass resolution, large acceptance and relative low background. The most significative background in the dilepton channels comes from Drell-Yan (DY) processes, which have been studied extensively. This background has the same topology as the hypothetical $Z'$ events, but it can be largely suppressed since its reconstructed mass lays mainly around the $Z$ boson mass. Nevertheless, the reconstructed mass distribution of DY+jets events has a long tail extending to large values, which makes them an important and irreducible background for $Z'$ searches.

The experimental signature of the $Z' \rightarrow \tau\tau$ channel consists on events with high $p_T$, almost back-to-back and oppositely charged $\tau$-pairs. The search for $Z'$ decaying into tau-pairs usually involves four experimental signatures, since the $\tau$ lepton has different decay modes: it can decay leptonically into a lighter lepton and two neutrinos (we will refer to the taus that decay in this way as $\tau_e$ or $\tau_\mu$, depending on the lepton in which they decay); or it can decay hadronically into a set of pions and a neutrino (we will refer to the taus that decay in this form as $\tau_\nu$). Therefore, the search for $Z' \rightarrow \tau\tau$ includes the channels: $\tau_e\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_\mu$ and $\tau_h\tau_h$. The $\tau_e\tau_e$ and $\tau_\mu\tau_\mu$ channels are not included in the search due to their relative low branching ratio, and because of the difficulty to distinguish their signature from the dilepton channels ($Z' \rightarrow \mu\mu$, $Z' \rightarrow ee$). The $\tau$ identification represents an important experimental challenge for several reasons: first, since the $\tau$ can not be reconstructed fully due to neutrinos, the momentum can not be known; second, the leptonic decay of the tau can not be identified since it has the same signature than events
in which leptons were produced as a result of a collision; and finally, the hadronic decay has a similar signature to a QCD-jet. Therefore, sophisticated tau identification algorithms have been developed by CMS and ATLAS experiments in order improve the identification efficiency. The CMS experiment uses the Hadron Plus Strips (HPS) algorithm, that identifies the hadronic tau leptons with an efficiency of $\sim 60\%$. Therefore, the difficulty of the tau identification, makes the search for $Z' \rightarrow \tau\tau$ challenging.

Two of the most sensitive channels are the $Z' \rightarrow \tau\ell\tau_h$ since they have a relative high yield, a better acceptance due to the presence of light leptons, and a relatively low QCD multijet background. The $Z' \rightarrow \tau_h\tau_h$ channel has the highest expected yield but it also has a very high background coming from QCD multijet production. The purpose of this dissertation is to search for $Z'$ bosons decaying into two hadronic taus in proton-proton collisions at a centre-of-mass energy of 13 TeV, with the data collected by the CMS experiment during 2016. This analysis makes part of the search for heavy resonances in the ditau final state, in which article I am one of the main authors/analysts.

During the Run I at the LHC, the CMS and ATLAS experiments performed searches for heavy resonances in the ditau final states, using the combined channels mentioned above. CMS has excluded the $Z'_{SSM}$ for masses below 1.4 TeV, using pp collisions at $\sqrt{s} = 7$ TeV; while ATLAS has set an exclusion limit for the $Z'_{SSM}$ mass of 2.02 TeV, using pp collisions at $\sqrt{s} = 8$ TeV. CMS Collaboration has also carried out a search for a $Z' \rightarrow \tau_\ell\tau_\mu$ boson, using collisions at $\sqrt{s} = 8$ TeV, excluding the $Z_{SSM}$ for masses below 1.3 TeV. With the high luminosity and the new energy range reached by the LHC during Run II, CMS has performed also the $Z' \rightarrow \tau\tau$ search at $\sqrt{s} = 13$ TeV using the 2015 data, and reporting a new upper limit of 2.1 TeV for the $Z'_{SSM}$ mass. The higher luminosity reached by the LHC during 2016 made possible an improved search. This document presents this search, using the data collected by the CMS experiment during 2016 at a centre-of-mass energy of 13 TeV. This analysis is currently unblinded, i.e., the CMS Collaboration has approved to explore with the real data the region where the $Z' \rightarrow \tau\tau$ is expected. However, the results using real data are still being reviewed internally and, therefore, they are not included in this document. Instead, expected exclusion limits have been calculated. We expect to exclude $Z'_{SSM}$ and $Z'_{TAT}$ bosons (decaying into two hadronic taus), in the case in which no signal is observed, for masses below 2.7 TeV and 2.2 TeV. These results show that the sensitivity reached in this analysis has improved over the one obtained in the previous search.

This document is organized as follows: Chapter 1 presents the theoretical bases of the physics of the $Z'$ boson, including the most relevant models as well as the results of previous searches. In Chapter 2, the CMS experiment is described, starting with a brief description of the LHC and, afterwards, presenting a full description of the CMS detector and its sub-detectors, making emphasis in the tau detection. Chapter 3 presents a detailed description of the physics objects reconstruction algorithms used by the CMS Collaboration, focusing mainly in the tau algorithm. The identification of the experimental signature of the $Z' \rightarrow \tau_h\tau_h$ events and its background is discussed in Chapter 4. All the experimental techniques involved in the identification of the signal and the reduction of the backgrounds are presented in Chapter 5. Finally, the results of the search and conclusions are presented in the Chapter 6.
1 The Physics of the Z’ Boson

In the search for physics BSM at the LHC one of the most exciting possibilities is the observation of neutral vector bosons with masses in the range of the TeV’s, generally known as Z’. Many models of physics BSM include this kind of states as a result of the breaking of higher symmetries upon which the models are built.

The Standard Model is a gauge field theory that has as its underlying symmetry group $G_{SM} = SU(3)_C \otimes SU(2)_L \otimes SU(1)_Y$ and a breaking mechanism mediated by the Higgs field, that leaves, as the remaining symmetry the group, $SU(3)_C \otimes U(1)_{EM}$. This is an example of how the breaking of a larger symmetry results on one or more $U(1)$ factors. These $U(1)$ factors are associated to gauge bosons that can be massive.

Models beyond the SM have larger symmetry groups than $G_{SM}$, and various symmetry breaking mechanisms that include $G_{SM}$ as remaining symmetry. But, in many of these scenarios extra $U(1)$ factors can also remain. They would be associated to neutral and massive vector bosons. For instance, a group $SU(N)$ can be broken by a Higgs-like mechanism, and the unbroken subgroup could contain up to N-1 $U(1)$ factors [2]. Therefore, the emergence of remaining $U(1)$ symmetries in the breaking of larger groups is a common feature of models BSM. This is why experimental signatures of new physics would, most likely, involve neutral vector bosons.

In the case of Grand Unified Theories (GUT), groups like $SU(5)$, $SO(10)$ or $E_6$ have been used as the main symmetry whose breaking would, eventually, take us to $G_{SM}$. But, this does not discard the presence of extra $U(1)$ factors, whose associated gauge bosons have large enough masses, so that they have not been observed so far, but that could be within the reach of the LHC experiments.

Not just plain GUT theories, but also superstring theories have large symmetry groups that break into $G_{SM} \otimes U(1)^n$ with $n \geq 1$. A similar situation can be found in GUT-SUSY models. Even some versions of theories with extra space dimensions involve massive Z’ in their particle spectra.

The breaking of the larger symmetry, resulting in $U(1)$ factors, would typically demand for an extended Higgs sector, that would have implications in particle physics and cosmology [2]. This is why the observation of a Z’ would provide very valuable information about the new physics involved.

The theoretical implications of the discovery of a Z’ at the LHC could range from flavor changing neutral current processes, rare B decays, production of other massive exotic particles in the decay chain, neutrino masses, etc [2].

1.1 Z’ Couplings

To understand how a new Z’ field would couple to other fields, specially to those of the SM, one can look at the case of the two known neutral vector bosons: $A_\mu$ and $Z_1\mu$ (here a sub-index 1
has been included to distinguish it from new Z-like states). The interaction of these fields with fermions is given by \[2\]:

\[
g J^\mu_3 W_3^\mu + g_t J^\mu_t B^\mu = e J^\mu_{em} A^\mu + g_1 J^\mu_1 Z_1^\mu ,
\]

where \(g\) and \(g_t\) are the gauge couplings associated to SU(2) and U(1)\(_Y\) respectively.

Here we have the interaction written in two basis:

- the gauge-eigenstates basis: \(W_3^\mu, B^\mu\),
- the mass-eigenstates basis: \(A^\mu, Z_1^\mu\),

related by:

\[
A^\mu = \sin \theta_W W_3^\mu + \cos \theta_W B^\mu ,
\]

\[
Z_1^\mu = \cos \theta_W W_3^\mu - \sin \theta_W B^\mu ,
\]

where,

\[
\tan \theta_W = \frac{g_t}{g} ; \quad e = g \sin \theta_W ,
\]

\[
g_1^2 = g^2 + g_t^2 = \frac{g^2}{\cos^2 \theta_W} .
\]

The fermion currents are given by:

\[
J^\mu_3 = \sum_i \bar{f}_i \gamma^\mu [t_{3L} P_L + t_{3R} P_R] f_i ,
\]

\[
J^\mu_t = \sum_i \bar{f}_i \gamma^\mu [y_{iL} P_L + y_{iR} P_R] f_i ,
\]

in the gauge-eigenstates basis, and by:

\[
J^\mu_{em} = \sum_i q_i \bar{f}_i \gamma^\mu f_i ,
\]

\[
J^\mu_1 = \sum_i \bar{f}_i \gamma^\mu [\epsilon^1_L (i) P_L + \epsilon^1_R (i) P_R] f_i ,
\]

in the mass-eigenstates basis.

Here \(f_i\) are the fermion fields, \(P_{L,R} = (1 \pm \gamma^5)/2\), and \(t_{3iL,R}\) is the third component of weak isospin of the fermion field, given by:

\[
t_{3uL} = t_{3uL} = +\frac{1}{2} ,
\]

\[
t_{3dL} = t_{3cL} = -\frac{1}{2} ,
\]

\[
t_{3R} = 0 .
\]
The $y_{L,R}$ are the weak hypercharges given by:

$$t_{3iL} + y_{iL} = t_{3iR} + y_{iR} = q_i, \quad (1.7)$$

where $q_i$ is the electric charge of the fermion field.

The chiral couplings $\epsilon^L_1(i)$ and $\epsilon^R_1(i)$ are given by:

$$\epsilon^L_1(i) = t_{3iL} - \sin^2 \theta_W q_i, \quad \epsilon^R_1(i) = t_{3iR} - \sin^2 \theta_W q_i. \quad (1.8)$$

In the case of extra $U(1)'$ symmetries, the interaction terms become:

$$eJ^\mu \epsilon_{em} A^\mu + \sum_{\alpha=1}^{n+1} g_{\alpha} J^\mu_{\alpha} Z^\mu_{\alpha}, \quad (1.9)$$

where $n$ is the number of extra $U(1)'$ factors; $g_1, Z_{1\mu}$, and $J_{1\mu}$ are the gauge coupling, the boson field, and the SM neutral current, respectively. Similarly, $g_{\alpha}, Z_{\alpha\mu}, \alpha = 2 \cdots n + 1$, are the gauge couplings and boson fields for the additional $U(1)'$ factors. The currents in equation (1.9) are:

$$J^\mu_{\alpha} = \sum_i \bar{f}_i \gamma^\mu [\epsilon^\alpha_L P_L + \epsilon^\alpha_R P_R] f_i, \quad (1.10)$$

where, the chiral couplings $\epsilon^\alpha_{L,R}(i)$ are respectively the $U(1)_{\alpha}$ charges of the left and right handed components of fermion $f_i$, and $g_{\alpha} V, A(i) = \epsilon^\alpha_L(i) \pm \epsilon^\alpha_R(i)$ are the corresponding vector and axial couplings.

Following a spontaneous breaking mechanism similar to the one of the SM, complex $SU(2)$ scalar multiplets $\phi_i$ are included:

$$\phi_i = \left( \begin{array}{c} \phi_i^+ \\ \phi_i^0 \end{array} \right). \quad (1.11)$$

If the neutral component of $\phi_i$ acquires a vacuum expectation values (VEV), different from zero, $A_\mu$ would remain massless, while the $Z_{\alpha\mu}$ develop mass terms of the form:

$$\frac{1}{2} M^2_{\alpha\beta} Z_{\alpha\mu} Z^\mu_{\beta},$$

where $M_{\alpha\beta}$ is the mass matrix for the $Z$ vector bosons.

In the case $n = 1$ (only one $Z'$), the matrix takes the form:

$$M_{\alpha\beta} = \left( \begin{array}{cc} M^2_{Z0} & \Delta^2 \\ \Delta^2 & M^2_{Z'} \end{array} \right), \quad (1.12)$$

This matrix can be diagonalized with the rotation:

$$U = \left( \begin{array}{cc} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{array} \right). \quad (1.13)$$
\[ \theta = \frac{1}{2} \arctan \left( \frac{2\Delta^2}{M_{Z_0}^2 - M_{Z_1}^2} \right) . \]  

The mass eigenvalues are given by:
\[ M_{1,2}^2 = \frac{1}{2} \left[ M_{Z_0}^2 + M_{Z_1}^2 \mp \sqrt{(M_{Z_1}^2 - M_{Z_0}^2)^2 + 4\Delta^4} \right] , \]  

where \( \Delta^2 \) is a quantity that depends on the model, and on the VEV of the Higgs doublets. If \( M_{Z_1} >> M_{Z_0}, |\Delta| \) then
\[ M_1^2 \sim M_{Z_0}^2 - \frac{\Delta^4}{M_{Z_1}^2} \sim M_{Z_0}^2 << M_{Z_1}^2 \]  

In summary, due to the needed symmetry breaking, there is a mass mixing between \( Z_0 \) and \( Z_1 \). But if \( M_{Z_1} >> M_{Z_0} \), the mixing is small [36].

Another source of mixing comes from the kinetic terms. The most general form of these terms is:
\[ -\frac{1}{4} F_{\alpha \mu}^{\alpha \nu} F_{\alpha \nu} - \frac{1}{4} F_{\beta \mu}^{\alpha \nu} F_{\beta \nu} - \frac{c_{\alpha \beta}}{2} F_{\alpha \mu}^{\alpha \nu} F_{\beta \nu} \]  

where \( F_{\alpha \mu} = \partial_{\mu} Z_{\alpha \nu} - \partial_{\nu} Z_{\alpha \mu} \). As one can see, there is a cross term that generates a kinetic mixing between the fields. In the case of \( n = 1 \), one can write \( c_{\alpha \beta} = \sin \chi \), and kinetic mixing generates a transformation:
\[ \begin{pmatrix} Z_{1\mu} \\ Z_{2\mu} \end{pmatrix} = \begin{pmatrix} 1 & -\tan \chi \\ 0 & \frac{1}{\cos \chi} \end{pmatrix} \begin{pmatrix} \hat{Z}_{1\mu} \\ \hat{Z}_{2\mu} \end{pmatrix} , \]  

that, in turn, will affect the mass eigenvalues. Again, in most of the cases, \( \chi \) is small, and the effect over the mass eigenvalues is also small.

### 1.2 A Heavy Gauge Boson \( Z' \) in BSM Theories

As mentioned, there are many theoretical models that include \( U(1)' \) symmetries. These models differ in coupling constants, even though most of them assume the electroweak strength values of the SM. The other aspect that differentiates between models is the symmetry breaking scheme and scale. Some examples of these kinds of theories are:

- **The Sequential Standard model (SSM):** is a simplified model in which the \( Z' \) has the same couplings and quantum numbers as the \( Z^0 \), but higher mass. It serves as a reference for experimental searches [37].

- **\( E_6 \) models:** These models are based on the \( E_6 \) group as gauge symmetry, which breaks at GUT scale down into the SM gauge group and extra \( U(1) \) symmetry groups, usually two of them. Depending on the features of the particular model, one can have \( Z' \) bosons with masses at TeV scale, which might be produced at the LHC [38]. Two scenarios have been used as a benchmark for the experimental search: the first one predicts the existence of the \( Z'_{\psi} \) and the \( Z'_{\chi} \) bosons, which arise from models when the \( E_6 \) group breaks into \( SO(10) \times U(1)_{\psi} \) and, subsequently, the \( SO(10) \) breaks into \( SO(5) \times U(1)_{\chi} \); in the second one, a \( Z'_{\phi} \) boson is predicted from a breaking symmetry of the \( SO(10) \) group in GUT scales [2].
• **Left-Right symmetric models**: These models involve the gauge group $SU(2)_L \times SU(2)_R \times U(1)_{B-L}$, where $B$ and $L$ refer to Baryon and Lepton numbers. These kind of theories predict a new heavy $Z'$, but also a new heavy $W^\pm'$ bosons \cite{39}. The new $W'$ is always lighter than the $Z'$, providing a good experimental test.

• **Technicolor models**: This kind of models include a new gauge force with properties similar to those of QCD. They predict new particles, such as technigluons and techniquarks. In extended versions, new gauge bosons couple to the SM fermions and the technifermions. A $Z'$ boson emerges from the symmetry breaking induced by technifermion condensates. This boson couples only to left handed fermions, and has enhanced couplings to fermions of the third generation. Therefore, the search for $Z' \rightarrow \tau\tau$ is of particular importance for testing these models \cite{40}.

• **Topcolor assisted technicolor models**: TAT models interpret the large value of the top quark mass as an evidence of a dynamical electroweak breaking mechanism that depends on the fermion generation. They assume an extended gauge sector of the form $SU(3)_1 \times SU(3)_2 \times U(1)_1 \times U(2)_2$. Fermions of the first and second generation transform under $U(1)_1$, while fermions of the third generation transform under $U(2)_2$. The $Z'$ arises from the $U(1)_1$ group. In these models the couplings to fermions of the third generation are also enhanced \cite{4,5}.

• **Little Higgs models**: These are the most popular non-GUT models that include new heavy gauge bosons. New $W'$ and $Z'$ bosons, with masses of the order of TeV are predicted as a result of the symmetry breaking of a $[SU(2)_L \times U(1)]^2$ gauge group that is part of the models \cite{41,42}.

### 1.3 Searches for $Z'$ Gauge Bosons

The $Z'$ bosons predicted by the many theoretical models mentioned above would have couplings to SM quarks, therefore, they might be produced by hadron colliders and might be observed directly as additional resonances in invariant mass distribution plots. Tevatron and LHC experiments have performed direct searches for $Z'$ resonances. LEP performed indirect searches, since the $Z'$ could not be produced on-shell, given the energy range reached by this accelerator (209 GeV). The searches performed by the LEP, Tevatron, and LHC experiments have not shown any evidence of a $Z'$ boson, and have constrained its existence on a wide range of mass. In this section the state of the art in the indirect and direct experimental searches for the $Z'$ boson are presented.

#### 1.3.1 Indirect Searches

LEP experiments performed indirect searches using accurate electroweak measurements around the $Z$ peak, looking for interference effects that might reveal a possible mixed state between $Z'$ and $Z$ bosons, predicted by some BSM scenarios (see Section 1.1). LEP experiments looked for evidences of the mixing, using measurements of cross sections and forward-backward asymmetries in the dilepton channels. The LEP experiments did not report any deviation from the SM expectations, and indirect limits on the $Z'$ mass were established (see Table 1.1). The results were obtained with the data sample collected by the four LEP experiments, using $e^+e^-$ collisions data centre-of-mass energy of 209 GeV \cite{43,44}.
1.3.2 Direct Searches

Direct searches for a $Z'$ boson have been performed by the CDF [6–9] and D0 [10–12] experiments at the Tevatron, and by the CMS [14–24] and ATLAS [25–34] experiments at the LHC. The $Z'$ bosons have been searched in many channels, since there are several theoretical models and several couplings in each scenario. The Sequential Standard Model, or SSM, is one of the most frequently used models since it predicts a $Z'_{SSM}$ whose couplings are identical to those of the $Z$ boson (see Section 1.2). Furthermore, other searches have included hypothetical $Z'$ bosons coming from GUT theories, such as the $Z'_\psi$ and $Z'_\chi$ bosons which arise from a symmetry breaking of the $E_6$ group, and the $Z'_\eta$ boson which comes from a breaking symmetry of the SO(10) group in GUT scales (see Section 1.2). Additionally, models where the $Z'$ boson decays preferentially to the third generation of leptons (see Section 1.2), as the topcolor-assisted technicolor (TAT) models, have motivated the searches in the ditau channel. In these scenarios, the $Z'$ boson has a higher coupling to the third generation than to the other lepton generations and, therefore, if a $Z'$ would exist, it would be observed first in the ditau channel.

A direct evidence of a $Z'$ boson should show up as a heavy resonance in the mass spectrum of its decay products, since it would decay into two leptons or jets with high momentum and opposite charge. No direct evidence has been found yet and its existence has been excluded on a wide range of mass. The constraints on the $Z'$ mass are usually presented as limits in the production cross section times the branching fraction of its decay products, for instance: $\sigma(pp \rightarrow Z') \times B(Z' \rightarrow \ell\ell)$. Several direct searches have been performed according to its decay products:

- **Dilepton searches ($Z' \rightarrow \ell\ell$ channel, $\ell = e, \mu$).**

  The $Z' \rightarrow \ell\ell$ channel has been widely explored since light leptons are usually reconstructed with high efficiencies in a large range of acceptance, which leads to a high resolution in the dilepton invariant mass; additionally, the dilepton channel has a negligible QCD multijet background contamination. These features make this channel the most sensitive one, providing the tightest exclusion limits on the $Z'$ mass. The main background comes from Drell-Yan (DY) processes since they have the same topology than the hypothetical $Z'$ events, and their reconstructed mass distribution can extend to large values where the $Z'$ boson is expected; therefore, DY processes are an irreducible source of background. However, since this background have been very well studied and understood, its contribution is well modeled and simulated.

  The CDF and D0 experiments performed the searches in the dilepton channel [6,7,11], using data samples of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. In the dielectron channel, CDF used a sample with an integrated luminosity of 2.5 fb$^{-1}$ and did not report a significant excess for the masses below 923 GeV in the case of the $Z'_{SSM}$ and for masses below 822 GeV in the case of the $Z'_\psi$ (see Figure 1.1, left). The data collected by D0, with an integrated

<table>
<thead>
<tr>
<th>$Z'$ model</th>
<th>$\chi$</th>
<th>$\psi$</th>
<th>$\eta$</th>
<th>SSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass limit for $Z'$ (GeV)</td>
<td>673</td>
<td>481</td>
<td>434</td>
<td>1787</td>
</tr>
</tbody>
</table>

Table 1.1: The 95% confidence level lower limits on the $Z'$ mass for $\chi$, $\psi$, $\eta$ and SSM models set by the experiments ALEPH, DELPHI, L3 and OPAL [44].
luminosity of 5.4 fb$^{-1}$, excluded the existence of the $Z'_{SSM}$ for masses below 1023 GeV and the $Z'_{ψ}$ for masses below 891 GeV (see Figure 1.1, right). The CDF Collaboration also searched in the $Z' → μμ$ channel with an integrated luminosity of 4.6 fb$^{-1}$, which resulted in a lower limit of 1071 GeV for the $Z'_{SSM}$ mass [7].

The CMS and ATLAS experiments have performed searches for high-mass resonances in the dilepton channel using each data taking period [13–16, 25–28]. The most recent search performed by the CMS experiment was performed using the data collected during 2016 [13]. For the dielectron channel, the possible existence of a $Z'$ resonance is excluded up to 4.1 TeV for the $Z'_{SSM}$ and up to 3.5 TeV for the $Z'_{ψ}$. For the dimuon channel, the CMS Collaboration reports an absence of resonances in the mass spectrum below 4.3 TeV for $Z'_{SSM}$ and 3.7 TeV for $Z'_{ψ}$. The combined analysis provides an exclusion limit of 4.5 TeV for $Z'_{SSM}$ and 3.9 TeV for $Z'_{ψ}$, respectively (see Figure 1.2 left). On the other hand, the most recent search for $Z'$ bosons decaying into the dilepton channel performed by the ATLAS experiment uses the data collected during 2015 and 2016 [25]. The ATLAS Collaboration has excluded, with a 95% Confidence Level (CL), the mass of $Z'_{SSM}$ below 4.3 TeV in the dielectron channel and 4.0 TeV in the dimuon channel, while the mass of $Z'_{ψ}$ has been excluded for masses below 3.9 TeV and 3.6 TeV in the dielectron channel and in the dimuon channel, respectively. The analysis with the combined channels results in an exclusion limit of 4.5 TeV for the $Z'_{SSM}$ boson and 3.8 TeV for $Z'_{ψ}$ boson (see Figure 1.2 right). The summary of the exclusion limits set on the $Z'_{SSM}$ and $Z'_{ψ}$ for the dilepton searches performed by the CMS and ATLAS experiments are shown in Table 1.2.

Figure 1.1: The figures show the 95% C.L. upper limits on $σ(pp → Z') \times B(Z' → ee)$ as a function of $M_{Z'}$, compared to the theoretical predictions of the cross section for the $Z'_{SSM}$ and the bosons arising from $E_6$. (Left) Upper limits observed by CDF [6]. (Right) Upper limits observed by D0 [11].
Figure 1.2: (Left) Upper 95\% CL limits as a function of the resonance mass $M$ on the ratio of the product of cross section and branching fraction into lepton pairs relative to that of $Z$ bosons, for the 13 TeV data collected by CMS during 2016. Theoretical predictions for $Z'_{\text{SSM}}$ and $Z'_{\psi}$ are shown for comparison [13]. (Right) Observed upper cross-section times branching ratio limits at 95\% CL for $Z'$, E6-motivated $Z'_{\psi}$ and $Z'_{\chi}$ bosons using the combined dilepton channel, for the combined 2015 and 2016 data collected by ATLAS. In addition, theoretical cross-sections are shown for the same models [25].

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\sqrt{s}$ [TeV]</th>
<th>$\mathcal{L}$ [fb$^{-1}$]</th>
<th>Upper Limit for $Z'_{\text{SSM}}$ [TeV]</th>
<th>Upper Limit for $Z'_{\psi}$ [TeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMS</td>
<td>13</td>
<td>35.9 for $ee$ and 36.3 for $\mu\mu$</td>
<td>4.5</td>
<td>3.9</td>
</tr>
<tr>
<td>ATLAS</td>
<td>13</td>
<td>36</td>
<td>4.5</td>
<td>3.8</td>
</tr>
</tbody>
</table>

Table 1.2: Summary of the upper 95\% CL limits set for $Z'_{\text{SSM}}$ and $Z'_{\psi}$ masses in the dilepton channel by the CMS and ATLAS experiments.

- **Ditau searches ($Z' \rightarrow \tau\tau$ channel).**

The CMS and ATLAS Collaborations have provided the most stringent limits on the $Z'$ mass using the dilepton channel. However, some models predict $Z'$ bosons which would couple preferentially to the third-generation of fermions, and hence, they would decay typically into tau pairs (see TAT models in section 1.2). In consequence, in these kind of models if a $Z'$ boson would exist, it would be observed first in the ditau channel and, therefore, they are the main motivation of searching for $Z'$ bosons decaying into taus. Additionally, in the case of models which predict the universality of couplings, or in which the difference of couplings of the $Z'$ boson to the third and the other lepton generations is small, these hypothetical bosons could be observed first using another fermion channel, such as $Z' \rightarrow \ell\ell$, and the searches for $Z'$ bosons in the ditau channel would reveal the nature of its couplings.
The benchmark models used for the search in this channel are $Z'_{SSM}$ and $Z'_{TAT}$.

The hypothetical $Z' \rightarrow \tau\tau$ events consist of two oppositely-charged high $p_T$ taus, produced back-to-back in the transversal direction. The searches for $Z'$ bosons in this channel involve different experimental signatures since, as will be explained further in section 1.4, the tau can decay leptonically or hadronically, depending if its decay products are leptons ($\tau_\ell$) or hadrons ($\tau_h$). Then, this search usually is performed with the combination of four of the possible signatures: when both taus decay hadronically $Z' \rightarrow \tau_h\tau_h$; when one tau decays leptonically and the other one decays hadronically, $Z' \rightarrow \tau_\ell\tau_h$ and $Z' \rightarrow \tau_\mu\tau_h$; and when both taus decay leptonically $Z' \rightarrow \tau_\ell\tau_\mu$. They contribute approximately to a 94% of the $Z'$ decay branching fraction (this thesis is focused on the $Z' \rightarrow \tau_h\tau_h$ channel). These searches are challenging from the experimental point of view since taus decay 66% of the times into hadrons; then, their signatures are similar than the ones produced by QCD-jets.

In consequence, the main background for these channel comes from QCD processes, which are not well modeled by simulation, and data-driven techniques must be used to estimate their contribution.

Several searches for $Z'$ bosons decaying into $\tau\tau$-pairs have been performed by the CDF [8] experiment at the Tevatron, and by the CMS [17–19] and ATLAS [29–32] experiments at the LHC. The CDF Collaboration excluded the $Z'_{SSM}$ boson in the ditau channel for masses below 399 GeV, using 195 pb$^{-1}$ of collected data [8].

The tightest constraints on the $Z'$ mass using the ditau channel have been set by the CMS and ATLAS Collaborations. For this channel, CMS has performed the search combining the four tau signatures mentioned above ($\tau_h\tau_h$, $\tau_\ell\tau_h$, $\tau_\mu\tau_h$ and $\tau_\tau\tau_\mu$) with the data collected during 2011, which corresponds to pp collisions at $\sqrt{s} = 7$ TeV with an integrated luminosity of 4.9 fb$^{-1}$. Upper limits were reported on the production of $Z'$ bosons, excluding its existence with 95% CL below 1.4 TeV for $Z'_{SSM}$ and 1.1 TeV for $Z'_{ψ}[17]$. Additionally, the CMS Collaboration has carried out the search for $Z'$ bosons in the $\tau_\ell\tau_\mu$ channel; this search was performed with the data collected during 2012, which corresponds to 19.7 fb$^{-1}$ of pp collisions at $\sqrt{s} = 8$ TeV. As a result, $Z'_{SSM}$ and $Z'_{ψ}$ were excluded for masses below 1.3 TeV and 0.81 TeV respectively [18]. The $Z' \rightarrow \tau\tau$ search was also performed with the data collected by CMS during 2015, combining the four ditau signatures. This search, which uses pp collisions at $\sqrt{s} = 13$ TeV with an integrated luminosity of 2.2 fb$^{-1}$, set upper limits with 95% CL for $Z'_{SSM}$ masses below 2.1 TeV and $Z'_{TAT}$ masses below 1.7 TeV (see Figure 1.3 right) [19]. The exclusion limits set by the $\tau_h\tau_h$ channel constrain the existence of the $Z'_{SSM}$ for masses up to 1.92 TeV and the $Z'_{TAT}$ for masses up to 1.51 TeV (see Figure 1.3 left). The search performed in the $\tau_h\tau_h$ channel is specially important since it will serve as a reference for the analysis performed in this thesis.

The ATLAS Collaboration has reported the tightest exclusion limits on the $Z'$ mass using the ditau channel since the search was performed with the data collected during 2015 and 2016, using pp collisions at $\sqrt{s} = 13$ TeV with an integrated luminosity of 36.1 fb$^{-1}$; ATLAS has excluded the $Z'_{SSM}$ for masses below 2.42 TeV with 95% CL (See Figure 1.4) [29].

The summary of the exclusion limits set on the $Z'_{SSM}$ mass in the ditau searches, performed by CMS and ATLAS experiments, are shown in Table 1.3.
Figure 1.3: Upper limit at the 95% CL on the product of the cross section and branching fraction into $\tau$ pairs as a function of the $Z'$ mass. The bands represent the one and two standard deviations. The figure shows the results by the CMS Collaboration for the $\tau_h \tau_h$ channel (left) and the combination of four ditau channels: $\tau_h \tau_h$, $\tau_e \tau_h$, $\tau_\mu \tau_h$ and $\tau_e \tau_\mu$ (right) [19].

<table>
<thead>
<tr>
<th>Data</th>
<th>$\sqrt{s}$ [TeV]</th>
<th>$\mathcal{L}$ [fb$^{-1}$]</th>
<th>ditau channel</th>
<th>Upper Limit for $Z'_{SSM}$ [TeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMS</td>
<td></td>
<td></td>
<td>$\tau_h \tau_h$, $\tau_e \tau_h$, $\tau_\mu \tau_h$, $\tau_e \tau_\mu$</td>
<td>1.4</td>
</tr>
<tr>
<td>2011</td>
<td>7</td>
<td>4.9</td>
<td>$\tau_h \tau_h$, $\tau_e \tau_h$, $\tau_\mu \tau_h$, $\tau_e \tau_\mu$</td>
<td>1.4</td>
</tr>
<tr>
<td>2012</td>
<td>8</td>
<td>19.7</td>
<td>$\tau_e \tau_\mu$</td>
<td>1.3</td>
</tr>
<tr>
<td>2015</td>
<td>13</td>
<td>2.2</td>
<td>$\tau_h \tau_h$, $\tau_e \tau_h$, $\tau_\mu \tau_h$, $\tau_e \tau_\mu$</td>
<td>2.1</td>
</tr>
<tr>
<td>ATLAS</td>
<td></td>
<td></td>
<td>$\tau_h \tau_h$, $\tau_e \tau_h$, $\tau_\mu \tau_h$, $\tau_e \tau_\mu$</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Table 1.3: Summary of the upper 95% CL limits set for $Z'_{SSM}$ and $Z'_{\psi}$ masses in the di tau channels by the CMS and ATLAS experiments.

- **Dijet ($Z' \rightarrow q\bar{q}$) searches.**

Since $Z'$ bosons might decay into quark-antiquark pairs, the CMS and ATLAS Collaborations have reported several searches that involve a jet-pair such as: $t\bar{t}$ [22], $b\bar{b}$ [23,34] and $jj$ [24,33]. The copious contribution of background from QCD multijet processes degrades the mass-resolution, making these channels less sensitive than the dilepton ones. In these searches the exclusion limits go from 1 TeV to 2.7 TeV depending on the model (see the references quoted before for more details.)
Summary.

In summary, the searches for $Z'$ bosons have not resulted in any confirmed observation, nor in any significant deviations from SM predictions, that would suggest their existence. Only lower exclusion limits on their masses have been set. LEP experiments set indirect exclusion limits on the $Z'$ mass for BSM in which they were predicted as a possible mixed state between $Z'$ and $Z$ bosons. In direct searches, the tightest limits on the $Z'$ mass have been set by the CMS and ATLAS Collaborations in the dilepton channel, where the $Z'_{\text{SSM}}$ and $Z'_{\psi}$ bosons have been excluded for masses below 4.5 TeV and 3.9 TeV respectively. Nevertheless, several BSM scenarios predict $Z'$ bosons that may have lower couplings with electrons and muons, and might decay preferably into $\tau$ pairs. The CMS and ATLAS experiments performed searches on $Z'$ decaying into $\tau\tau$ using samples from Run II, and have excluded the $Z'_{\text{SSM}}$ mass below 2.1 TeV (CMS) and 2.4 TeV (ATLAS). The purpose of this thesis is to search for $Z'$ bosons decaying into $\tau_h\tau_h$ using the data recorded by CMS during 2016.

1.4 The Physics of the tau-lepton

Since the final state of the channel analyzed in this work involves taus, some of the properties of this particle are presented in this section. The tau-lepton belongs to the third generation of fermions and is the heaviest lepton, with a mass of 1.777 GeV. It has a lifetime of $2.9 \times 10^{-13}$ s, and its decay length, $c\tau$, is 87 $\mu$m. It is the only lepton with enough mass to decay into leptons but also into hadrons. At the fundamental level, a $\tau$ decays into a neutrino by the emission of a virtual W boson ($\tau \rightarrow \nu \ W^*$). The $W^*$ decays 33% of the times leptonically, $W^* \rightarrow \nu \ell \ ($$\ell = e, \mu$); and it decays hadronically the remaining 67% of the times, $W^* \rightarrow q\bar{q}'$. In the hadronic case the $q\bar{q}'$ pair will form hadrons (mostly pions). Figure 1.5 shows the Feynman diagrams of
the tau decays and their branching fraction are listed in Table ??.

<table>
<thead>
<tr>
<th>Final State</th>
<th>Braching Fraction [%]</th>
<th>Resonance</th>
<th>Mass [GeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e^- \bar{\nu}<em>e \nu</em>\tau$</td>
<td>$17.83 \pm 0.04$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu^- \bar{\nu}<em>\mu \nu</em>\tau$</td>
<td>$17.41 \pm 0.04$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\pi^- \nu_\tau$</td>
<td>$10.83 \pm 0.06$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\pi^- \pi^0 \nu_\tau$</td>
<td>$25.52 \pm 0.09$</td>
<td>$\rho$</td>
<td>770</td>
</tr>
<tr>
<td>$\pi^- \pi^0 \pi^0 \nu_\tau$</td>
<td>$9.30 \pm 0.11$</td>
<td>$a_1$</td>
<td>1200</td>
</tr>
<tr>
<td>$\pi^- \pi^- \pi^0 \nu_\tau$</td>
<td>$1.05 \pm 0.07$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\pi^- \pi^- \pi^0 \pi^0 \nu_\tau$</td>
<td>$8.99 \pm 0.06$</td>
<td>$a_1$</td>
<td>1200</td>
</tr>
<tr>
<td>$\pi^- \pi^- \pi^0 \nu_\tau$</td>
<td>$2.70 \pm 0.08$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.4: Branching fraction of $\tau^-$ modes [45].

The tau identification is challenging from the experimental point of view for several reasons. As was mentioned above, the final states of the tau decays involve one or two neutrinos; however, it is not possible to observe neutrinos without an enormous amount of material and therefore, they can not be detected in experiments like CMS. As consequence, a fraction of the tau 4-momentum goes undetected, not allowing the full reconstruction of its kinematical parameters. Additionally, due to the relative small decay length, in the leptonic case, it is not possible to distinguish the tau decay vertex from the primary vertex of the collision. Therefore, in the leptonic decays, the presence of neutrinos and the absence of a secondary vertex makes it impossible to distinguish a charged lepton coming from a tau decay than one coming from other processes. In consequence, taus can only be identified directly from their hadronic decays.

The hadronic identification of taus is also challenging since the experimental signature is similar to the one of hadronic-jets. An hadronic-jet, or QCD-jet, comes from the strong interactions between quarks and leptons, resulting in an abundant amount of charged and neutral pions.
emitted within a cone. Since the $\tau_h$ decay produces a neutrino plus pions (charged and neutral ones), they can be misidentified as a QCD-jets. Besides, these jets are produced copiously in pp collisions and their yield is eight orders of magnitude higher than the one of $\tau_h$, constituting a large source of background for the tau identification. Nevertheless, there are some differences between a $\tau_h$ and a QCD-jet, that allow for a good level of discrimination: for instance, a QCD-jet has a wider energy profile than a $\tau_h$; the hadronic tau decay also has a narrower cone and fewer charged particles than a QCD-jet. These differences can be exploited to implement algorithms in order to distinguish between them.

Since the purpose of this work is to search for $Z'$ bosons in the dihadronic tau decay channel, the tau identification in the CMS experiment is crucial. Chapter 2 describes the CMS experiment, with a special focus on the experimental signatures of the tau decay products, and Chapter 3 shows the algorithms used for CMS in order to identify the particles that emerge from a collision, making emphasis in the tau identification algorithm, which is described in section 3.7.
2 The CMS experiment

The CMS and ATLAS experiments are the biggest multi-purpose particle detectors ever built in the world and are part of the LHC [46]. The LHC is a proton collider located at the CERN laboratory in Switzerland. The unprecedented centre-of-mass energy of the proton-proton collisions produced by this accelerator and its very high luminosity, make possible the search for physics BSM in a new kinematic regime. The data used for this dissertation comes from pp collisions produced by the LHC at $\sqrt{s} = 13$ TeV, and recorded by the CMS experiment during 2016. In the present Chapter, the LHC accelerator and the CMS experiment will be described, making emphasis on the CMS sub-detectors involved in the tau-lepton detection, which plays an important role in this work.

2.1 LHC Accelerator

The LHC is a particle accelerator designed to collide protons (or lead ions) at a centre-of-mass energy up to 14 TeV. The LHC accelerates protons along two rings with a circumference of 27 km, installed in a tunnel approximately 100 m underground. Bunches of protons (or ions) are accelerated in opposite directions using ratio frequency (RF) cavities along the rings. They collide in four different points, where dedicated experiments are placed in order to detect the products of the collisions. The four experiments are: CMS [47, 48], ATLAS [49], LHCb [50] and ALICE [51] (see Figure 2.1). CMS and ATLAS are multi-purpose detectors, optimized for the discovery of new physics BSM. The aim of the LHCb detector is to study the charge-parity (CP) symmetry violation, which has been postulated to explain the origin of matter-antimatter asymmetry in our universe. ALICE is specialized in studying the quark-gluon plasma. Besides these experiments, there are two additional smaller ones: TOTEM and LHCf. The TOTEM main goal is the accurate measurement of total, elastic and diffractive pp cross sections. LHCf uses the particles emitted forward by collisions in order to simulate the cosmic rays behavior in controlled conditions.

2.1.1 LHC Proton Accelerator Chain

In order to achieve the very high energy of the LHC proton beams, several pre-acceleration stages are used as shown in Figure 2.2. Before protons are injected into the two LHC rings, the process starts with the extraction of protons in the Duoplasmatron Proton Ion Source; in this source, the protons are extracted in “bunches” by ionization of hydrogen gas. Then, these bunches are injected into a linear accelerator, Linac2, where their energy is increased up to 50 MeV. Once the protons reach such energy, they are delivered to the Proton Synchrotron Booster (PSB) and then to the Proton Synchrotron (PS) where they are accelerated up to 1.4 GeV and 25 GeV, respectively. Additionally, in the PS the bunches are spaced in time by 25 ns [†]. The pre-acceleration chain finishes in the Super Proton Synchrotron (SPS), where the protons reach an energy up to 450 GeV, and then they are injected into the LHC rings. The bunches are injected into the LHC rings in opposite directions where they reach an energy up to 7 TeV per bunch. The time spacing between bunches is known as Bunch Crossing (BX). For the LHC Run I, the BX was 50 ns; for Run II the BX was reduced to its design value of 25 ns.

†The time spacing between bunches is known as Bunch Crossing (BX). For the LHC Run I, the BX was 50 ns; for Run II the BX was reduced to its design value of 25 ns.
beam, i.e., the LHC could produce proton-proton collisions up to 14 TeV in the center-mass frame ($\sqrt{s} = 14$ TeV). The protons are accelerated using 8 RF cavities, which operate with a frequency that goes up to 400 MHz in the final acceleration stage and an electric field gradient of 5 MeV per meter. The RFs also ensure the longitudinal stability of the beams. The proton bunches are radially focused by 392 quadrupole magnets placed along the LHC rings. The radial focusing is important to increase the collision probability. Additionally, the LHC uses 1232 superconductor dipole magnets, where each of them produce a magnetic field strength up to 8.3 T, in order to bend the beam along the circular path. The dipoles are cooled by superfluid helium at a temperature of 1.9 K, that also helps to improve the vacuum inside the beam pipes.

2.1.2 LHC Operational Parameters

The main quantities that describe the performance of a particle accelerator are the beam energy and the luminosity. The instantaneous luminosity measures the number of collisions per unit of area and per unit of time; it depends on the accelerator’s design and, in the case of the LHC, it is given by:

$$L = \frac{N^2 k_b f_{rec} \gamma}{4\pi\epsilon\beta^*} R,$$

(2.1)

†The units of instantaneous luminosity are cm$^{-2}$s$^{-1}$.
where $N_b$ is the number of protons per bunch, $k_b$ is the number of bunches per beam, $f_{rev}$ is the number of revolutions per second, $\gamma$ is the relativistic factor, $\epsilon$ is the normalized transverse beam emittance, $\beta^*$ is the optical $\beta$-function at the collision point and $R$ is a geometrical factor related with the crossing angle of the two beams. The optimization of the instantaneous luminosity is achieved in several ways, such as by increasing the number of protons per bunch ($N_b$) or by increasing the number of bunches per beam ($k_b$). For the statistical accuracy of the physics measurements, the important quantity is the integrated luminosity $L$ (instantaneous luminosity integrated in time, $L$) because the number of events produced ($N_{ev}$) is proportional to it, and it is given by the expression:

$$N_{ev} = \mathcal{L}_{e} \sigma_p,$$

(2.2)

where $\sigma_p$ is the cross-section of the process of interest, for example the cross-section of an hypothetical $Z'$ production, $\sigma(pp \rightarrow Z')$. If the cross-section that we want to measure is small, we need a large integrated luminosity in order to achieve a reasonable number of events, even more, if the process demands for a complex discrimination between signal and background. For instance, a $Z'$ decaying into two taus should have a cross-section of the order of pb, while Drell-Yan process, that constitute an important background, has a cross-section of around 5780 pb$^{11}$. Therefore, the signal discrimination in this case should be of the order of $10^{-3}$, demanding even higher luminosities. In addition, in each BX there are several pp collisions (of the order of 20, during 2016 data taking period), and only a very small fraction of these collisions correspond to

$^1$The units of the integrated luminosity are inverse barn, $b^{-1}$.

$^{11}$In this example, the cross-sections are estimated for pp collisions at $\sqrt{s} = 13$ TeV and for a $Z'$ with a mass of 500 GeV.
hard proton-proton interactions that could produce interesting physics. Most of the collisions correspond to elastic and diffractive scattering as well as soft scattering. This means that in a BX with an interesting hard scattering collision there would be another 20 soft scattering ones, that will have to be identified and removed from the event. This phenomenon is known as “pile-up” (PU) and makes difficult the identification of the hard interaction. The level of PU depends on the number of protons per bunch $N_b$, as well as on the geometrical factors of the beam.

In order to achieve the high statistics for the identification of BSM signals, a good performance of the LHC accelerator is required. Operational quantities such as beam energy, luminosity, bunch crossing, etc., have been improved from one running period to the next. Table 2.1 shows the most relevant quantities of the LHC operation per year.

<table>
<thead>
<tr>
<th></th>
<th>Design</th>
<th>Run I</th>
<th>Run II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beam Energy [TeV]</td>
<td>7</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>Number of protons per bunch [$10^{11}$]</td>
<td>1.15</td>
<td>1.0</td>
<td>1.3</td>
</tr>
<tr>
<td>Number of bunches per beam</td>
<td>2808</td>
<td>368</td>
<td>1380</td>
</tr>
<tr>
<td>Bunch Spacing [ns]</td>
<td>25</td>
<td>150</td>
<td>50</td>
</tr>
<tr>
<td>Average Pile-up in CMS</td>
<td>0.021</td>
<td>0.35</td>
<td>0.77</td>
</tr>
<tr>
<td>Maximum peak luminosity [$10^{34}$ cm$^{-2}$s$^{-1}$]</td>
<td>1.0</td>
<td>0.048</td>
<td>5.5</td>
</tr>
<tr>
<td>Integrated Luminosity [fb$^{-1}$]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Some relevant operational parameters for Run II, compared with values reached during Run I [52].

Run I and Run II of the LHC

The LHC operational period spanning from 2010 to 2012 is known as Run I. During 2010, the LHC reached pp collisions at centre-of-mass energy of 7 TeV and an integrated luminosity of around 50 pb$^{-1}$. During 2011 with the same centre-of-mass energy, an integrated luminosity of approximately 6 fb$^{-1}$ was obtained. For the 2012 run, the energy per beam was increased up to 4 TeV ($\sqrt{s} = 8$ TeV), reaching approximately 20 fb$^{-1}$. The integrated luminosity delivered by the LHC and recorded by CMS during Run I is shown in Figure 2.3. Afterwards, a two-years technical shut-down took place.

After the shut-down, a new data taking period, known as Run II, started on May 2015. That year, LHC delivered 4.2 fb$^{-1}$ of pp collisions at $\sqrt{s} = 13$ TeV. During 2016, LHC operated at the same centre-of-mass energy than during 2015 and reached 40.82 fb$^{-1}$, exceeding in 53% the expected integrated luminosity (∼26 fb$^{-1}$). The integrated luminosity delivered by the LHC and recorded by CMS, during 2015 and 2016, is shown in Figure 2.4. The success of the 2016 run was due to the LHC operational stability and the high peak luminosity reached ($1.4 \times 10^{34}$ cm$^{-2}$s$^{-1}$, which represented an improvement of 40% over the design value); this was achieved due to the shortening of the beam size from the injectors and the reduction of the crossing angle between the two beams. Compared with the 2015 run, the 2016 run reduced the number of protons per bunch to $1.1 \times 10^{11}$ and the number of bunches per beam from 2244 to 2200, keeping the BX at 25 ns. As a result, the average PU went from 14 in the 2015 run to 27 in the 2016 run (see Table 2.1). As already mentioned, the data used for this dissertation is the one recorded by CMS during 2016 run.
2.2 The CMS Detector

The Compact Muon Solenoid (CMS) is, along with ATLAS, a multi-purpose detector designed with a broad physics program, which includes the understanding of the electroweak symmetry breaking through the Higgs mechanism, and the search for physics BSM, such as SUSY, extra dimensions, etc. The CMS detector is located 100 m underground in “Point 5” of the LHC (see Figure 2.1). It is a hermetic detector around the collision point, with a cylindrical shape that has a length of 21.6 m and a diameter of 14.6 m. One of the especial features of the detector is the superconducting solenoid which produces an inner magnetic field of $3.8 \, \text{T}$ (over a volume of $341.7 \, \text{m}^3$). The strong magnetic field bends the tracks of the charged particles coming from the interactions, with the purpose of identifying their electric charge and to accurately measure their momentum. Besides the solenoid, the detector is composed by four subsystems: encased inside the solenoid are the Tracker System, the Electromagnetic Calorimeter (ECAL) and the Hadron Calorimeter (HCAL); and outside the magnet are the Muon Chambers. The purpose of the Silicon Tracker is to reconstruct the collision vertices and the tracks of the charged particles emerging from the collision. The Calorimeter system (ECAL and HCAL) allows to measure the energy of hadrons, electrons and photons. The Muon Chambers, embedded inside an iron-yoke structure, reconstruct the muon tracks and provide an accurate information for their momentum measurement. The overall layout of CMS detector is shown in Figure 2.5. A more detailed description can be found in Ref. [54].

All the information produced in an event is stored in the readout electronics of each sub-detector. Once the event information is compressed, the data size for one bunch crossing is around 1 MB and, therefore, with the nominal LHC luminosity, CMS would produce 40 TB of information per second; this high rate makes impossible the data storage with the current technology. Nevertheless, interesting physics can be produced only in hard-interaction collisions, which occur approximately once each one million collisions. In consequence, CMS uses a Trigger System to select only the hard scattering events which are reduced from 40 million collisions per second to...
only 100 events per second, making feasible the data storage. Once the Trigger System identifies the interesting events, they are stored in order to be analyzed afterward.

**Coordinate System**

In order to define the position of any detector components and, in consequence, of any particle signal, CMS has defined a cartesian and spherical coordinate systems. The origin of both systems is the nominal interaction point, which is at the center of the detector. The x-axis points towards the center of the LHC ring, the y-axis points upwards and the z-axis points along the beam pipe in the counterclockwise direction. The polar angle \( \theta \) and azimuthal angle \( \phi \) are defined in the usual way.

In order to parametrize the direction in which the particles are emitted, the pseudorapidity \( \eta \) is better than \( \theta \) because its distribution is more uniform. Pseudorapidity is given by:

\[
\eta = -\ln \left( \tan \left( \frac{\theta}{2} \right) \right)
\]  

A more detailed description of the CMS coordinate system can be found in Ref. \[54\].

**2.3 Superconducting Solenoid**

The superconducting solenoid produces a uniform inner magnetic field with the value of 3.8 T. In the outer region, the returning magnetic flux is compactified by the iron yoke, resulting on an average magnetic field strength of 2 T. The magnetic field provided by the solenoid bends the tracks of the charged particles emerging from the collisions (see Figure 2.6), which is crucial for their charge identification and their momentum measurement. The electric charge is identified...
The momentum of a charged particle that moves through a uniform magnetic field is given by:

\[ p = \gamma m v = q B r, \]  

(2.4)

where \( \gamma \) is the relativistic factor; \( m, v, q \) are its mass, rapidity and charge, respectively; \( B \) is the magnetic field strength; and \( r \) is the ratio of the bending. A strong magnetic field is necessary for the momentum measurement of very energetic charged particles, in consequence the momentum resolution depends on the magnetic field strength and the spatial resolution of the detectors that reconstruct the tracks. The momentum resolution is:

\[ \frac{\sigma_p}{p} \propto \frac{p}{BL^2}, \]  

(2.5)

where \( L \) is the length of the trajectory. The strong magnetic field provided by the solenoid and the high spatial resolution of the tracker detectors, allows CMS to have a very good momentum resolution; for instance, the momentum resolution for muons is 1% up to 100 GeV [48].

The CMS superconducting solenoid is the biggest one built ever. The coil is made of NbTi with 4 layers of winding. In order to generate the strong magnetic field, this solenoid operates with a nominal current of 19.5 kA, storing an energy up to 2.6 GJ. A cooling system keeps its superconducting state using liquid helium at 4.65 K. The main parameters of the CMS magnet are summarized in Table 2.2.
Figure 2.6: Schematic view on CMS transverse plane for trajectories and energy deposits of several particles moving through the solenoid magnetic field. Taken from [55].

<table>
<thead>
<tr>
<th>Magnet Parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Inner magnetic field</td>
<td>3.8 T</td>
</tr>
<tr>
<td>Diameter</td>
<td>5.9 m</td>
</tr>
<tr>
<td>Length</td>
<td>12.5 m</td>
</tr>
<tr>
<td>Nominal Current</td>
<td>19.5 kA</td>
</tr>
<tr>
<td>Stored Energy</td>
<td>2.7 GJ</td>
</tr>
<tr>
<td>Inductance</td>
<td>14.2 H</td>
</tr>
</tbody>
</table>

Table 2.2: Main parameters of the CMS Solenoid [47].

2.4 The Tracker System

The Tracker System is the inner-most detector system in CMS. It has a cylindrical shape covering an acceptance range of $|\eta| < 2.5$. This system is designed to reconstruct accurately the charged-particle tracks, which is essential for achieving a high momentum resolution as well as for identifying the primary and the secondary vertices. Secondary vertices are the evidence of the decay of long-lived particles coming out the collision, for instance jets originated from b-decays, known as b-jets, which have a vertex that can be distinguished from the primary vertex. Therefore a high resolution of secondary vertex reconstruction is required in order to identify b-jets (see Section 3.5). In this dissertation the Tracker System is crucial due to the following reasons:

- The searches for heavy resonances in the dilepton channels require a high momentum resolution for leptons with transverse momentum greater than 1 TeV.
- Since the tau-leptons can decay into charged hadrons (see Section 1.4), the reconstruction of
these hadrons in the Tracker System is crucial for the tau identification and reconstruction. Charged hadrons are reconstructed with an efficiency of at least 95% for $p_T$ greater than 10 GeV (see Section 3.1.1) [56].

- The transverse impact parameter and secondary-vertex resolution are comparable with the distances that tau-leptons travel before decay [57], in consequence, there is a probability that a b-jet can fake the tau signature (see Section 4.2). The high b-jet identification efficiency provided by the Tracker System will reduce this source of background.

- Due to the material of the Tracker System, there is a high probability that photons coming from $\pi^0 \rightarrow \gamma \gamma$ convert into electron-positron pairs. The interaction of these photons with the Tracker material is important for jet and tau reconstruction, since $\pi^0$'s are copiously produced in QCD-jets and they are also one of the tau decay products (see Section 1.4).

Since the Tracker System is the closest detector to the interaction point, it is exposed to the highest radiation doses in CMS; in average 1000 particles from 27 proton-proton collision are passing through this system each 25 ns. In consequence, the Tracker System was designed to achieve a high granularity and a fast response in order to identify the tracks and associate them to the proper BX. Additionally, this system was designed to have a high level of radiation hardness. Since these requirements are fulfilled by silicon detectors, the Tracker System is based on this technology. This system is composed of two subsystems: the Pixel Tracker and the Strip Tracker.

The Pixel Tracker

The Pixel Tracker is the closest subsystem to the beam pipe. It is composed by three concentric layers with a radius of 4.4 cm, 7.3 cm and 10.2 cm, and each one has a length of 53 cm. In the forward regions, there are two disks in each endcap which are located at $|z| = 34.5$ cm and $|z| = 46.5$ cm from the interaction point (see Figure 2.7). The whole subsystem has approximately 66 million pixels, covering an active surface area of around one squared meter. Each pixel cell has a size of $100 \times 150$ $\mu$m$^2$, providing a high spatial resolution in the $r-\phi$ plane and also in the $z$ direction. The Pixel detector has an acceptance range of $|\eta| < 2.5$ [54].

The Strip Tracker

Surrounding the Pixel Tracker is the Strip Tracker, which is made of 9.6 million strips sensors, covering an active detection area of 198 m$^2$. This subsystem is divided in two components: the inner tracker (20 cm $< |z| < 55$ cm) and the outer tracker (55 cm $< |z| < 116$ cm). The inner tracker is made of 4 cylindrical layers, called Tracker Inner Barrel (TIB), and 3 disks installed in each endcap, known as Tracker Inner Disks (TID). The outer tracker is composed of 6 layers, called Tracker Outer Barrel (TOB), and 9 disks in each endcap, known as Tracker EndCaps (TEC). The whole Strip Tracker has a diameter of 2.4 m and a length of 5.5 m, covering an acceptance range of $|\eta| < 2.5$ (see Figure 2.7). The spatial resolution depends on the location of each strip sensor within the subsystem, for instance, the spatial resolution provided by the TIB which varies from 23 to 34 $\mu$m in the $r-\phi$ plane and 23 $\mu$m in the $z$ direction.

Particle interaction with the Tracker material

When a particle passes through the Tracker System it interacts not only with the active volume of the detector, but also with the other components such as the read-out electronic, the mechanical structure, the services and the cooling system. This amount of material is significant and the interaction with the crossing particles must be considered. Figure 2.8 shows the thickness (in
terms of number of radiation lengths) of tracker material that a particle must pass through before reaching the ECAL. As a result, there is a high probability that photons, coming from $\pi^0 \rightarrow \gamma\gamma$, can convert into electron-positron pairs which, as mentioned above, is important for jet and tau reconstruction.

## 2.5 The Calorimeter System

The purpose of the Calorimeter System is to stop most of the particles coming out from the collision and to measure their energy with a good granularity. Electrons, photons, and hadrons interact with the calorimeter material, depositing all their energy on the detectors which is sampled and measured. The only SM particles that scape from the Calorimeter System are muons and neutrinos: muons deposit a very low amount of energy in this system, and continue traveling beyond to the detectors installed in the outer part of CMS, designed to reconstruct their tracks. In the case of neutrinos, they escape without detection since they only interact weakly, leaving an energy imbalance in the event; therefore, this imbalance is an indirect evidence of their presence or the presence of any particle that only interacts weakly.\[1\]

The CMS Calorimeter is divided in two subsystems: the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). The ECAL is the closest calorimeter subsystem to the interaction point; it is designed to measure the energy of electrons and photons. The reconstruction of these particles is essential for many physics analyses, for instance for those in which jets or taus are involved. In these analyses, the photon and electron reconstruction is crucial for the energy measurement of the jet and for the tau identification. The HCAL, the other calorimeter subsystem, which is located just outside of the ECAL, is designed to measure the energy of the hadrons. It plays a fundamental role in jet-like objects: QCD-jets, b-jets, tau-jets; being able to differentiate among these objects, is crucial, in particular for these analyses. Figure 2.6 shows the arrangement of the Calorimeter System and the energy deposits of SM particles on it.

---

†A big amount of energy imbalance would be an evidence of dark matter candidates, since they only participate on weak interactions.
2.5.1 The Electromagnetic Calorimeter

The Electromagnetic Calorimeter is designed to absorb the total energy of electrons and photons when they pass through it. The interaction between these particles and the calorimeter material produces a cascade of electrons and photons in a process called electromagnetic shower. An electromagnetic shower starts when an energetic electron, or photon, enters into the high density medium of the calorimeter and starts to lose energy due to interactions with the medium. Electrons and positrons lose their energy by bremsstrahlung radiation, while photons lose their energy by electron-positron conversions. Therefore, a cascade of secondary particles is created and the energy losses continue through these two processes until the energy of the particles is not enough to produce new ones. As a result, the energy deposits of the electromagnetic shower are spread over the calorimeter material and are sampled and measured using scintillators. The profile (transverse and longitudinal) of the electromagnetic shower allows to determine the energy of the incoming particle. Electromagnetic showers are then described by two parameters which depend on the calorimeter material: the radiation length ($X_0$) and the Molière radius. The radiation length is the distance than an electron or photon travels until its energy is reduced by a factor of $1/e$, while the Molière radius is the radius of a cylinder where 90% of the electromagnetic shower is contained. In consequence, the calorimeter is designed with enough thickness (in terms of radiation lengths) to measure the total energy of the particles, and with a Molière radius small enough to achieve a good granularity. The schematic view of an electromagnetic shower is shown in Figure 2.9.

The CMS ECAL has cylindrical and hermetic design with scintillators made of lead tungstate crystals (PbWO$_4$), see Figure 2.10. The PbWO$_4$ crystals are characterized by its high density (8.28 g/cm$^3$), its short radiation length (0.89 cm) and its small Molière radius (2.2 cm), making them appropriate to achieve accurate energy measurements with a good granularity [54]; additionally, these crystals have excellent radiation-hardness and have a fast response, which make
them suitable for the LHC environment. When electrons or photons cross these crystals, they emit a light pulse with a time response of $\sim 25$ ns (about 80% of the times). The scintillator light is collected by photo-detectors that convert it into an electric signal. The photo-detectors used in the barrel are Avalanche Photo-Diodes (APDs) while in the endcaps are Vacuum Photo-Triodes (VPT).

The ECAL has 61200 crystals in the barrel and 7324 crystals in each endcap [47], covering a $|\eta|$ range up to 3. The crystals are installed in a quasi-projective geometry in such a way that each of them points toward the center of the detector, plus an additional angle of $3^\circ$, with the purpose of avoiding particles passing through inactive regions of the ECAL. In the barrel region, the calorimeter (ECAL Barrel or EB) has a inner radius of 129 cm, covering a pseudorapidity region of $|\eta| < 1.479$. The EB crystals are 230 mm long (25.8 $X_0$) and cover a cross-section at the front face of $0.0174 \times 0.0174$ in the $\eta - \phi$ plane (which corresponds to $22 \times 22$ mm$^2$). In the forward region, the ECAL Endcaps (EE) are located at 314 cm from the collision point, covering a pseudorapidity range of $1.479 < |\eta| < 3.0$. The EE crystals are 220 mm long (24.7 $X_0$) and cover a cross-section of $28.62 \times 28.62$ mm$^2$. In addition to the crystals, Preshower Detectors are installed in front of endcaps, covering a pseudorapidity range of $1.65 < |\eta| < 2.6$. This detector
consists of two lead layers, each one followed by silicon sensors. Their main goal is to identify the two photons produced by neutral-pion decays in the forward region. The complete ECAL arrangement is shown in Figure 2.11.

![Figure 2.11: Arrangement of the ECAL components.](image)

The energy resolution reached by the ECAL is given by:

\[
\left( \frac{\sigma_E}{E} \right)^2 = \left( \frac{2.8\%}{\sqrt{E}} \right)^2 + \left( \frac{0.12\%}{E} \right)^2 + (0.3\%)^2 ,
\]

where the first term comes from the stochastic nature of electromagnetic showers, the second term is related to electronic noise and the third term is related to the systematic uncertainty produced by the calibration of the apparatus. As can be inferred from Eq. 2.6, the energy resolution for energetic particles is dominated by the systematic uncertainty, while the stochastic and noise terms are dominant for low energies. As an example, the energy resolution obtained in the EB is close to 1% for all electrons that come from Z decays [59].

### 2.5.2 The Hadronic Calorimeter

The Hadron Calorimeter (HCAL) is designed to stop the hadrons and to measure their energy. It is composed of fluorescent scintillators inserted in layers of a dense material called the absorber. The absorber is made of stainless steel and copper layers. When a hadron hits the absorber it produces a cascade of particles, known as hadronic shower. The hadronic shower is a cascade of secondary particles, mainly pions, originated by strong interactions between the hadron and the atomic nuclei of the absorber material. Due to copious pion production, the shower also acquires electromagnetic components through neutral pion decays into photons. As the hadronic shower develops, the particles are detected by the scintillators, producing light pulses that are collected by optical fibers. The amount of light collected is proportional to the amount of energy deposited in the scintillator material, allowing the energy to be measured. A hadronic shower is described by the absorption length, which is the average distance traveled by the hadron through the medium before it interacts with the absorber; it is given by:

\[
\lambda = \frac{A}{N_A \sigma_{abs}} ,
\]

where \( A \) is the nuclei weight, \( N_A \) is the Avogadro’s number and \( \sigma_{abs} \) is the absorption cross-section.
In the central pseudorapidity region (HCAL Barrel, HB) the volume is restricted due to the solenoid, for this reason, part of the calorimeter is installed outside (Hadron Outer Calorimeter, HO) to achieve the desired level of the hermeticity. In the forward region there are two subsystems installed: the HCAL Endcaps (HE), which covers the pseudorapidity range of $1.4 < \eta < 3$; and the HCAL Forward (HF), which extends the $\eta$ coverage up to 5.2. The HCAL structure is shown in Figure 2.12.

The HB is divided into two identical halves in the $z$-direction. Each half is composed of 18 wedges, covering a range of $|\eta| < 1.4$. Each wedge consists of brass absorber plates staggered with plastic scintillators. The brass is a non-magnetic material and it has a small radiation length ($\lambda = 16.42$ cm), making it appropriate for the CMS environment. Between every two layers of absorbers, there is plastic scintillator with a thickness of 3.7 mm; the scintillator sends light pulses through optical-fibers (WaveLenght Shifter, WLS) to the photo-detectors (Hibryd Photodiodes, HPD). The photo-detectors convert the light into an electric signal. The innermost and outermost plates of the wedge are made of stainless steel in order to strengthen the structure. Each wedge is segmented in the $\eta$-direction by 16 structures called towers; each tower covers an identical solid angle whose area in the $\phi - \eta$ plane is $0.087 \times 0.087$.

The HO is located outside the solenoid, covering $|\eta| < 1.26$. It uses as absorber material the solenoid itself and the inner most layer of the iron-yoke structure. The HO is divided into 5 wheels in the $z$-direction, similar to the geometrical disposition of the Muon Chambers (see Section 2.6). In the central wheel of the detector, there are two layers of plastic scintillators staggered between the solenoid and the inner layer of the iron yoke. Additionally, there is one scintillator installed in each remaining wheel just behind the solenoid. Figure 2.12 shows the disposition of the HO plastic scintillators.

The HE is a sampling detector similar to the HB, composed of brass layers staggered with plastic scintillators of 3.7 mm of thickness. The HF is located at 11.2 m from the interaction point in the...
z-direction, covering an extended η region up to 5. The purpose of the HF is to achieve a better hermeticity, which improves the reconstruction of very forward jets as well as the measurement of the energy imbalance of the event.

Due to the nature of a sampling detector like the HCAL, its energy resolution is worse than the one obtained by the ECAL. For example, for pions, it is [54]:

\[
\left( \frac{\sigma_E}{E} \right)^2 = \left( \frac{138\%}{\sqrt{E}} \right)^2 + (13\%)^2.
\]  

Although the HCAL does not provide a high energy resolution, it is not crucial since this information is combined with the one of other detector systems.

### 2.6 The Muon System

The purpose of the Muon System is to identify and reconstruct muons but also, to generate trigger signal for the data acquisition system. This system is located outside the solenoid since the muons are the only known charged particles than can cross the tracker, the calorimeters and the solenoid material. In order to achieve the muon reconstruction and to trigger with them, CMS has three different kinds of Muon Detectors: Drift Tubes (DTs) in the barrel, Cathode Strip Chambers (CSCs) in endcaps and Resistive Plate Chambers (RPCs) in both, in the barrel and the endcaps. The DTs and the CSCs are used for tracking purposes. The RPCs are used for triggering due to their high time resolution. The Muon System structure is shown in Figure 2.13. In summary, the three technologies used in the Muon Chambers are:

- **Drift Tubes**, or DTs, are gaseous detectors used for tracking due to their excellent spatial resolution. The DT basic cell is a tube filled with a gas mixture of 85% Ar and 15% CO₂. The anode is an aluminum wire placed longitudinally in the center of the tube. The tube has a cross section of 13 × 22 mm² and a length between 2 to 3 m, depending on the position of the chamber in the barrel. When a charged particle passes through the cell the gas is ionized and produces an avalanche of electrons, because of the high electric field close to the wire. The signal deposit by the electrons in the wire is later amplified.

- **Cathode Strip Chambers**, or CSCs, perform the muon track reconstruction in the endcaps. Each CSC chamber has a trapezoidal shape in order to cover the 12 azimuthal sectors of the detector. A CSC Chamber has 6 planes of wires along the azimuthal direction, defining the radial coordinate of the track. The wire planes are intercalated with 7 panels of cathode strips that run radially and give the φ measurement with a resolution of 10 mrad. There are a total of 540 CSCs in the CMS detector.

- **Resistive Plate Chambers**, or RPCs, are gaseous parallel-plate detectors with a time resolution of about 1 ns. These detectors have an important roll for the triggering system, since they have excellent time resolution suitable for the identification of the BX associated to the detected muon. There are 6 RPC layers in the barrel and 4 disks in each endcap (see Figure 2.13).

In the barrel region, the Muon Chambers (DTs and RPCs) are inserted into the five wheels of the iron yoke; they form four concentric cylindrical layers, each of them divided on 12 sectors in φ. Then, the geometrical arrangement of the Muon System consists on 5 wheels in the z-direction,
Figure 2.13: Longitudinal view of CMS, showing the Muon System layout: DT Detectors are represented in pink, CSCs are green and RPCs are blue. GEM Detectors, that will be installed in LS2, are in red.

4 layers in the $r-$direction and 12 sectors in the $\phi-$direction. In the forward region, the Muon Chambers (CSCs and RPCs) are arranged into 4 disks in each endcap, where each disk is divided into 12 sectors in the $\phi-$direction and 4 rings in the $r-$direction.

The use of two types of detectors, DTs and RPCs in the barrel and CSCs and RPCs in the endcap improve the muon reconstruction resolution and efficiency. The RPCs provides a fast response ($< 25$ ns), allowing the identification of the BX, while the DTs and CSCs provide a high spatial resolution. The information obtained from the Muon Chambers is combined with the one provided by the Tracker system in order to improve the muon identification efficiency and the muon reconstruction. For instance, the momentum resolution for muons of $p_T = 10$ GeV is 10% using only the Muon Chambers information, while it becomes 1% when it is combined with the Tracker System information [47].

During the Long Shut Down 2, scheduled to 2019 (LS2), Gas Electron Multipliers (GEMs) will be installed in the endcaps in order to improve the muon identification efficiency and the momentum resolution in these regions (see Figure 2.13). In addition to their high granularity, GEMs will be used for triggering, due to their high time resolution.

### 2.7 The Trigger and Data Acquisition Systems

The Trigger and the Data Acquisition (DAQ) Systems use the information provided by the Tracker, the Calorimeter, and the Muon Systems in order to identify the events with interesting physics and to store them to be analyzed afterward. The information of an event is stored in
the read-out electronics of each sub-detector system; when these pieces are combined, the data size of an event is around 1 MB (1 MB per BX). Considering that the LHC operates with a collision rate of 40 MHz at nominal luminosity, CMS could deliver up to 40 TB of information per second. The data storage for such high rate is impossible with the current technology, but also only the hard scattering collisions (interesting events) should be selected and stored. These collisions happen very seldom, in consequence, the Trigger System is used to select only those events, reducing the rate from 40 MHz to 100 Hz, and making it feasible to store that data using the Data Acquisition System (DAQ).

2.7.1 The Trigger System

The Trigger System performs the selection of the hard scattering events in two steps: the Level-1 Trigger (L1 Trigger) and the High Level Trigger (HLT). The first step, the Level-1 Trigger, is performed by programmable read-out electronics in order to reduce the event rate from 40 MHz to 100 kHz. The second step, the HLT, is performed by about one thousand processors, which through software algorithms reduce the event rate up to 100 Hz.

The Level-1 Trigger

The Level-1 Trigger performs the first event selection through hardware criteria programmed in the read-out electronics of each sub-detector. A hard scattering event typically involves heavy particle production such as b-quarks, t-quarks, heavy bosons or high-p_T muons; then, interesting physics events usually contain high p_T jets and muon signals. In consequence, the Level-1 Trigger searches for signals into the read-out electronics of the Calorimeter System and the Muon System. The information provided by these two sub-detectors is combined and sent to the Global Trigger System; this system decides if the event is accepted sending a signal, known as L1-Accept, to the read-out electronics, so that it stores the information of the event†. Once the event is stored by the DAQ System, the data is moved to the HLT, where a second level of triggering is performed. Figure 2.14 shows the data flux through the L1 Trigger System.

The High Level Trigger

As mentioned above, the events selected by the L1 Trigger are delivered to a farm of about one thousand processors, where the High Level Trigger is executed. Unlike the L1 Trigger, which is implemented in hardware, the HLT selection is implemented in software since more complex algorithms are applied. The selection performed by the HLT starts with the readout of the information coming from each sub-detector system; then, the processors apply algorithms to the data in order to perform a quick reconstruction of the physics objects in the event; finally, the hard scattering events are identified. The events that pass the HLT trigger, about 100 events per second, are stored in disk. The stored data is labeled according to the physics objects identified on each event; for instance the triggered events due to the presence of a muon are labeled as SingleMuon dataset, and hence analyses which involves muons in the final state only use the muon dataset instead of all the dataset stored by CMS.

†If the L1-Accept signal is not produced for an event, the data associated to that event is erased from the readout electronics memory.
2.7.2 Data Acquisition System

The Data Acquisition System (DAQ) stores the events selected by the Trigger System. Since the read-out electronics of each sub-detector can store the information of an event only for 3.2 µs, the data will be erased unless a L1-Accept signal is received before that time. However, while the Trigger System is deciding if the event is selected or discarded, the read-out electronics must keep storing the information of the subsequent collisions; for this purpose electronic devices known as Front End Systems (FES) are used, which store data continuously in 40-MHz pipelined buffers. Once the L1-Accept signal reaches the FES, the data is pushed into the DAQ system via Front End Driver devices (FED). The DAQ System uses Front-end Read-out Link devices (FRL) in order to combine the information provided by several FEDs of each sub-detector, resulting in a data file suitable for the HLT processing. Finally, the event information is delivered to the farm of processors through optical links, where the HLT trigger is applied. The DAQ system is designed to operate with same input frequency than the L1 Trigger and to deliver data to the HLT System.

2.7.3 Data Processing

The data recorded by the DAQ system results in a data file which contains all the information for the accepted events. The data delivered by the HLT have a RAW format, which is not suitable for physics analysis; therefore a reconstruction process is performed on it in order to identify all the physics objects and determining their kinematics variables. CMS uses the Particle Flow algorithm in order to reconstruct every particle coming from the pp collision; the Particle Flow algorithm and the particle reconstruction are described in Chapter 3. As a result, the reconstructed physics objects are stored in a RECO data format; nevertheless, the RECO data is still not appropriate for physics analysis due to its large storage size on disk. Hence, two skimming processes are performed on the RECO data, resulting in a data file known as mini-AOD. The mini-AOD contains only the relevant information of the physics objects, making them appropriate for physics analysis.
3 Physics Object Reconstruction

The CMS event reconstruction is performed combining the information provided by each sub-detector described in Chapter 2. This combination allows to identify almost all the final states produced in the hard interaction, providing the measurements of their kinematic variables. In the CMS experiment, the global event reconstruction is usually performed with the Particle Flow (PF) algorithm, which is suitable for the detector design. This algorithm identifies the final states individually and classifies them into muons, electrons, photons, and hadrons. Higher level objects, such as jets, missing transverse energy and taus are reconstructed using the information of all the particles identified in the event. The PF algorithm shows a better performance for the physics objects reconstruction compared to alternative algorithms, especially in the case of jet reconstruction. The PF jet reconstruction is an important key when searching for Z' bosons in the dihadronic tau channel since jets are used as input for the tau-lepton reconstruction algorithm. Besides, b-jets, electrons, and muons reconstruction is also important for the search since they can fake the tau signature. Additionally, the missing transverse energy reconstruction is also important for this search since the taus decay into tau neutrinos.

This Chapter is organized as follows: first, a description of the PF algorithm is presented in section 3.1; the muon and electron reconstruction are described in sections 3.2 and 3.3; then, the jet and b-jet reconstruction are presented in sections 3.4 and 3.5; the missing transverse energy reconstruction is presented in section 3.6; finally, the tau reconstruction and identification are presented in detail in section 3.7.

3.1 Particle Flow

The CMS Particle Flow algorithm (PF) reconstructs and identifies individually all the stable visible particles that are produced in the hard interaction and that are within the acceptance of the detector, by combining the information collected by the sub-detectors. The PF technique performs the global event reconstruction classifying all the visible particles into five mutually exclusive groups: muons, electrons, charged hadrons, photons and neutral hadrons. These individual particles, called “PF Candidates”, are used as an input in further algorithms to reconstruct higher level objects such as jets, missing transverse energy (MET) and tau-leptons.

The capabilities of the CMS detector are ideal for using the PF technique for global event reconstruction. The high granularity of the inner tracker and of the ECAL, the hermeticity of the HCAL, the high performance of the muon system, and the strong magnetic field allow the PF algorithm to identify individual particles, reaching a high performance reconstruction even for charged particles with very low momentum (of the order of 100 MeV). The main advantage of using the PF technique, instead of traditional event reconstruction methods, is the excellent performance of the jet reconstruction. Alternative methods perform the jet reconstruction using only the energy deposits on the calorimeter system, while the PF technique uses the combined information coming from all sub-detectors in order to reconstruct the particles individually, and to cluster them into jets. This more completed information results in an improved performance of the jet reconstruction. Additionally, the individual particle reconstruction provides a de-
tailed information about the compositeness of the jets, allowing to determinate their hadron profile and their origin. This is important for the tau identification since the knowledge of the jet-compositeness allows to distinguish between QCD-jets and tau-jets. In summary, the PF technique provides a high reconstruction performance for all physics objects, in particular jets, MET and tau-leptons [60].

The PF technique involves three steps:

1. The algorithm builds the so-called “PF elements”, which consist on the tracks reconstructed in the Inner Tracker, the energy clusters reconstructed in the calorimeters and the tracks reconstructed in the muon system.

2. The algorithm performs a topological association of the basic PF elements using the so-called “link-algorithm”.

3. Individual particles are identified and reconstructed from the content of the linked elements.

Muons are reconstructed connecting together the tracks provided by the inner tracker and the muon system, as will be described in section 3.2; the electron reconstruction is performed combining the track information and the energy deposits in the ECAL (see Section 3.3); charged hadrons are identified linking their tracks with their energy deposits on the Calorimeter System; finally, the particles, like the photons and the neutral hadrons, are reconstructed from the information provided by the calorimeter system.

3.1.1 Track and Vertex Reconstruction

The track reconstruction is performed using the information provided by the Tracker system. The high granularity of this system allows for an accurate reconstruction of the charged-particle tracks, which is crucial for the momentum resolution and for the identification of the primary and secondary vertices. Additionally, the track-reconstruction resolution is also important in order to distinguish between primary and secondary vertices and to discard any vertex produced from PU processes.

**Track Reconstruction**

The track reconstruction is performed in several steps by a process called *iterative tracking*, which looks iteratively for hits (signals in each layer of the detector) associated to a track. In the firsts iterations, the algorithm takes the list of all the hits in the Tracker for the event and applies tight criteria in order to search for easy-to-identify tracks (tracks with relative high $p_T$ produced near the interaction region). The hits associated to these tracks are removed from the list with the purpose of reducing the combinatorial complexity in the subsequent iterations. Then, the selection criteria are loosened in order to identify low-$p_T$ tracks. In the first three iterations the algorithm reaches an efficiency up to 99.5% for isolated muons and larger than 90% for charged hadrons within jets [60]. In the last iterations, constraints on the origin of the vertex are relaxed in order to reconstruct the tracks originated outside the beam spot, for instance, secondary charged particles produced from photon conversions in the tracker material, and to reconstruct the remaining tracks associated to a small number of hits. Each iteration can be summarized in four steps:

- Tracks are seeded using 2 or 3 hits, giving rise to the initial track candidates and their initial trajectory parameters.
• The track parameters are recalculated using the Kalman filter algorithm [61] to account for the energy loss and the multiple Coulomb scattering produced by the interaction of the charged particle and the tracker material. The Kalman filter performs an extrapolation outwards of the inner tracker layers with the purpose to find additional hits associated to the track and to estimate the track parameters with more precision.

• An accurate information of the trajectory is obtained applying a fit on the track.

• Tracks are selected based on the quality flags, using criteria such as the $\chi^2$ of the fit, the number of hits associated to the track and the distance between the origin of the track and the primary vertex.

The estimation of the track reconstruction efficiency is performed comparing the reconstructed tracks with those generated from simulated samples; the simulated events contain only single muons or pions: muons are ideal for this purpose since, unlike electrons, they have a negligible energy loss by Bremsstrahlung radiation. On the other hand, pions not only lose energy due to Coulomb scattering but also due to nuclear interactions with the tracker material. The last kind of interactions are not taken into account in the track finder algorithm, reducing the track reconstruction efficiency in the angular region covered by the tracker. The tracking efficiency is higher than 99% for isolated muons with a $p_T > 1$ GeV while, in the case of charged pions, it is close to 95% (see Figure 3.1). The pileup interactions degrade significantly the efficiency for tracks with $p_T < 1$ GeV [62].

### Vertex Reconstruction

Vertex reconstruction is performed with the purpose to measure the position and the uncertainty of all the vertices produced in the pp collision. It is performed in three stages: first, the algorithm selects the tracks that might come from the interaction region, according to their impact parameter, their number of the reconstructed hits associated to the tracks (at least 2 hits in the pixel detector and at least 5 hits in the whole Tracker System) and the $\chi^2$ of their trajectory ($\leq 20$). In the second step, the tracks that might be produced from the same vertex are grouped into clusters; the track clustering is performed using the deterministic annealing (DA) algorithm [63], which identifies all the interaction points in a pp collision (even the inelastic collisions). Finally, the clusters of tracks associated to the same interaction point are fitted using an adaptive vertex fitter technique [64] with the purpose of estimating the vertex parameters, in particular its position. The primary vertex is identified as the vertex with the higher sum of $p_T^2$ of the clusters of tracks.

The resolution of the primary vertex depends on the number of tracks associated to it. Figure 3.2 shows the resolution of the x and z positions for the primary vertex reconstruction, using minimum-bias sample† and jet-enriched data samples [62]. The resolution of the x position is less than 20 $\mu$m and of the z position is less than 25 $\mu$m, for the minimum-bias samples, while for jet-enriched samples are less than 10 $\mu$m and less than 12 $\mu$m, respectively. Figure 3.2 (c) shows the efficiency of primary-vertex reconstruction as a function of the number of tracks.

### 3.1.2 Clustering

Other important PF elements, besides of the reconstructed tracks, are the calorimetric clusters. This clustering is used for several purposes: to identify the energy deposits of photons and

---

†Minimum-bias events pass a suit of triggers and minimum requirements on hit or track multiplicity.
hadrons, allowing to discriminate neutral hadrons from charged hadrons; to provide additional information in order to determine the track parameters of the charged hadrons that were not measured accurately in the track reconstruction, for example, charged hadrons that pass outside of the tracker acceptance, or charged hadrons with high $p_T$; and, to reconstruct electrons along with their associated Bremsstrahlung radiation.

![Figure 3.1: Efficiency of reconstructed tracks as a function of the $p_T$ (left) and the $\eta$ (right) for muons (top) and charged pions (bottom). The efficiencies were obtained using data from pp collisions with a centre-of-mass energy of 7 TeV, which corresponds to the 2011 data taking period. For simulated data an average of 8 pileup events was used, which is roughly the amount delivered by the LHC in 2011. Figure taken from [62].](image-url)
The clustering is performed for each part of the CMS calorimeter system: ECAL, HCAL, HF and PS. The clustering algorithm proceeds from the “cluster seeds” identified at local level. These seeds are selected from individual calorimeter cells whose energy deposits surpass a well defined threshold. Once a cluster seed is selected, the algorithm searches for energy deposits in the boundary cells. The algorithm adds to the cluster surrounding cells that have a maximum energy higher than two standard deviations from the electronic noise: 80 MeV for the ECAL Barrel, 300 MeV for the ECAL Endcaps and 800 MeV for the HCAL [60]. These combined cells are known as “topological clusters”. The PF clusters are constructed from the topological clusters, avoiding any overlapping between them.

![Resolution of the x position](image1)

(a) Resolution of the x position

![Resolution of the z position](image2)

(b) Resolution of the z position

![Reconstruction efficiency](image3)

(c) Reconstruction efficiency

Figure 3.2: Resolution of the x and z positions for primary vertex reconstruction, using a) minimum-bias samples and b) jet-enriched samples. The resolution is improved using jet-enriched samples, since in these sample the $p_T$ mean is larger than minimum-bias samples. For the primary vertex reconstruction efficiency shown in c), minimum-bias MC samples were used. These results were estimated with pp collisions with a centre-of-mass energy of 7 TeV. Figures were taken from [62].
3.1.3 Link algorithm

As mentioned earlier, a particle is expected to produce signatures in different sub-detectors, giving rise to one or more PF elements: Tracker tracks, clusters and Muon tracks; for example, a charged hadron, as the pion, would produce a track in the inner tracker along with energy deposits in both calorimeters. PF uses the link algorithm to perform a topological combination of the PF elements coming from the different sub-detectors with the aim of fully reconstructing each particle in the event.

The link between a charged-particle track and a calorimeter cluster is performed as follows:

- The track is extrapolated from the last hit reconstructed in the tracker to the two detection layers in the case of PS, or the expected depth of a typical electron shower in the case of the ECAL, or one interaction length for a typical hadron shower in the case of the HCAL.
- The track is linked to a calorimeter cluster if its extrapolated position is within the cluster boundaries.
- The distance between the extrapolated track position and the cluster position, in the \( \eta - \phi \) plane, defines the quality of the link.

Similarly, the link of two calorimeter clusters (i.e., links between PS and ECAL clusters or links between ECAL and HCAL clusters) is performed when the extrapolated position from the more granular calorimeter (PS or ECAL) is within the boundaries of the less granular calorimeter (ECAL or HCAL). For example, the link between the ECAL and HCAL clusters produced by one pion is established when the extrapolated position from the ECAL is within the HCAL cluster boundary. Finally, the link between a charged-particle track reconstructed in the Tracker System and a track reconstructed in the Muon System is established by a global fit between the two tracks; its \( \chi^2 \) defines the quality of the link. This link is known as global muon (see Section 3.2).

3.1.4 PF Candidates

Once the links are performed, the physics objects are identified by the Particle Flow algorithm according to their signatures in each sub-detector. The following list shows the physics objects reconstructed individually:

- **Muons** are reconstructed from the link between their signatures in the Tracker System and the Muon Chambers. A PF muon is selected if the momentum measured by the muon chambers is consistent with that measured by the Tracker System within three standard deviations.

- **Electrons** are reconstructed from the link between a track and an ECAL cluster. The identification of electrons relies on their short tracks and their energy losses due to Bremsstrahlung. Their reconstruction is a complex task since the emitted Bremsstrahlung converts into electron-positron pairs, whose tracks are bended due to the strong magnetic field; this leads to energy deposits enlarged in the \( \phi \)–direction. Therefore, the PF electrons are built from short tracks consistent with clusters enhanced in the \( \phi \)–direction.

- **Charged hadrons** are reconstructed using a link between a Tracker track and clusters in the ECAL and the HCAL, for each hadron. However, there are two scenarios in which the momentum of the track is not consistent with the energy measured in the calorimeters: First, if the energy measured is less than the momentum of the track, cleaning procedures
are performed in order to discard spurious or mis-reconstructed tracks, and then if there is consistency, a PF charged hadron is built; in cases when there is still a difference between the energy measured and the momentum, a relaxed muon search is performed. Second, if the energy measured is greater than the momentum of the track, the energy excess is attributed to the presence of neutral particles: photons or neutral hadrons.

- **Neutral particles (photons and neutral hadrons),** as mentioned above, are reconstructed from the energy excess measured by the calorimeters compared with the momentum measured by the Tracker. Specifically, if this excess surpasses the total energy deposited in the ECAL, a PF photon is built with the ECAL energy, while a PF neutral hadron is built using the remaining energy excess, otherwise only one PF photon is reconstructed. Additionally, PF photons and PF neutral hadrons can be reconstructed from isolated ECAL clusters and isolated HCAL clusters, respectively.

Once all the visible particles are reconstructed individually, specific algorithms are used to reconstruct higher level objects such as jets, missing transverse energy and tau-leptons. The reconstruction of these objects will be described in the following sections, making emphasis in the case of the taus since they are crucial for this thesis. Besides, the jet, muon and electron reconstruction will also be described in detail since these physics objects can fake the tau signature.

### 3.2 Muon Reconstruction and Identification

Muons are minimum ionizing particles with a long lifetime ($c\tau = 659$ m); for this reason, they are capable to cross through the whole CMS detector without significant energy deposits and before their decay. In consequence, the muon identification is performed using the tracking information provided by the Tracker System (*tracker tracks*) and by the Muon Chambers (*standalone-muon tracks*). The tracker tracks were described in section 3.1.1; the standalone-muon tracks are reconstructed through a linear fit of the hits detected in the DT and CSC sub-systems, using the Kalman fitting technique [61]. Based on these two PF objects, the muon reconstruction [65] is performed with three different approaches:

- **Global muon reconstruction (outside-in).** Once a standalone-muon track or hit is identified, an extrapolation from the innermost muon stations to the outer tracker surface is performed. If a tracker track is found and its parameters (as its momentum) are consistent with those of the standalone-muon track, both trajectories are associated to the same muon. Then, all the hits coming from the Tracker System and the Muon Chambers are combined and fitted using the Kalman filter technique [61], which results in a reconstruction known as *Global Muon*. This is particularly important for high-$p_T$ muons ($p_T > 200$ GeV), since the global fit can improve the momentum resolution compared with the one measured only with the Tracker System information.

- **Tracker muon reconstruction (inside-out).** In this approach, the muon reconstruction starts from the tracker tracks, with $p_T > 0.5$ GeV, and with a total momentum higher than 2.5 GeV; these tracks are considered as possible muon candidates and, consequently, they are extrapolated towards the Muon Chambers. In order to match a tracker track with a standalone-muon track, additional information is taken into account, such as the magnetic field, the energy losses and the multiple Coulomb scattering in the detector material [65]. If at least one segment reconstructed in the Muon Chambers is associated with the tracker

*A segment refers to the fit among the hits reconstructed in only one Muon Chamber (DT or CSC).*
track, the match is classified as Tracker Muon. The tracker muon reconstruction is especially important for low-\(p_T\) muons (\(p_T > 5\) GeV and \(p_T < 200\) GeV) due to its high efficiency (since it requires only one muon-segment), and to its high energy resolution in this kinematic range.

- **Standalone muon reconstruction.** When no tracker track is matched with the standalone muon track, the reconstruction results in a Standalone Muon. They are not commonly used in CMS analyses since they have the worst momentum resolution compared with the Global Muons and the Tracker Muons. Additionally, their signature can come from cosmic-ray muons.

  The high efficiency of the tracker tracks and the standalone tracks allows to reconstruct 99% of the muons produced in the pp collisions within the geometrical acceptance \([65]\). A muon can be built as a Global Muon or as a Tracker Muon using either of both approaches; besides, in the case that one muon is reconstructed using both of them, the information is merged into one single muon candidate. Once the muon candidate is selected, several algorithms are applied in order to identify it from background processes that can fake the muon signature, such as charged hadrons. Charged hadrons can cross the whole Calorimeter System and can reach the Muon Chambers, producing similar signatures than those of a muon, with the exception of the energy deposits in the calorimeters. Therefore, an energy-loss requirement allows to discriminate muons from charged hadrons. There are three basic muon identification algorithms: the soft muon selection, the tight muon selection, and the global selection (a more detailed description of each algorithm can be found in reference \([65]\)). The PF algorithm uses these identification requirements and, along with the energy deposits in the Calorimeter System, it selects the so-called PF Muon Candidates.

  The PF Muons are identified using three different selections known as “isolation”, “pf-tight” and “pf-loose” \([66]\). The isolation selection looks for muons with low neighboring activity, defining a cone of \(\Delta R = 0.4\) around the muon trajectory. If the sum of the \(p_T\) of the tracks and the transverse energy deposited in the calorimeters, within the cone, is less than 15% of the muon \(p_T\), an “isolated” muon is selected. The main goal of pf-tight and pf-loose selections is the identification of muons within jets. The pf-tight selection requires a minimum amount of hits in the Muon Chambers and searches for compatibility between the momentum measured and the energy deposited in the calorimeters. When the momentum is significantly higher than the energy deposits, which is compatible with the muon signature, but not with the charged hadron signature, the pf-loose selection is applied. In consequence, the last selection discriminates between muons and charged hadrons in a jet, relaxing the number of hits required and looking for matches between the tracker track and the standalone-muon track.

### 3.3 Electron Reconstruction and Identification

The electron reconstruction relies on the association of an ECAL cluster and a track reconstructed in the Tracker System. This is challenging since electrons lose energy through Bremsstrahlung radiation while crossing the Tracker material, therefore, their energy deposits are spread out in several ECAL crystals, mostly in the \(\phi\)-direction, due to the bending of the electron tracks caused by the magnetic field (see Figure 3.3). Because of its interactions with the Tracker material, an electron can lose of the order of 33% of its energy crossing the central region of the Tracker System (\(|\eta| \sim 0\)) where the material budget is minimum, whereas 86% of its energy is lost by photon radiation if it passes through the Tracker at \(|\eta| \sim 1.4\) \([67]\). In consequence, the collection
of the energy coming from Bremsstrahlung radiation is crucial for an accurate measurement of the initial energy of the electron.

![Figure 3.3: Schematic view of energy deposited by an electron in the ECAL.](image)

There are two algorithms to measure the initial energy of the electron using the information provided by the ECAL: the “hybrid” algorithm for the barrel and the “multi-5×5” algorithm for the endcaps. The hybrid algorithm searches for several clusters consistent with energy deposits enlarged in the $\phi$-direction, as was mentioned above. It starts selecting the seed crystal\textsuperscript{†} and then, it looks progressively for crystals, in both $\phi$ directions, that exceed a threshold energy. These crystals are grouped, resulting in the so-called supercluster (SC). In the case of the “multi-5×5” algorithm, since the endcap crystals are not arranged in a $\eta-\phi$ geometry, a crystal is selected based on the high energy deposit and an arrange of $5\times5$ crystals centered on it are selected. The cluster reconstruction is complemented with the PF algorithm, adding the contiguous crystals to the seed, whose energy collected exceeds two standard deviations above the electronic noise (230 MeV in the barrel and 600 MeV in the endcaps).

The electron track reconstruction is also challenging due to the significant energy losses by Bremsstrahlung in each Tracker layer, which can cause a considerable variation in the curvature of its trajectory; this results in a degradation of the performance of the track reconstruction. For this reason, the Kalman Fitter (KF) algorithm used for all charged particles is not appropriated to reconstruct the electron track and instead a Gaussian Sum Filter (GSF)\textsuperscript{[68]} algorithm has been developed with this purpose. Since the pattern recognition procedure can be very time consuming, the algorithm starts preselecting tracks seeds. There are two algorithms used for the electron track reconstruction: the first one is called seeding and the other one is called tracking. The seeding algorithm performs an extrapolation from the SC to the Tracker System in order to select hits compatible with either, positive or negative charge hypotheses; then, the selected hits are used as input for a preliminary track reconstruction, which is performed with the general algorithm for charged particles, extrapolating the track towards the ECAL and matching it to a SC. The tracking algorithm looks for additional hits consistent with the electron track, discarding any ambiguity; it uses the Kalman Fitter method in each successive layer, including the energy loss by the electron. Once the hits collection is set, without ambiguities, a GSF fit is performed to estimate the track parameters. As a result, the algorithm provides an electron track that can be extrapolated to the ECAL, whose energy loss by Bremsstrahlung can be es-

\textsuperscript{†}The seed crystal has the higher collected energy in a defined region.
estimated as: \( f_{\text{br}} = (p_{\text{in}} - p_{\text{out}})/p_{\text{in}} \), where \( p_{\text{in}} \) is the momentum measured in the inner-most layer of the Tracker system, while \( p_{\text{out}} \) is the momentum at the ECAL surface, estimated from the extrapolation. This variable is useful for the electron identification.

Once the PF cluster candidates and the GSF tracks are selected, the algorithm builds the PF electron candidates, looking for geometrical matches between them. The matching is set when \( |\Delta \eta| = |\eta_{\text{SC}} - \eta_{\text{extra}}| < 0.02 \), where \( \eta_{\text{SC}} \) is the \( \eta \) position of the SC energy and \( \eta_{\text{extra}} \) is the \( \eta \) position of the extrapolated track from the inner track layer to the ECAL surface. Similarly, for the \( \phi \)-direction, \( |\Delta \phi| = |\phi_{\text{SC}} - \phi_{\text{extra}}| < 0.15 \) is required.

CMS uses several algorithms to identify isolated electrons (signal) and to distinguish them from processes that can fake their signature, such as photon conversions, jets misidentified as electrons, hadronic taus misidentified as electrons, or electrons produced from semileptonic decays of b and c quarks. In this analysis, a MultiVariate Analysis (MVA) algorithm is used for electron identification, exploiting the information of the track-cluster matches, the SC geometrical structure, the \( f_{\text{br}} \) fraction, among others. Different working points are defined according to the desired reconstruction efficiency for each analysis (see Table 3.1). In order to reduce the electron-to-tau misidentification rate, the 90\% efficiency working point of the MVA electron ID is selected for this analysis.

<table>
<thead>
<tr>
<th>Category</th>
<th>( \text{MV}_{\text{min}} ) cut (80% signal eff)</th>
<th>( \text{MV}_{\text{min}} ) cut (90% signal eff)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barrel (( \eta &lt; 0.8 )) ( p_{T} &gt; 10 )</td>
<td>0.941</td>
<td>0.837</td>
</tr>
<tr>
<td>Barrel (( \eta &gt; 0.8 )) ( p_{T} &gt; 10 )</td>
<td>0.899</td>
<td>0.715</td>
</tr>
<tr>
<td>Endcap ( p_{T} &gt; 10 )</td>
<td>0.758</td>
<td>0.357</td>
</tr>
</tbody>
</table>

Table 3.1: Electron ID Selections.

An additional background reduction is obtained through an isolation requirement. Jets or electrons coming from semileptonic decays of b or c quarks can fake the signature of electrons coming from the primary interaction. Since both cases have a significant energy deposits around their tracks, an isolation requirement will reduce considerably these background contaminations. The relative isolation is defined as:

\[
\text{Iso}_{\text{PF}} = \frac{\sum_{\text{charged hadron}} p_{T} + \max(0, \sum_{\text{neutral hadron}} p_{T} + \sum_{\gamma} p_{T} - 0.5 \times \sum_{\text{charged hadron from PU}} p_{T})}{p_{T_{\text{electron}}}},
\]

where the sums run over the charged hadron candidates, neutral hadrons and photons, within a \( \Delta R \) cone around the electron direction. The charged hadron candidates are required to be originated from the primary vertex \(^{67}\) and a correction due to PU contributions is included \( (p_{T_{\text{charged hadron from PU}}}) \). In this analysis, the isolation requirement used is \( \Delta R < 0.4 \).

### 3.4 Jet Reconstruction

Quarks and gluons produced by the pp collisions, at parton level, have a well defined color charge and, due to the confinement of the strong interaction, they can not exist as free particles; instead, they will follow a process of fragmentation and combination with quarks and gluons from the vacuum to form colorless bound states called hadrons; this process is known as hadronization.
The hadronization will give rise to showers of particles, which propagate in a similar direction forming a cone; this set of particles is known as a jet (see Figure 3.4). The aim of a jet reconstruction algorithm is to determine the energy and the direction of the initial parton. The PF technique reconstructs all the particles individually, from the information collected by the CMS sub-detectors, and then cluster them into the jet, determining the compositeness and, when possible, its the initial parton.

The signature of the jet in each layer is expected to be located roughly into an area of $\pi R^2$, where $R$ is used as an input parameter of the reconstruction algorithms. The parameter $R$, and consequently the jet area in a given layer, determine the sensitivity of the algorithm to discriminate between jets and soft radiation processes. If the $R$ parameter is too large, not only all possible particles coming from the hadronization are included in the reconstruction, but also those coming from the underlying event (UE) or pileup collisions (PU) might be included; this will lead to an overestimation of the jet energy. Additionally, the algorithm must consider other important circumstances that degrade the jet-reconstruction efficiency, such as the infrared and the collinear scenarios. The infrared processes refer to soft QCD radiation that might be wrongly included in the jet reconstruction, altering the jet components. The collinear processes refer to parton splitting into collinear components, leading to the reconstruction of a different number of jets. The algorithms that avoid including these processes are known as infrared- and collinear-safe algorithms (ICR).

There are two different types of algorithms that have been used for jet reconstruction: the cone algorithms (SISCone) and the sequential clustering algorithms (kT, Cambridge-Aachen, anti-kT). The main difference between them is the way how the particles which belong to a jet are clustered. The cone algorithms consider that the particles in a jet are spread out in a rigid conical region, which results in a fixed definition of the jet area, while the sequential algorithms assume fluctuations on the jet area, where the boundaries are defined with a fixed maximum radius. The main features of these algorithms are:

- **SISCone algorithm**: checks all the particles in the event in order to identify stable combinations that are coherent with a jet. This algorithm explores randomly the $\eta-\phi$ plane using a circle, with a well defined area, until a particle matches with the circumference; then, the circumference is pivoted until a second particle comes in contact with it. The matches define the stable cones. The particles associated to an stable cone are removed from the list of particles of the event and the procedure is performed again until no more
stable cones are found. Finally, the stable cones are split or merged, according to an overlap parameter, in order to reconstruct the jets [69].

- **kT-algorithm**: estimates the following distance variables for all pair of particles \((i,j)\):

\[
d_{ij} = \text{min}(p_{T_i}^2, p_{T_j}^2) \Delta R^2_{ij} / R^2,

\]
\[
d_{iB} = p_{T_i}^2,
\]

where \(\Delta R^2_{ij} = \Delta(\eta_i, \eta_j)^2 + \Delta(\phi_i, \phi_j)^2\) is the distance in the \(\eta-\phi\) plane between the particles \(i\) and \(j\), \(R\) is the radius parameter that defines the size of the jet, and \(d_{iB}\) is the distance between the beam axis and the particle \(i\). Then, the algorithm proceeds by finding the minimum distances \(d_{ij}\) and \(d_{iB}\); if \(d_{ij}\) is a minimum, the pair of particles \(i, j\) are recombined adding their four-momenta; if \(d_{iB}\) is a minimum, the algorithm identify the particle \(i\) as a jet. The distances are recalculated iteratively until all particles are part of the jet [69,70].

- **Cambridge-Aachen algorithm**: proceeds similar to the kT algorithm, but defining the variable distances:

\[
d_{ij} = \Delta R^2_{ij} / R,

\]
\[
d_{iB} = 1.
\]

Since \(d_{ij}\) is independent of the momentum, the jet size will not be clearly defined and UE and PU processes will be included in the jet reconstruction. In spite of that disadvantage, this algorithm is commonly used since it allows to determine the jet compositeness [71].

- **anti-kT algorithm**: defines the following distance variables for all particles pairs \((i,j)\):

\[
d_{ij} = \text{min} \left( \frac{1}{p_{T_i}^2}, \frac{1}{p_{T_j}^2} \right) \Delta R^2_{ij} / R^2,

\]
\[
d_{iB} = 1 / p_{T_i}^2.
\]

The anti-kT algorithm follows the same procedure as the kT algorithm. It calculates iteratively the variables \(d_{ij}\) and \(d_{iB}\), and searches for the minimum of the set \(\{d_{ij}, d_{iB}\}\). If \(d_{ij}\) is a minimum, the particles \(i\) and \(j\) are merged by the algorithm and they are treated as one particle; the iterative procedure continues until the \(d_{iB}\) minimum is found. If \(d_{iB}\) is a minimum distance, the particle \(i\) is identified as a jet, which means that all the particles merged by the algorithm associated with that particle \(i\) are the constituents of the jet. Since the anti-kT \(d_{ij}\) variable is dominated by high \(p_T\) particles, the algorithm will first merge them, resulting in an accurate jet area estimation [72].

The anti-kT algorithm is used in most of the CMS and ATLAS analyses. The advantage of this algorithm is related with the jet area since, as mentioned above, it calculates first the distance for the high \(p_T\) particles and then recalculates this distance, including low \(p_T\) particles; this results in a slight fluctuation in the jet area, which is important for removing any contributions from UE and PU processes. Figure 3.5 presents an example of the jet area profile using the algorithms described above, showing the advantage of the anti-kT algorithm.
Figure 3.5: Jets reconstruction using the algorithms: SISCone (left-top), kT (right-top), Cambridge-Aachen (left-bottom) and anti-kT (right-bottom). The jet reconstruction was performed with simulated samples using the same $R$ for all the algorithms. The SISCone algorithm reconstructs jets in smaller well defined area but it identifies two different jets instead of the one (see jet in color gray); the kT algorithm results an irregular jet area since it starts the reconstruction using the low $p_T$ particles; the Cambridge-Aachen algorithm also shows an irregular jet area, since the distances between particles are independent of their momenta; and the anti-kT algorithm shows a well defined area, reconstructing the proper number of jets [71].

The CMS jet reconstruction is performed with the anti-kT algorithm, setting the distance parameter $R$ in the $\eta - \phi$ plane to 0.4. The four-momentum of the reconstructed jet corresponds to the sum of the four-momenta of all the PF objects associated to it. However, due to detector response and experimental effects, the reconstructed PF jet four-momentum does not correspond to the one at parton or hadron level, in consequence, corrections must be applied to the jet energy. These corrections are known as Jet Energy Corrections (JEC) [73,74]. Different levels of JEC (see Figure 3.6) are applied in a fixed sequence:

- **Pileup correction**: Also referred as “L1 correction”. This correction is applied to discard any contribution to the jet reconstruction from PU processes. It corrects for the additional tracks and the excess of energy deposits in the calorimeters due to pileup events. This
contribution is estimated using simulated dijet events with and without PU, and it is parameterized as a function of the offset energy density $\rho$, the jet area, the jet $\eta$-direction and the jet $p_T$.

- **True response**: The second level of JEC (known as L2L3 MC-truth correction) is related to the detector response and to its effect over the hadron distributions. It corrects for their non-uniformity in the $\eta$ direction and their non-linearity in $p_T$, by comparing with distributions obtained with simulated QCD-multijet events.

- **Residual corrections**: L2L3 residual corrections are applied to address the remaining differences between the jet response on data and simulated samples (of the order of 1%). These corrections are achieved with data-driven methods, using dijet samples for the $\eta$-dependent part corrections, and $\gamma/Z$+jets samples for the $p_T$ part.

- **Flavor correction**: This is an optional correction, useful for some analyses focused in identifying, at some level, the initial parton of the jet. Since the true response correction is performed using QCD simulated samples with flavor mixture at parton level, this correction assumes a specific flavor hypothesis for the initial parton of the jet and considers the jet response for different initial parton flavors. For example, jets coming from light quarks have higher momentum particles than the ones coming from gluons.

![Figure 3.6: Levels of corrections for PF jet four-momentum. Figure taken from 74](image)

The jet reconstruction is important for this analysis, since jets are taken as an input in the algorithm to reconstruct the hadronic tau decay. In this work, the jet reconstruction is performed with the anti-kT algorithm, using the distance parameter $\Delta R(\phi, \eta) = 0.4$. Jets are required to have $p_T > 30$ GeV and $|\eta| < 2.4$. For the jet identification, the loose ID working point is used, which has a reconstruction efficiency in simulation greater than 98%. The jet energy corrections recommended by the CMS collaboration were applied.

### 3.5 b-jet Identification

The identification of jets originated from bottom quark decays, or *b-tagging*, exploits the properties of the $B$-hadrons, such as their large decay life-times ($c \tau \approx 450 \mu$m), which leads to displaced tracks as well as the presence of leptons in the final state coming from semileptonic decays. CMS has developed several algorithms for b-jet identification based on features such as the large impact parameter of the b-jet candidate, with respect to the primary vertex; a possible secondary vertex reconstructed inside the jet; the mass and the number of tracks associated to the secondary vertex; the number of tracks in the jet and the possible presence of soft leptons.

In this analysis we use the Combined Secondary Vertex algorithm (CSV), which shows the best
performance for b-jet identification \cite{78}. The CSV algorithm is MVA-based, combining the in-
formation of displaced tracks, secondary vertices and the jet kinematics. The method provides a
single discriminator to evaluate the compatibility of a given jet with a b-jet. The CSV discrimi-
nator defines three possible working points: the CSVL, or loose working point, allows to select a
high b-tagging efficiency; the CSVT, or tight working point, allows to select a high purity b-jet
sample; and the CSVM, or middle working point, shows a high efficiency preserving high purity.
The three working points have been defined according to their b-tagging misidentification rate
(probability of a non b-jet being tagged as a b-jet) and their b-tagging efficiency (probability
of a real b-jet being tagged by the algorithm). Table \ref{table:3.2} shows the b-tagging efficiency for the
three working points.

<table>
<thead>
<tr>
<th>Working Point</th>
<th>CSVv2 discriminator</th>
<th>b-tagging efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loose</td>
<td>≥ 0.5426</td>
<td>≈ 83 %</td>
</tr>
<tr>
<td>Medium</td>
<td>≥ 0.8484</td>
<td>≈ 69 %</td>
</tr>
<tr>
<td>Tight</td>
<td>≥ 0.9535</td>
<td>≈ 49 %</td>
</tr>
</tbody>
</table>

Table 3.2: CSVv2 discriminator threshold and corresponding efficiency for the three working
points. The efficiencies have been determined using b-jets with transverse momentum
above 30 GeV, with simulated $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13 \text{ TeV}$. The numbers
are presented just for illustration since they also depend on the $p_T$ and $\eta$ distributions
of the jets \cite{79}.

The b-tagging algorithm has been improved for Run II (CSVv2) with the main aim of reducing
the computing time. CSVv2 includes an updated multivariate algorithm that estimates the CSV
discriminator using a neural network method instead of a simple likelihood rate method; besides,
it also uses a new algorithm for the secondary vertex reconstruction (Inclusive Vertex Finder,
IVF) as well as an improved track selection. The improvement of the CSVv2 algorithm lead to
an increase of about 10\% in the b-jet identification efficiency compared to the CSV algorithm
used for Run I (see Figure 3.7) \cite{80}. The loose working point, which provides an efficiency of
83\%, is used in this analysis.

### 3.6 Missing Transverse Energy

Almost all the final states coming from a pp collision can be detected and identified by the
CMS experiment with the exception of neutral particles that only interact weakly with matter,
such as neutrinos or hypothetical neutralinos. Although these particles scape from CMS without
detection, it is possible to infer the transverse momenta carried by all of them using the transverse
momentum conservation. The transverse momentum carried out by all the weakly interacting
particles, known as the missing transverse momentum, and denoted by $\vec{E}_T$, is determined by
the total momentum imbalance of the event in the orthogonal plane to the beam line. There
are several algorithms for the $\vec{E}_T$ reconstruction in CMS \cite{81}; the most common are: the PF
$\vec{E}_T$ algorithm, which is based on the PF technique and the Calo-$\vec{E}_T$ algorithm which uses only
the energy deposited in the calorimeter towers. The PF $\vec{E}_T$ reconstruction is widely used in the
CMS analyses, and it is defined as the negative of the vectorial sum of transverse momenta of all PF particles. Its magnitude is known as missing transverse energy (MET or $\vec{E}_T$):

$$\vec{E}_T = - \sum_{i \in \text{tracks}} \vec{p}_{T,i}, \quad \vec{E}_T = \left| \vec{E}_T \right|. \quad (3.5)$$
Figure 3.7: Efficiency of non-b jets to be misidentified as b-jet, as a function of the b-jet efficiency for several jet-identification algorithms. An enriched $t\bar{t}$ sample with 2.6 fb$^{-1}$ from the data collected in 2015 at $\sqrt{s} = 13$ TeV and a BX of 25 ns. Although cMVAv2 algorithm shows the best performance, the CSVv2 algorithm is used for many analysis and shows an improvement compared with the CSV algorithm used in Run I. Figure taken from [79].

Similar to the case of the jet reconstruction, the missing transverse momentum can be overestimated due to experimental effects such as the pileup and the non-linearity and non-uniformity of the detector response to hadrons (see Section 3.4). With the purpose to address this overestimation, three kinds of corrections can be applied to the $\vec{E}_T$ measurement:

- **Type-0 Correction**: the purpose of this correction is to remove any contribution from PU interactions to $\vec{E}_T$. Although, PU processes produce few invisible particles (mainly neutrinos coming from kaon decays), they degrade the $\vec{E}_T$ estimation due to the incapability to identify all the tracks coming from PU interaction. In consequence, this correction removes all charged hadrons that might come from pile-up interactions from the $\vec{E}_T$ estimation.

- **Type-I Correction**: this correction includes the JEC for the jets, in the $\vec{E}_T$ reconstruction. All the particles that can be clustered into jets are identified and their momenta are replaced with the sum of the transverse momenta of the jets that include JEC [82]. The correction is given by:

$\vec{E}_{T^{corr}} = \vec{E}_T - \sum_{jets} (p_T^{JEC} - p_T^{jets}) \tag{3.6}$

- **xy-Shift Correction**: the $E_T$ measurement should be independent of the $\phi$-direction, due to the cylindrical symmetry of the CMS detector. However, a dependency on $\phi$ is observed, which might come from anisotropic detector responses, inactive calorimeter cells or detector misalignments. Besides, the $\phi$ dependency on $E_T$ is also related with the pileup contributions and can be mitigated by shifting the origin of the transverse momentum plane.
The “type-I” correction to $\vec{E}_T$ is widely preferred by CMS analyses. For Run II, the “type-I” correction requires jets with $p_T$ greater than 15 GeV (including JEC) whose energy fraction deposited in the ECAL is less than 0.9 \[82\].

The performance of $E_T$ reconstruction is estimated from Z bosons in dilepton events. The $E_T$ resolution is dominated by the hadron activity in the event since the leptons have high momentum resolution, which varies from 1-6\% for muons, 1-4\% for electrons/photons and 5-15\% for jets.

Figure 3.8 shows a very good agreement between data and simulated events for the $\vec{E}_T$ distribution in the $Z \rightarrow \mu^- \mu^+$ and $Z \rightarrow e^- e^+$ channels \[82\].

In this analysis, the “type-I” correction to $\vec{E}_T$ is applied. In order to reject any anomalous event with high-$\vec{E}_T$, the filters recommended by the CMS Collaboration were applied \[83\].

Figure 3.8: $E_T$ distribution for $Z \rightarrow \mu^- \mu^+$ (left) and $Z \rightarrow e^- e^+$ (right). Data comes from pp collisions at $\sqrt{s} = 13$ TeV collected by CMS detector during 2016 (integrated luminosity up to 12.9 fb$^{-1}$). Figure taken from \[82\].

### 3.7 Tau Reconstruction and Identification

As was mentioned in Section 1.4, the leptonic decays of the tau cannot be distinguished from the leptons originated in the hard interaction and, therefore, only the hadronic decays can be identified directly. The hadronic tau reconstruction is challenging, from the experimental point of view, since its signature is pretty similar to the one of a QCD-jet, which is produced with a cross section eight orders of magnitude larger. Additionally, the tau kinematic variables cannot be fully reconstructed due to the presence of neutrinos in its decay modes. In most of the CMS analyses, the tau reconstruction and identification is performed using the Hadron-Plus-Strips algorithm (HPS) \[84\]. This algorithm has two stages:

- **Reconstruction**: The algorithm searches for charged and neutral PF objects that are compatible with the tau-decay modes, in order to reconstruct a tau and to compute its kinematic variables.
Identification: Discriminators are applied on the reconstructed tau in order to distinguish it from background processes, such as QCD-jets, electrons and muons, reducing the misidentification rates.

3.7.1 Tau Reconstruction Algorithm

The HPS algorithm takes as input jets reconstructed with the anti-kT algorithm (described in section 3.4) and reconstructs individually the hadronic tau decay modes, using the charged and neutral constituents of the jet built by the PF algorithm. As can be seen in Table 3.3, the hadronic decays are composed by one neutrino, neutral pions and charged hadrons (mostly pions). In consequence, the HPS algorithm looks for charged hadrons using the information provided by the Tracker and Calorimeter Systems, and looks for energy deposits in the calorimeters consistent with the signature of neutral pions. As was mentioned in section 2.4, there is a high probability that a photon, coming from a \( \pi^0 \rightarrow \gamma \gamma \) decay, can convert into an electron-positron pair due to its interaction with the Tracker material; then, the electron and the positron are spread apart in opposite directions due to the bending of their tracks by the magnetic field, resulting in energy deposits enlarged in the \( \phi \)-direction (see Figure 3.9). Therefore, the HPS algorithm looks for ECAL strips in the \( \eta - \phi \) plane consistent with the \( \pi^0 \) signature. These strips are combined with the charged hadrons with the aim to reconstruct the possible tau decay modes.

<table>
<thead>
<tr>
<th>Final State</th>
<th>Branching Fraction [%]</th>
<th>Resonance</th>
<th>Mass [GeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( e^- \nu e \nu )</td>
<td>17.8</td>
<td>( \pi^0 )</td>
<td></td>
</tr>
<tr>
<td>( \mu^- \nu \mu \nu )</td>
<td>17.4</td>
<td>( \rho )</td>
<td>770</td>
</tr>
<tr>
<td>( h^- \nu )</td>
<td>11.5</td>
<td>( a_1 )</td>
<td>1200</td>
</tr>
<tr>
<td>( h^- \pi^0 \nu )</td>
<td>25.9</td>
<td>( a_1 )</td>
<td>1200</td>
</tr>
<tr>
<td>( h^- h^+ \pi^0 \nu )</td>
<td>9.8</td>
<td>( a_1 )</td>
<td>1200</td>
</tr>
<tr>
<td>( h^- h^+ h^- \pi^0 \nu )</td>
<td>4.8</td>
<td>Others</td>
<td>3.3</td>
</tr>
</tbody>
</table>

Table 3.3: Branching fraction of \( \tau^- \) modes [45].

Strip reconstruction.

The electrons and photons constituents of the jet, which is used as an input for the \( \tau_h \) reconstruction, are clustered into \( \Delta \eta \times \Delta \phi \) strips. The clustering starts with an iterative procedure in order to select the highest-p\( _T \) electron, or photon, that has not been included yet in any strip; the \( e/\gamma \) selected seeds the clustering algorithm. The position of the new strip corresponds to the \( \eta \) and the \( \phi \) positions of the \( e/\gamma \) seed. Then, the next-larger p\( _T \) electron, or photon, is added to the strip, if it is within a \( \Delta \eta \times \Delta \phi \) window. The iterative procedure ends when no electrons, nor photons, are found within the size of the window. In the previous versions of the HPS algorithm, the size of the \( \Delta \eta \times \Delta \phi \) window was fixed to \( 0.05 \times 0.20 \) in the \( \eta - \phi \) plane [57]. In some cases, all electrons and photons originated from the tau decay products are not included in the fixed size window, reducing the isolation efficiency. For instance, a charged pion, coming from a \( \tau_h \) decay, can produce secondary low-p\( _T \) particles due to its nuclear interaction with the tracker material, resulting in low-p\( _T \) electrons and photons that can lie outside of the window. Therefore, it might be
convenient to increase the window size, but it would represent a decrease in the isolation efficiency for high-p$_T$ taus, since their decay products tend to be boosted and localized into a smaller strip size. In consequence, a dynamic-strip reconstruction algorithm has been developed based on these considerations (see Figure 3.10).

The dynamic-strip reconstruction selects iteratively the next-larger p$_T$ electron, or photon, and adds it to the strip, if it is within a range of:

$$\Delta \eta = 0.20 \cdot \frac{(p_{T,e/\gamma})^{-0.66}}{\text{GeV}} + 0.20 \cdot \frac{(p_{T,\text{strip}})^{-0.66}}{\text{GeV}},$$

$$\Delta \phi = 0.35 \cdot \frac{(p_{T,e/\gamma})^{-0.71}}{\text{GeV}} + 0.35 \cdot \frac{(p_{T,\text{strip}})^{-0.71}}{\text{GeV}},$$

(3.7)

where p$_T$,$e/\gamma$ is the e/\gamma-momentum and p$_{T,\text{strip}}$ is the momentum associated to the strip. The window size is constrained up to 0.15 in $\Delta \eta$ and up to 0.3 in $\Delta \phi$. As can be inferred from the equations (3.7), the window size depends on the momentum of both, the strip and the added electrons or photons. This results in an improvement in the isolation efficiency compared to the fixed window size used in Run I since it reduces the QCD contamination. Then, the strip position is recalculated using a p$_T$-weighted average of all electrons and photons included in the strip, as follows:

$$\eta_{\text{strip}} = \frac{1}{p_{T,\text{strip}}} \sum \left( \frac{p_{T,e/\gamma} \cdot \eta_{e/\gamma}}{\text{GeV}} \right),$$

$$\phi_{\text{strip}} = \frac{1}{p_{T,\text{strip}}} \sum \left( \frac{p_{T,e/\gamma} \cdot \phi_{e/\gamma}}{\text{GeV}} \right),$$

(3.8)
where $p_{T,\text{strip}} = \sum p_{T,e/\gamma}$. The strip reconstruction ends if no electrons, nor photons, are found within the window. In that case, the clustering looks for the highest-$p_T$ electron, or photon, which is not associated to any strip, and selects it in order to reconstruct a new strip.

Strips, whose $p_T$ is larger than 2.5 GeV, are kept as $\pi^0$ candidates. A $\pi^0$ candidate can be identified as a tau decay product if the strip position is within the signal cone associated to the tau candidate.

- **Charged hadron reconstruction.**

As was mentioned above, the constituents of the jet are taken as input for the HPS algorithm, which looks for charged hadrons consistent with the tau signature. The algorithm requires charged particles with $p_T$ larger than 0.5 GeV, whose transverse impact parameter, $d_{xy}$, is less than 0.1 cm. The $p_T$ requirement on the hadrons ensures the selection of high quality tracks, which have a reconstruction efficiency of $\sim 95\%$, and have passed previous requirements, such as the number of reconstructed hits on the Tracker System and the $\chi^2$ of the fit (see Section 3.1.1). The criterion on the transverse impact parameter reduces the background contamination coming from PU and spurious tracks; however, such criterion is not so restrictive in order to avoid any rejection of high-$p_T$ taus, which would have a long lifetime.

The set of charged hadrons and strips obtained from the PF jet are combined in order to reconstruct the tau decay modes individually. The visible tau decay products are mainly one or three hadrons plus neutral pions. The algorithm aim is to reconstruct all the hadronic decay modes with exception of $h^\pm h^\mp h^\pm \pi^0$, this decay mode is not considered currently by the algorithm because, besides of its relative low branching ratio (4.8%), it has a significant background contribution from QCD-jets. In summary, the reconstructed decay modes are: $h^\pm$, $h^\pm \pi^0$, $h^\pm \pi^0 \pi^0$, and $h^\pm h^\mp h^\pm$ (see Table 3.3).
There are three requirements on the hadron plus strip combination in order to reconstruct the tau candidate:

1. A mass criterion is applied since the tau candidate mass ($m_\tau$) must be consistent with the mass of either of the resonances, $\rho$ or $a_1$. The mass is optimized for each decay mode in order to increase the reconstruction efficiency; the requirements are [84]:
   - a mass window of $(0.3 \text{ GeV} - \Delta m_\tau) < m_\tau < (1.3 \text{ GeV} \cdot \sqrt{p_T(\text{GeV})} + \Delta m_\tau)$ for the $h^\pm \pi^0$ decay mode. The mass upper limit is constrained further to be between 1.3 to 4.2 GeV;
   - a mass window of $(0.4 \text{ GeV} - \Delta m_\tau) < m_\tau < (1.2 \text{ GeV} \cdot \sqrt{p_T(\text{GeV})} + \Delta m_\tau)$ for the $h^\pm \pi^0 \pi^0$ decay mode. The mass upper limit is constrained further to be between 1.2 to 4.0 GeV;
   - a mass window of $0.3 \text{ GeV} < m_\tau < 1.3 \text{ GeV} \cdot \sqrt{p_T(\text{GeV})}$ for the $h^\pm h^\mp h^\pm$ decay mode. The mass upper limit is constrained further to be between 1.3 to 4.2 GeV.

   where $\Delta m_\tau$ corresponds to the mass change due to the addition of $e/\gamma$ candidates to the strip reconstruction.

2. The electric charge of the tau candidate must be $\pm 1$, otherwise they will be rejected.

3. The tau decay products must lie within the signal cone. The signal cone is centered in the direction of the tau momentum, which initially is assumed as the direction of the hadron with highest momentum, and it has a radius of $R_{sig} = 3.0/p_T(\text{GeV})$, where $p_T$ corresponds to the sum of the charged hadrons momenta. The lower and upper limits on $R_{sig}$ are 0.05 and 0.1, respectively.

When multiple decay modes pass the tau requirements described above for the input jet, only the hypothesis with the highest-$p_T$ is selected. Then, a single tau candidate is reconstructed for each jet. The 4-momentum of the $\tau_h$ candidate is the sum of the 4-momenta of the charged hadrons and the one reconstructed from the strips.

**Decay Mode Reconstruction**

In the tau reconstruction jargon, the number of charged particles in a tau decay mode is known as prong. Hereafter, the decay modes will be referred as:

- **1 prong**: for $h^\pm$.
- **1 prong + $\pi^0$**: for $h^\pm \pi^0$ and $h^\pm \pi^0 \pi^0$.
- **3 prongs**: for $h^\pm h^\mp h^\pm$.

In the previous version of the HPS algorithm, the decay mode reconstruction (oldDMF) only considers the final states listed above (see Figure 3.11). However, an additional 2-prong unphysical final state can be considered, which results from high-$p_T$ taus in the 3-prong decay mode; in this case, the tracks of the boosted charged hadrons are very close, and the distance between two of them can be smaller than the spatial resolution of the Tracker System. Therefore, two of the three tracks are merged and reconstructed as a single charged hadron, resulting in an apparent 2-prong final state. In the latest version of the HPS algorithm, the decay mode reconstruction (newDMF) also considers the 2-prong final state. For this analysis, the newDMF
was used. The decay modes considered in the tau reconstruction affect the sensitivity of the $Z'$ identification, since they are correlated with the QCD contamination. The decay modes used for the tau reconstruction are 1or3-prongs, instead of 1or2or3-prongs, since a better significance is obtained.

![Image](image_url)

Figure 3.11: Tau decay modes reconstructed by the HPS algorithm. Taken from [85].

### 3.7.2 Tau Identification

Once the tau candidates are reconstructed, discriminants dedicated to identify them from each background contamination are applied. As it has been mentioned before, jets, muons and electrons can fake the tau signature. In this section these discriminator algorithms are described.

**Tau Isolation Discriminator**

The purpose of the tau isolation discriminator is to reduce the background contribution from QCD-jets. Since the QCD-jet energy is carried out by charged hadrons ($\sim 65\%$), photons from $\pi^0$ decays ($\sim 15\%$) and neutral hadrons ($\sim 20\%$) [86], their signatures are similar to those produced by the hadronic tau decay modes; besides, QCD-jets have a cross section around eight orders of magnitude larger, making the discrimination against jets crucial for the $\tau_h$ identification. The isolation discriminator algorithm mainly exploits two variables that distinguish hadronic taus from QCD-jets. The first one is the distance between the decay products of a hadronic tau, since it is correlated with the tau energy; therefore, the decay products of a high-$p_T$ tau lie within a narrower cone than the constituents of a QCD-jet. The second one is the multiplicity of the particles, since the hadronic taus have few number of particles, each one with high momentum, whereas the QCD-jet has a higher multiplicity of particles with a wider energy profile. For these reasons, the so-called isolation-sum discriminator uses a cone whose size should be big enough to contain all the decay products coming from the $\tau_h$ decay and small enough to reject the QCD events. The isolation cone has a size of $\Delta R = 0.5$ and it is centered in the $\tau_h$ direction (see Figure 3.12).

The isolation-sum discriminator is computed using all the charged particles and photons reconstructed by the PF algorithm that lie within the cone, excluding those that are identified as constituents of the tau candidate. It is defined as:

$$I_\tau = \sum p_T^{charged}(d_z < 0.2\text{cm}) + \max \left(0, \sum p_T^2 - \Delta\beta \sum p_T^{charged}(d_z > 0.2\text{cm})\right), \quad (3.9)$$
Figure 3.12: Cone definition for tau isolation discriminator. Taken from [85].

where any contribution coming from PU interactions are suppressed through requirements on the production vertex of the charged hadrons ($d_z$):

- The charged hadrons that come from PU are suppressed, requiring that the tracks are originated in the production vertex of the $\tau_h$ ($d_z < 0.2\text{cm}$).

- The PU contribution to the photon reconstruction is suppressed requiring that the charged particles, within a cone of size $\Delta R = 0.8$, are not originated in the production vertex of the $\tau_h$ ($d_z > 0.2\text{cm}$). The so-called $\Delta\beta$-factor makes the tau efficiency independent of the PU. For Run II, the $\Delta\beta$ value used is 0.2.

The loose, medium and tight working points (WP) for the isolation-sum discriminant are defined requiring $I_\tau$ to be less than 2.5 GeV, 1.5 GeV, and 0.8 GeV, respectively [84]. Figure 3.13 shows a comparison of the isolation-sum discriminators’ performance for the previous and the current versions of the HPS algorithm. As was mentioned in the previous section, the main difference between them is the size used for the strip reconstruction, which was fixed for Run I, while for Run II it depends on the $p_T$ of the electrons and photons (dynamic-strip reconstruction). The dynamic-strip reconstruction allows to include the photons (coming from the tau decay products) that might lie outside of the signal cone, reducing the jet $\rightarrow \tau_h$ misidentification rate (see Figure 3.13). In order to compare the performance of the working points, simulated signal samples corresponding to $H \rightarrow \tau\tau$ and $Z' \rightarrow \tau\tau$ events were used, while simulated background samples corresponding to QCD multijet processes were used. This study was performed using pp collisions at $\sqrt{s} = 13$ TeV. For the current version of the HPS algorithm, the tight isolation-sum discriminator has a $\tau$ identification efficiency of $\sim60\%$, and the QCD-jet rejection rate is $\sim99.7\%$.

$\Delta\beta$-factor is defined as the ratio between the energy of all charged hadrons and the energy of all photons that come from PU.  

---

$^\dagger$The $\Delta\beta$-factor is defined as the ratio between the energy of all charged hadrons and the energy of all photons that come from PU.
Figure 3.13: Misidentification probability as a function of the \( \tau_h \) identification efficiency, evaluated using \( H \to \tau\tau \) (left) and \( Z' \to \tau\tau \) (right), and multijet simulated samples. Four configurations of the reconstruction and isolation method are compared (three of them for strip reconstruction using a fixed size, and the remaining one, for the dynamic-strip reconstruction). The three points on each curve correspond to, from left to right, the tight, medium, and loose WPs. The misidentification probability is calculated relative to jets that pass minimal \( \tau_h \) reconstruction requirements. For the current version of the HPS algorithm (dynamic-strip reconstruction) the tight isolation discriminator shows a better tau identification efficiency and a lower misidentification rate. Taken from [84].

Additionally to the isolation-sum discriminator, in Run II, an MVA-based discriminator is applied for the tau identification. It has a better efficiency and a lower misidentification rate than the cutoff-based discriminator used in Run I. The MVA-based discriminator uses the following variables as input:

- the \( p_T \) and the \( \eta \) direction of the \( \tau_h \) candidate,
- the transverse impact parameter \( d_0 \) of the leading (highest-\( p_T \)) track of the \( \tau_h \) candidate and its significance \( (d_0/\sigma_{d_0}) \),
- the isolation variable defined in equation 3.9,
- the reconstructed decay mode,
- the multiplicity of photons an electrons, with \( p_T > 0.5 \) GeV, that lie within the signal and the isolation cones,
- the distance between the primary and secondary vertices and its significance, when the secondary vertex is identified.

The MVA-based discriminator is trained to distinguish the \( \tau_h \) candidate (signal) from QCD-jets (background). The simulated samples used are: in the case of the signal, \( Z/\gamma \to \tau\tau \), \( H \to \tau\tau \),
\( Z' \rightarrow \tau\tau \), and \( W' \rightarrow \tau\nu \); and in the case of the background, multijet, \( W+jets \), and \( t\bar{t} \). All the events were weighted such that the training of the MVA-discriminator was independent of the \( p_T \) and \( \eta \) distributions of the \( \tau_h \). Once the discriminator was trained, six working points were defined according to the \( \tau_h \) identification efficiency. Figure 3.14 shows a comparison among the working points of the MVA-based and cutoff-based (Run I) discriminators, which were estimated with \( H \rightarrow \tau\tau \) and \( Z' \rightarrow \tau\tau \) simulated samples for the signal, and QCD multijet simulated samples for the background.

![CMS Simulation Preliminary](image1)

![CMS Simulation Preliminary](image2)

Figure 3.14: Misidentification probability as a function of \( \tau_h \) identification efficiency, evaluated using \( H \rightarrow \tau\tau \) and QCD simulated samples (left), and \( Z' \) (2 TeV) and QCD simulated samples (right). The MVA-based discriminators (red line) are compared to that of the isolation sum discriminators (pink line). The points correspond to working points of the discriminators. The three working points of the isolation-sum discriminator curve are, from left to the right, tight, medium, and loose working points. The five working points of the MVA-based discriminator curve are, from left to the right, very tight, tight, medium, loose, and very loose. The misidentification probability is calculated with respect to jets, which pass minimal \( \tau \) reconstruction requirements. Taken from [84].

Figure 3.15 shows the expected identification efficiency and the misidentification rate, as a function of \( p_T \), for each MVA-based working point. In this analysis, the MVA-based tight discriminator WP is used in order to achieve a high purity in the di-\( \tau \) selection, while keeping a considerable QCD-jet background rejection. For the current version of the HPS algorithm, the MVA-based tight discriminator has a \( \tau \) identification efficiency of \( \sim 60\% \), and the QCD-jet rejection rate is \( \sim 99.8\% \).

**Tau Discrimination against electrons**

An electron signature has some probability to be misidentified as a \( h^- \), or a \( h^+ \pi^0 \), decay modes of a tau. The probability is higher in the last case, since electrons emit bremsstrahlung that can convert into photons in the tracker material, faking the \( \pi^0 \) signature. In consequence, a MVA-algorithm has been developed in order to discriminate electrons from taus, exploiting \( \tau_h \) features
Figure 3.15: Efficiency of the $\tau_h$ identification estimated with simulated $Z/\gamma \rightarrow \tau\tau$ events (left) and the misidentification probability estimated with simulated QCD multi-jet events (right) for the very loose, loose, medium, tight, very tight, and very very tight working points of the MVA based $\tau_h$-isolation algorithm. The efficiency is shown as a function of the $\tau_h$ transverse momentum while the misidentification probability is shown as a function of the jet transverse momentum. Taken from [84].

like the amount of Bremsstrahlung associated to the leading track and the low multiplicity of particles. The MVA-based against electrons discriminator uses the following variables as input:

- the $p_T$ and the $\eta$ direction of the $\tau_h$ candidate,
- the mass of the $\tau_h$ candidate,
- the Electromagnetic energy fraction, $E_{\text{ECAL}}/(E_{\text{ECAL}} + E_{\text{HCAL}})$, of all PF candidates associated to the $\tau_h$,
- the ECAL and HCAL energies relative to the momentum of the leading charged-hadron track ($E_{\text{ECAL}}/p_{\text{lead}}^T$ and $E_{\text{HCAL}}/p_{\text{lead}}^T$),
- the $p_T$-weighted root-mean-squared distances in the $\eta$ and the $\phi$ directions ($\sqrt{\sum (\Delta \eta)^2 p_T}$ and $\sqrt{\sum (\Delta \phi)^2 p_T}$) between the strips and the leading charged-hadron track,
- the fraction of the tau energy carried by the photons ($\sum E_\gamma/E_\tau$),
- the ratio between the ECAL energy and the inner tracker momentum, $(E_e + \sum E_\gamma)/p_{\text{in}}$,
- the Bremsstrahlung fraction $f_{\text{brem}}$ measured by the GSF (see Section 3.3),
- the $\chi^2/N_{\text{dof}}$ of the GSF,
- the fraction ($N_{\text{hits}}^{\text{GSF}} - N_{\text{hits}}^{\text{KF}}$)/($N_{\text{hits}}^{\text{GSF}} + N_{\text{hits}}^{\text{KF}}$), where $N_{\text{hits}}^{\text{GSF}}$ and $N_{\text{hits}}^{\text{KF}}$ are the number of hits associated to the track by the GSF and the KF algorithms, respectively since they will show up any Bremsstrahlung emission in the Tracker System.
The MVA-based against discriminator is trained with the simulated samples: $Z/\gamma \rightarrow \tau\tau$, $H \rightarrow \tau\tau$, $Z' \rightarrow \tau\tau$, and $W' \rightarrow \tau\bar{\nu}_\tau$ for the signal, while $Z/\gamma \rightarrow ee$, $H \rightarrow ee$, $Z' \rightarrow ee$, and $W' \rightarrow e\bar{\nu}_e$ for the background. Six WP are defined according to the tau-reconstruction efficiency. The efficiency and the misidentification rate are uniform over $p_T$ (see Figure 3.16). In this analysis the loose MVA-based against electron discriminator is used since it keeps a high tau reconstruction efficiency (~83%) and a relative low misidentification rate ($10^{-2}$).

Figure 3.16: Efficiency of the $\tau_h$ identification estimated with simulated $Z/\gamma^* \rightarrow \tau\tau$ events (left) and the $e \rightarrow \tau_h$ misidentification probability estimated with simulated $Z/\gamma^* \rightarrow ee$ events (right) for the very loose, loose, medium, tight and very tight working points of the MVA-based anti-$e$ discrimination algorithm. The efficiency is shown as a function of the $\tau_h$ transverse momentum while the misidentification probability is shown as a function of the $e$ transverse momentum. Both, efficiency and misidentification probability, are calculated for $\tau_h$ candidates with a reconstructed decay mode and passing the loose working point of the isolation sum-discriminator. Taken from [84].

**Tau Discrimination against muons**

A muon signature can fake the one produced by a $h^\pm$ decay mode of a $\tau_h$; due to this, discriminants have been developed in order to distinguish between them. The cutoff-based discriminator is used since it has a similar efficiency than the MVA-based one. Two working points have been defined:

- **Loose**: A $\tau_h$ candidate fails this discriminant when at least two track segments are found in the Muon Chambers within a cone of size $\Delta R < 0.3$ and centered in the $\tau_h$ direction, or when the sum of the ECAL and HCAL energy deposits are less than 20% of the momentum of the leading track of the $\tau_h$ candidate.

- **Tight**: A $\tau_h$ candidate passes this discriminant when it has passed the loose WP and no hits are found in the CSCs, DTs and RPCs located in the two outermost Muon Stations, within a cone of size $\Delta R < 0.3$ and centered in the $\tau_h$ direction.
The tau identification efficiency for both WPs is more that 99% and the misidentification probability is $3.5 \times 10^{-3}$ and $1.4 \times 10^{-3}$ for the loose and tight WP, respectively. In this analysis, the tight working point is used.

**Summary**

The particle object reconstruction used in this analysis can be summarized as:

- **Tau Reconstruction**: The hadronic taus are reconstructed using the HPS algorithm, requiring 1or3-prongs with the newDMF discriminator. The signal cone has a radius of $R_{\text{sig}} = 3.0/p_T$(GeV) and it is centered in the direction of the tau momentum. In order to identify the hadronic taus from QCD-jets, the tight WP of the MVA-based isolation discriminator was used, reaching a $\sim$60% of $\tau$ identification efficiency and a $\sim$99.8% of rejection rate against QCD-jets. The isolation cone has a size of $\Delta R = 0.5$ and it is centered in the $\tau_h$ direction. In order to discriminate taus from electrons, the loose MVA-based discriminator was used, achieving a tau reconstruction efficiency of $\sim$83% and keeping a misidentification rate of the order of $10^{-2}$. In order to discriminate taus from muons, the tight WP of cutoff-based discriminator was required, which reaches a 99% of tau reconstruction efficiency and a $1.4 \times 10^{-3}$ of misidentification rate.

- **Jet Reconstruction**: Since jets are used as input for the HPS algorithm, its reconstruction is also crucial in this analysis. Jets are reconstructed using anti-kT algorithm (loose WP), with the distance parameter $\Delta R(\phi, \eta) = 0.4$. Jets are required to have $p_T > 30 \text{ GeV}$ and $|\eta| < 2.4$.

- **b-jet Reconstruction**: The b-jets are reconstructed using the loose working point of the CSVv2 algorithm, which provides an efficiency of 83%.

- **Electron Reconstruction**: The electrons are reconstructed using the 90% efficiency working point of the MVA-based electron ID algorithm, with the isolation requirement of $\Delta R < 0.4$.

- **Muon Reconstruction**: PF “global muons” with an isolation cone of $\Delta R < 0.4$ are required.

- **Missing Transverse Energy Reconstruction**: The $\vec{E}_T$ is reconstructed with the PF algorithm, where the “type-I” correction is applied.
4 Experimental signature of $Z' \to \tau_h \tau_h$

Since a $Z'$ boson is a hypothetical massive, colorless and electrically neutral particle, which might couple to the SM fermions, it would decay into two taus with opposite electric charged (see Figure 4.1). Moreover, taus can decay leptonically or hadronically (see Section 1.3) and, therefore, there are six possible ditau final state signatures: $\tau_\mu \tau_\mu$, $\tau_\tau \tau_e$, $\tau_\tau \tau_\mu$, $\tau_\mu \tau_h$, $\tau_\tau \tau_h$, and $\tau_h \tau_h$, as can be seen in Table 4.1. However, in the $\tau_\mu \tau_\mu$ and $\tau_\tau \tau_e$ cases, it is not possible to distinguish experimentally between the signatures of electrons or muons coming from tau decays and the signatures of those coming directly from the hard interaction; in consequence, the searches for $Z'$ bosons in the ditau channel usually involve the other four final states: $\tau_\tau \tau_\mu$, $\tau_\mu \tau_h$, $\tau_\tau \tau_h$ and $\tau_h \tau_h$. The most sensitive channel is the dihadronic tau final state since, even though there is a high QCD contamination, it has the highest branching ratio. The $\tau_\tau \tau_h$ channels also have a significant sensitivity, due to the high efficiency of the light lepton reconstruction, and a relative low QCD background contamination. In this thesis, the search for $Z'$ bosons was performed in the dihadronic tau final state. This is one of four channels used in the $Z' \to \tau \tau$ analysis. The combination of the four ditau signatures allows to increase the statistics, which improves the sensitivity of the search; the methodology used for the combination will be described in Chapter 6.

Figure 4.1: Feynman diagram for the signal process ($Z' \to \tau \tau$). Taken from [87].

This chapter is organized as follows: the signature of the expected $Z' \to \tau_h \tau_h$ events is described in section 4.1, the possible SM processes that can fake the $Z' \to \tau_h \tau_h$ signature are studied in section 4.2, the methodology used to distinguish between signal and background is described in section 4.3, finally, the strategy to search for $Z' \to \tau_h \tau_h$ events is presented in section 4.4.

4.1 Signature

Due to the large mass of the $Z'$ boson, the two taus coming from $Z' \to \tau \tau$ decay are expected to have a high transverse momentum, to be oppositely charged and to travel in opposite directions. Since these taus would be very energetic, their decay products are expected to be collinear with
Table 4.1: Six possible ditau signatures of the $Z'$ → $\tau\tau$ channel and their branching fractions.

<table>
<thead>
<tr>
<th>$Z'$ ditau signatures</th>
<th>Branching Ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_\mu \tau_\mu$</td>
<td>3.1</td>
</tr>
<tr>
<td>$\tau_e \tau_e$</td>
<td>3.1</td>
</tr>
<tr>
<td>$\tau_e \tau_\mu$</td>
<td>6.2</td>
</tr>
<tr>
<td>$\tau_\mu \tau_h$</td>
<td>22.5</td>
</tr>
<tr>
<td>$\tau_e \tau_h$</td>
<td>23.1</td>
</tr>
<tr>
<td>$\tau_h \tau_h$</td>
<td>42.0</td>
</tr>
</tbody>
</table>

the original $\tau_h$. Additionally, due to the presence of neutrinos in the tau decay, $Z'$ boson events would include missing transverse momentum, $\vec{E}_T$, which would point towards the direction of the less energetic tau. These features defines the topological region where the $Z' \rightarrow \tau_h \tau_h$ events (also known as “signal” events) are expected. Therefore, an event selection is performed considering these topological features in order to identify the signal and to reject, when possible, any background.

4.2 Expected Background

There are several SM processes that have similar signatures and topologies than the $Z' \rightarrow \tau_h \tau_h$ events. Such SM processes are known as “backgrounds”, and their understanding is crucial to quantify the amount of SM events expected in the high mass region, where the signal events would be observed. The background contamination in the signal region can be reduced with topological-event selections, considering the differences between the $Z' \rightarrow \tau_h \tau_h$ and the background signatures. The main contamination comes from the QCD multijets events which represent around 70% of the total background. The background processes considered in this analysis are:

- **QCD multijet production**: As was mentioned in section 3.7, the main source of background for the tau identification is the QCD-jet production. Although the jet → $\tau_h$ misidentification rate is small (of the order of $10^{-2}$), its background contribution is non-negligible since it has a significantly larger production cross section. This makes the QCD multijet events the dominant background in the $Z' \rightarrow \tau_h \tau_h$ channel (70% of the total background). Additionally, a QCD-jet fakes the tau signature with a probability at least one order of magnitude higher than the misidentification probability coming from electrons or muons, resulting in a high contamination of QCD multijet events in the $\tau_h \tau_h$ channels, in comparison with the semileptonic channels ($\tau_\ell \tau_h$). However, the QCD multijet events do not have the intrinsic momentum imbalance of the signal (resulting from the presence of neutrinos) and, therefore, a requirement on the missing transverse energy will reduce this background. Furthermore, the QCD multijet events do not have a significant contribution in the high mass region, where the signal events are expected; in consequence, they would not affect considerably the sensitivity of the analysis. Figure 4.2 shows a schematic view of the QCD dijet production.
• **Drell-Yan process** \((pp \rightarrow Z/\gamma^* \rightarrow \tau_h\tau_h)\): Other background contribution comes from Drell-Yan processes in which the tau pair comes from the \(Z/\gamma^* \rightarrow \tau_h\tau_h\) decay. Since it has the same signatures than the signal this background is irreducible, in other words, no topological criteria can be applied in order to suppress its contribution. However, this background can be distinguished from the signal, using the invariant mass distribution: the \(Z/\gamma^* \rightarrow \tau_h\tau_h\) process has a mass peak in the low region of the invariant mass distribution \((m(\tau_1, \tau_2) < 100 \text{ GeV})\), while the new resonance is expected in the high mass region. The properties of the Drell-Yan processes are known to a high degree of accuracy and, indeed, the \(Z/\gamma^* \rightarrow \tau_h\tau_h\) events are used to validate the tau identification criteria in the analysis (see Section 5.5.2). Figure 4.3 shows the Feynman diagram for the Drell-Yan process production.

![Feynman diagram for the Drell-Yan process](image)

Figure 4.3: Feynman diagram for the \(Z/\gamma^* \rightarrow \tau_h\tau_h\) process. Taken from [88].
• **W+Jets events:** The production of a W boson in association with jets (W+jets) can fake the signal events since the W decays 11.4% of the times into a tau plus a neutrino ($W \rightarrow \tau \nu$), while a jet can be misidentified as a hadronic tau; if such jet is in the opposite hemisphere than the W decay products, it would result into a misidentified back-to-back $\tau_h \tau_h$ signal with a momentum imbalance due to the neutrino. Similarly, W+Jets can mimic the signal when the W boson decays hadronically (faking the $\tau_h$ signature) and the jet (misidentified as a tau) is produced in the opposite direction; however, the last scenario can be suppressed since it does not have an intrinsic momentum imbalance. The W+Jets background contribution depends strongly on the jet $\rightarrow \tau_h$ misidentification rate and, therefore, it can be highly reduced through the tau identification criteria. The W+Jets events represent 7% of the total background contribution. Figure 4.4 shows the Feynman diagram for the W+Jets production process.

![Feynman diagram for the W+Jets process. Taken from [87](https://example.com).](image)

• **$t\bar{t}$ production:** Another background comes from the top-antitop production. The top quark decays most of the times into a bottom quark and a W boson ($t \rightarrow Wb$). As in the case of the W+Jets, the W can result into a genuine tau coming from the $W \rightarrow \tau \nu$ decay, or a misidentified tau coming from its hadronic decay. Therefore, each W boson, that comes from the top and the antitop decays, can mimic the dihadronic tau signature with a momentum imbalance due to the presence of neutrinos. Additionally, the b quark coming from the top decay can result into a hadronic jet, which can fake the $\tau_h$ signature. This background is highly suppressed through the rejection of any jet identified as a b-jet; with this purpose, the b-tagging algorithm described in section 3.5 is used. The $t\bar{t}$ production represents less than 1% of the total background. Figure 4.5 shows the Feynman diagram for the $t\bar{t}$ production process.

• **Diboson processes:** The diboson background (labeled as VV) refers to the $ZZ$, $WZ$, and $WW$ production. As was mentioned above, both, Z and W bosons, can result into a genuine or a fake tau in the final state. This background accounts for all the possible combinations
of tau pairs, coming from the diboson processes. The tau pairs can mimic the topological features of the signal. The dibosons SM process does not have a significant contribution to the total background (less than 1%). Figure 4.6 shows the Feynman diagram for the diboson production process.

- **H → τₜτₜ process**: the Higgs boson can decay into two-hadronic taus and, similarly to the Drell-Yan case, this background has the same topology than the signal. However, the contribution from H → τₜτₜ is not significant since the events lie in the low region of the invariant mass distribution. Also, in contrast with the Drell Yan process, it has a small production cross section. The H → τₜτₜ process only represents the 0.035% of the total background in the signal region.

In summary, the dominant background for the Z' → τₜτₜ signature comes from the QCD multijet events, which correspond to around 70% of the total; this background can be reduced since it does not have the momentum imbalance of the signal. The other important background comes from Drell-Yan processes, which have the same signature than the signal with exception of the low
invariant mass of the decay products; this background is known with a high degree of accuracy and it can be used to validate the tau identification criteria used in this analysis. The remaining backgrounds, W+Jets, $t\bar{t}$ and diboson, can be reduced through tau identification criteria and topological event selections, and they represent 8% of the total background. The $H \rightarrow \tau\tau$ background is not considered in this analysis since its contribution is negligible (0.035%).

### 4.3 Mass Reconstruction

As has been mentioned, if a $Z'$ boson exists, it would be observed as a peak in the invariant mass distribution of its decay products and, therefore, the search for this new boson consists on distinguishing this heavy resonance from the background processes that lie in the low mass region. In the particular $Z' \rightarrow \tau\tau$ case, it is not possible to reconstruct fully the ditau invariant mass due to the presence of neutrinos in the final state (see Section 3.7). In consequence, a narrow peak will not be observed and, instead, the $Z'$ events would manifest themselves as an enhanced mass above the expected level of background. Nevertheless, in order to consider the unmeasured energy of the neutrinos, the missing transverse energy can be included in the mass calculation. A $M(\tau_1, \tau_2, E_T)$ variable, known as the effective visible mass, defined as:

$$M(\tau_1, \tau_2, E_T) = \sqrt{(E_{\tau_1} + E_{\tau_2} + E_T)^2 - (p_{\tau_1} + p_{\tau_2} + \vec{E}_T)^2},$$  

includes the intrinsic missing transverse energy of the $Z' \rightarrow \tau\tau$ events and, therefore, allows an improved discrimination of the signal against the backgrounds when compared with the simple invariant mass variable. In summary, the search for $Z' \rightarrow \tau\tau$ events reduces to the search for enhanced heavy resonances in the $M(\tau_1, \tau_2, E_T)$ distribution of the dihadronic tau decay products. Figure 4.7 shows the $M(\tau_1, \tau_2, E_T)$ distribution of the dihadronic final state for the expected background and a simulated signal sample with an expected mass of 3 TeV. As can be seen in the figure, most of the backgrounds lie in the low mass region of the mass spectrum, while the expected signal is located in the high mass region.

### 4.4 Strategy

Since a $Z'$ is a neutral and massive boson, we look for two oppositely-charged taus with high transverse momentum, that travel in opposite directions. As was mentioned above, the signal events must have missing transverse energy due to the presence of neutrinos. An event selection is performed according to these topological features, in order to identify the signal with high efficiency; additionally, these selection criteria allow to optimize the rejection of any background process that can mimic the signal, reducing the influence of systematic uncertainties.

The discrimination between signal and background events is performed using the effective visible mass variable, since the $Z' \rightarrow \tau\tau$ events would give high values and the background processes would give low values. In order to interpret the data and to extract any possible excess from the SM expectation, or to establish exclusion limits, a shape analysis, based on binned likelihood functions, was implemented. In chapter 5 the whole analysis is presented, starting from the description of the data sets, the event selection, the estimation of background and the systematic effects.
Figure 4.7: $M(\tau_1, \tau_2, E_T)$ distribution of dihadronic tau final state for the expected backgrounds and a $Z'$ signal with an expected mass of 3 TeV. The distribution was obtained from the events that have passed the selection criteria described in section 5.3. The background estimation was performed according the method described in section 5.5.
5 Search for Z’ Bosons with 2016 Data

In this chapter, the search for Z’ bosons in the dihadronic tau channel performed using the data collected by CMS during 2016, is presented. This analysis makes part of the search for heavy resonances in the ditau final state, in which I am one of the main authors/analysts [35].

The structure of the chapter is as follows: first, the data sets used for the search are described in sections 5.1 and 5.2. Section 5.1 presents the selected triggered events, with at least two hadronic taus in the final state, while Section 5.2 lists the data collected by CMS during 2016, as well as the simulated samples used for the background. Once the data and simulated samples are defined, an event selection is performed according to the topological features of the $Z' \rightarrow \tau_h \tau_h$ signature; this event selection is specified in section 5.3. The estimation of the amount of background events that pass the signal selection criteria are described in section 5.5. Finally, once the background estimation is validated, the systematic effects that influence the expected signal and background events are discussed in section 5.6.

5.1 Trigger Selection

The $\tau_h$’s are reconstructed in both levels of the trigger system: the L1-Trigger and the HLT (see section 2.7). The L1 tau trigger only uses the information provided by the calorimeter system, in order to identify and to gather all the neutral pions that come from the tau decays; with the purpose of reducing the jet to tau misidentification rate at L1-level, a requirement on the tau isolation energy is applied [90]. The HLT tau candidate is built combining tracks and calorimeter clusters, using a simplified PF algorithm, where the track requirements are relaxed. At least three charged hadrons, located in a narrow cone, are required, while the calorimeter clusters are reconstructed in the same way as in the off-line case (see Section 3.7) but, at HLT-level there is no discrimination among the tau decay modes. Additionally, the cut-based isolation requirement is applied on the HLT candidates to reduce the jet misidentification rate.

Events that fired the $\text{HLT\_DoubleMediumIsoPFTau35\_Trk1\_eta2p1}$ trigger are considered for this analysis. This double-tau HLT trigger is seeded by a double-tau L1 trigger, and selects taus with a medium isolation WP, and a transverse momentum higher than 35 GeV in the $|\eta| < 2.1$ region. Due to the simplified track reconstruction performed in the HLT, the trigger constrains the tau identification acceptance to the $|\eta| < 2.1$ region. The double-tau HLT trigger paths changed for the 2016 run since the L1-trigger system, and the calorimeter system, were upgraded; this improves the performance in the selection of $\tau_h$ at L1-trigger and, consequently, the ditau trigger selection.

Since part of the trigger system is hardware-based, it is difficult to emulate the trigger behavior on simulated samples, in particular, the effects coming from the high luminosity and PU contributions. In consequence, the trigger selection was not applied on simulated samples and, instead, it was modeled as a weight on the events; the weight was obtained from the trigger efficiency measured with data events. Nevertheless, no impact on the sensitivity was found in the case of applying the trigger on the simulated samples (see appendix A), and therefore both methods (applying or not applying the trigger selection on simulated samples) could be used.
In this analysis, the HLT trigger selection \((\text{HLT}_\text{DoubleMediumIsoPFTau35_Trk1_eta2p1})\) was applied to data events only.

5.1.1 Trigger Efficiency

As was mentioned above, the trigger selection was applied only on data events, and on the simulated samples it was modeled applying an efficiency weight. The trigger efficiency weight was estimated from real data (full 2016 data sample) using the tag-and-probe method. The tag-and-probe method consists in selecting \(Z \rightarrow \tau \tau \rightarrow \mu \tau h\) events in order to obtain a relatively clean sample of \(\mu \tau\) leptons, taking advantage of the fact that the invariant mass of the final states is near to the \(Z\) mass peak. One of the final states of the \(Z\) events can be used to reduce the background (the tag), while the second one can be used to measure the efficiency of the selection (the probe). Then, in order to measure the trigger efficiency for one tau-lepton, the muon is used as the tag, while the tau is used as the probe. Then, the trigger efficiency per tau-lepton is given by:

\[
\epsilon = \frac{\text{events that fire a } (\mu + \tau)\text{-trigger}}{\text{events that fire a } \mu\text{-trigger}}.
\]

The \((\mu + \tau)\)-cross trigger used to select the “numerator” events in equation (5.1) is \(\text{HLT}_\text{IsoMu21_eto2p1_MediumIsoPFTau32_Trk1_eto2p1}\). Since this trigger has the same L1 requirements for the tau-lepton than those of the double-tau trigger used for the analysis, it can be used to calculate the single-tau trigger efficiency for the simulated sample selection. In order to select the \(Z \rightarrow \tau \tau \rightarrow \mu \tau h\) events for the “denominator” in the equation (5.1), the following conditions are required:

- events must fire the \(\text{HLT}_\text{IsoMu24}\) trigger,
- exactly 1 global muon with \(|\eta| < 2.1\) and \(p_T > 24\) GeV,
- the muon passes the medium isolation criterion,
- muon best track with \(d_{xy} < 0.2\) cm and \(d_z < 0.045\) cm with respect to the primary vertex,
- relative \(\mu\) isolation < 0.1 (with \(\delta \beta\) corrections),
- at least one \(\tau_h\) reconstructed using the HPS algorithm, with \(|\eta| < 2.1\) and \(p_T > 20\) GeV,
- MVA discriminator against electrons: “againstElectronVLooseMVA6”,
- cutoff-based discriminator against muon: “againstMuonTight3”,
- \(\text{newDMF}\) with 1 or 3 charged hadrons,
- MVA isolation discriminator: “byTightIsoMVARun2v1DBnewDMwLT”,
- \(p_T > 5\) GeV for the leading track of the \(\tau_h\), with \(d_{xy} < 0.2\) cm and \(d_z < 0.045\) cm with respect to the primary vertex,
- \(\Delta R(\mu, \tau_h) > 0.5\),
- \(Q(\mu) \times Q(\tau_h) < 0\),
- \(40 < m(\mu, \tau_h) < 80\) GeV,
- $m_T(\mu, E_T) < 30$ GeV,
- 0 jets tagged as b-jets.

The “numerator” events also pass the same event selection criteria but, additionally, they must fulfill the $(\mu + \tau)$ cross trigger. The reconstructed $\tau_h$ candidates that pass the $Z \rightarrow \tau\tau \rightarrow \mu\tau_h$ event selection, have the same identification criteria used for the $Z' \rightarrow \tau_h\tau_h$ analysis. Such criteria are described in section 5.3.1. Figure 5.1 shows the tau trigger efficiency, obtained from equation 5.1 as a function of the tau-p$_T$.

![Figure 5.1: Trigger efficiency per $\tau_h$ as a function of p$_T$ for data measured from the 2016 full data sample, using $Z \rightarrow \tau\tau \rightarrow \mu\tau_h$ events.](image)

The function used to perform the fit of the trigger efficiency curve is:

$$pdf = P[0] + P[1] \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{t^2}{2}} dt,$$

where,

$$y = \frac{\sqrt{x} - \sqrt{P[2]}}{2P[3]} , \quad P[2] = \mu, \quad P[3] = \sigma .$$

The trigger weight applied per tau-lepton is obtained from this fit.

In summary, the trigger selection is not applied on simulated samples, instead, it is modeled by weighting the simulated predicted events. The weight, for each tau-lepton, is obtained from the fit of the trigger efficiency curve from the data. Figure 5.1 shows that the trigger efficiency reaches a plateau value of $\epsilon=0.95$ at tau p$_T = 70$ GeV. In consequence, in order to obtain a well modeled trigger selection and to avoid the turn-on curve, the hadronic taus must have transverse momenta greater than 70 GeV, in other words, only the hadronic taus with p$_T > 70$ GeV are selected and only the plateau region of the trigger efficiency is relevant for this analysis. Due to
the inefficiency observed in the plateau region, a systematic uncertainty of 5% is assigned per hadronic tau due to the trigger selection. The overall trigger uncertainty, when considering both taus, is 10%.

5.2 Data and Simulated samples

This section presents the data samples and the simulated samples for signal and for background used in this analysis.

5.2.1 Data Samples

This analysis uses the data collected by the CMS experiment from the proton-proton collisions at centre-of-mass energy of 13 TeV delivered by the LHC during 2016. The data recorded corresponds to a total integrated luminosity of 37.76 fb$^{-1}$ (Figure 2.4 right). However, the data validated by the CMS Data Quality Monitoring team (CMSDQM) to be used for physics analyses has an integrated luminosity of 35.9 fb$^{-1}$. In order to guarantee that the right data is used, an official JSON file is issued containing all the “good” run ranges and lumin sections. The certified JSON file used in this analysis was:

\[\text{Cert}_2\text{71036-28}\text{4044}_{\text{13 TeV}}_{\text{03Feb2017ReReco_Collisions16}}_{\text{JSON.txt}}.\]

For this thesis, the datasets, which include at least one triggered tau-lepton at HLT level, are shown in Table 5.1. The trigger selection described in the previous section was applied to these datasets.

<table>
<thead>
<tr>
<th>Physics Sample</th>
<th>Official CMS Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run 2016C Tau Run2016C-03Feb2017</td>
<td>/Tau/Run2016C-03Feb2017-v1/MINIAOD</td>
</tr>
<tr>
<td>Run 2016F Tau Run2016F-03Feb2017</td>
<td>/Tau/Run2016F-03Feb2017-v1/MINIAOD</td>
</tr>
</tbody>
</table>

Table 5.1: Collision Data Samples

5.2.2 Simulated Samples

The event simulation consists of two steps: the event generation simulates the particle production of a specific physics process; in the second step, the detector response is simulated. These processes are very complex and no analytical calculations can be performed, therefore, Monte Carlo (MC) techniques are used.

**Event Simulation**

- **Event Generation**: The event generation is performed in four steps: the hard scattering, the parton shower, the hadronization and the underlying event simulations. The hard
scattering simulation consists of the calculation of the probability amplitude for a specific physics process as a result of a pp-collision, using the parton distribution functions (PDFs) and the matrix element (ME). A PDF is the probability density as a function of the momentum fraction carried by a given parton. These functions are extracted from deep inelastic scattering data and from extrapolations to high energies. The ME represents the probability amplitude of the process; its estimation depends on the level of the calculation performed. Event generations usually includes Leading Order terms (LO) or Next-to-Leading order terms (NLO), depending on the availability of the calculations. Once the physics process of interest is generated, then the subsequent hadronization processes are simulated using models built to describe them. Finally, the underlying event is simulated depending on the energy and the processes from which it was produced. As a result of the event generation, the four momentum of each particle emerging from the collision is produced. The most common generators used are: Pythia [91], Madgraph [92], POWHEG [93, 94], Comphep [95], Sherpa [96], among others. Pythia is widely used due to the accurate modeling of parton shower production and fragmentation, while POWHEG has a very accurate description of processes like top quark decay and Madgraph is very precise describing SM processes like Drell-Yan.

- **Detector Response Simulation:** The second step is the simulation of particles passing through the detector. In order to describe such interactions, a detailed modeled of the CMS geometry is needed. The software must be able to accurately describe the effects of the magnetic field, the electromagnetic and hadronic shower production, and so on. For this purpose, the CMS Collaboration uses the package GEANT4 [97, 98]. As a result all the observable final states produced from the pp collision, as their kinematic measured variables, are given.

The whole event simulation results into a RAW data file, which has the same format as the one delivered by the HLT system with real data. Then, the simulated events are reconstructed, following the data processing described in Section 2.7.3.

**Signal and Background Simulated Samples**

The CMS collaboration issues official simulated samples for all the Standard Model processes to be used in physics analyses. In this thesis, the official Spring _miniAODv2_ 2016 simulated samples were used. The signal samples of the Z’ production, as well as the samples for the SM processes, described in section 4.2, were generated using Pythia8, Madgraph-v5, and POWHEG, where the tau-lepton decays were simulated with a specialized package called TAUOLA [99]. The detector simulation was performed using GEANT4.

The background simulated samples used in this thesis are shown in Table 5.2. The production of the Drell-Yan and QCD multijet simulated samples was performed using the next-to-leading order Madgraph-v5 generator (NLO_MG5), while the same generator was used for the W+Jets background but with the leading order version (LO_MG5). The Pythia8 generator was used for the production of diboson background, while the POWHEG generator was used for the tt background.

The Drell-Yan simulated samples were binned using the mass of the generated lepton pair, while the QCD multijet and W+Jets simulated samples were binned using $H_T$.[79]

$H_T$, the transverse hadronic energy, is defined as the scalar sum of the transverse momenta of the jet
Similar to the background, the signal samples are simulated using Pythia8, and the tau decays are simulated using TAUOLA. The signal samples for different $Z'$ mass values, using the SSM, are shown in Table 5.3.

<table>
<thead>
<tr>
<th>Process</th>
<th>samples</th>
<th>MC generator</th>
<th>cross-section (pb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z \rightarrow l^+ l^-$</td>
<td>$m &gt; 50 \text{ GeV}$</td>
<td>NLO_MG5</td>
<td>5.765.4900</td>
</tr>
<tr>
<td></td>
<td>mass binned</td>
<td>NLO_MG5</td>
<td>7.9077</td>
</tr>
<tr>
<td></td>
<td>NLO samples</td>
<td>NLO_MG5</td>
<td>0.4263</td>
</tr>
<tr>
<td></td>
<td>$400 \text{ GeV} &lt; m &lt; 500 \text{ GeV}$</td>
<td>NLO_MG5</td>
<td>0.2390</td>
</tr>
<tr>
<td></td>
<td>$700 \text{ GeV} &lt; m &lt; 800 \text{ GeV}$</td>
<td>NLO_MG5</td>
<td>0.0341</td>
</tr>
<tr>
<td></td>
<td>$1000 \text{ GeV} &lt; m &lt; 1500 \text{ GeV}$</td>
<td>NLO_MG5</td>
<td>0.0150</td>
</tr>
<tr>
<td></td>
<td>$1500 \text{ GeV} &lt; m &lt; 2000 \text{ GeV}$</td>
<td>NLO_MG5</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>$2000 \text{ GeV} &lt; m &lt; 3000 \text{ GeV}$</td>
<td>NLO_MG5</td>
<td>0.0004</td>
</tr>
<tr>
<td>$W + jets$</td>
<td>$0 \text{ GeV} &lt; HT &lt; 70 \text{ GeV}$</td>
<td>LO_MG5</td>
<td>6.1526.508</td>
</tr>
<tr>
<td></td>
<td>HT binned</td>
<td>LO_MG5</td>
<td>16.009.976</td>
</tr>
<tr>
<td></td>
<td>LO samples</td>
<td>LO_MG5</td>
<td>16.325.34</td>
</tr>
<tr>
<td></td>
<td>$100 \text{ GeV} &lt; HT &lt; 200 \text{ GeV}$</td>
<td>LO_MG5</td>
<td>4.365.97</td>
</tr>
<tr>
<td></td>
<td>$200 \text{ GeV} &lt; HT &lt; 400 \text{ GeV}$</td>
<td>LO_MG5</td>
<td>59.366</td>
</tr>
<tr>
<td></td>
<td>$400 \text{ GeV} &lt; HT &lt; 600 \text{ GeV}$</td>
<td>LO_MG5</td>
<td>14.626</td>
</tr>
<tr>
<td></td>
<td>$800 \text{ GeV} &lt; HT &lt; 1200 \text{ GeV}$</td>
<td>LO_MG5</td>
<td>6.677</td>
</tr>
<tr>
<td></td>
<td>$1200 \text{ GeV} &lt; HT &lt; 2500 \text{ GeV}$</td>
<td>LO_MG5</td>
<td>1.613</td>
</tr>
<tr>
<td></td>
<td>$2500 \text{ GeV} &lt; HT &lt; \infty$</td>
<td>LO_MG5</td>
<td>0.039</td>
</tr>
<tr>
<td>$\tau$</td>
<td>DiBoson</td>
<td>POWHEG</td>
<td>8.31756</td>
</tr>
<tr>
<td></td>
<td>WW</td>
<td>PYTHIA8</td>
<td>6.321</td>
</tr>
<tr>
<td></td>
<td>WZ</td>
<td>PYTHIA8</td>
<td>2.282</td>
</tr>
<tr>
<td></td>
<td>ZZ</td>
<td>PYTHIA8</td>
<td>3.132</td>
</tr>
<tr>
<td>QCD multijet</td>
<td>$50 \text{ GeV} &lt; HT &lt; 100 \text{ GeV}$</td>
<td>NLO_MG5</td>
<td>2.4670000.0</td>
</tr>
<tr>
<td></td>
<td>HT binned</td>
<td>NLO_MG5</td>
<td>2.8080000.0</td>
</tr>
<tr>
<td></td>
<td>NLO samples</td>
<td>NLO_MG5</td>
<td>1.7120000.0</td>
</tr>
<tr>
<td></td>
<td>$300 \text{ GeV} &lt; HT &lt; 500 \text{ GeV}$</td>
<td>NLO_MG5</td>
<td>3.47500.0</td>
</tr>
<tr>
<td></td>
<td>$500 \text{ GeV} &lt; HT &lt; 700 \text{ GeV}$</td>
<td>NLO_MG5</td>
<td>3.2100.0</td>
</tr>
<tr>
<td></td>
<td>$700 \text{ GeV} &lt; HT &lt; 1000 \text{ GeV}$</td>
<td>NLO_MG5</td>
<td>6.823.0</td>
</tr>
<tr>
<td></td>
<td>$1000 \text{ GeV} &lt; HT &lt; 1500 \text{ GeV}$</td>
<td>NLO_MG5</td>
<td>12.80</td>
</tr>
<tr>
<td></td>
<td>$1500 \text{ GeV} &lt; HT &lt; 2000 \text{ GeV}$</td>
<td>NLO_MG5</td>
<td>12.00</td>
</tr>
<tr>
<td></td>
<td>$2000 \text{ GeV} &lt; HT &lt; \infty$</td>
<td>NLO_MG5</td>
<td>25.3</td>
</tr>
</tbody>
</table>

Table 5.2: Simulated Samples used for this analysis.

### 5.3 Event Selection

In the previous sections, the datasets presented were skimmed through a trigger selection, where the real data events passed the $HLT\_DoubleMediumIsoPFTau35\_Trk1\_eta2p1$ trigger and the simulated samples were weighted (see section 5.1). An additional selection of events was performed with these datasets, with the purpose to identify the $Z' \rightarrow \tau_h \tau_h$ signature; this selection was split in two categories: the $\tau_h$ identification and the topological selections.
5.3.1 $\tau_h$ Identification

In this analysis the hadronic taus were reconstructed using the HPS algorithm, seeded by jets reconstructed with the anti-kT algorithm and whose transverse momenta are higher than 30 GeV. Additionally, discriminators are applied in order to distinguish the $\tau_h$ signature from those produced by QCD-jets, electrons, and muons (see Section 3.7.2). The $\tau_h$-identification is crucial to reduce the QCD-jet background, which is the dominant SM contamination of the $Z' \to \tau_h \tau_h$ sample.

<table>
<thead>
<tr>
<th>Process (mass point [GeV])</th>
<th>MC generator</th>
<th>width [GeV]</th>
<th>cross-section (pb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSM $Z'$ (500)</td>
<td>PYTHIA8</td>
<td>14.73</td>
<td>5.75100</td>
</tr>
<tr>
<td>SSM $Z'$ (1000)</td>
<td>PYTHIA8</td>
<td>30.97</td>
<td>0.38650</td>
</tr>
<tr>
<td>SSM $Z'$ (1500)</td>
<td>PYTHIA8</td>
<td>47.05</td>
<td>0.06479</td>
</tr>
<tr>
<td>SSM $Z'$ (1750)</td>
<td>PYTHIA8</td>
<td>55.07</td>
<td>0.03104</td>
</tr>
<tr>
<td>SSM $Z'$ (2000)</td>
<td>PYTHIA8</td>
<td>63.10</td>
<td>0.01583</td>
</tr>
<tr>
<td>SSM $Z'$ (2500)</td>
<td>PYTHIA8</td>
<td>79.16</td>
<td>0.00468</td>
</tr>
<tr>
<td>SSM $Z'$ (3000)</td>
<td>PYTHIA8</td>
<td>95.24</td>
<td>0.00159</td>
</tr>
<tr>
<td>SSM $Z'$ (3500)</td>
<td>PYTHIA8</td>
<td>111.30</td>
<td>0.00059</td>
</tr>
</tbody>
</table>

Table 5.3: Simulated Signal Samples with Production Campaign: RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1

Events were required to have taus with 1or3-prongs, whose decay mode is reconstructed using the newDMF discriminator. We used 1or3-prongs selection since it represents 95% of the hadronic tau decays (see Table 3.3); additionally, the unphysical 2-prong decay mode was not included, since it produced a decrease on the sensitivity of the analysis (Appendix B). Figure 5.2 shows the n-prongs for the leading $\tau$ after all tau requirements. On the other hand, the selection of the newDMF discriminator was motivated by the agreement between the data and the level of background extracted from a control region dominated by the Drell-Yan; this control region, used to validate the background simulation samples, will be described in section 5.5.2. The comparison with the other discriminator (oldDMF) is presented in Appendix B.

Figure 5.2: n-prong distribution, normalized to unity, for the leading tau after all tau requirements. QCD multijet background was estimated via data-driven method (see Section 5.5.1)
Taus are required to have a transverse momentum greater than 70 GeV for two reasons: first, this high $p_T$ selection allows to reduce the QCD-contamination, which is expected to be considerable at low-$p_T$ values; second, because for this $p_T$ range, a constant 95% trigger efficiency is obtained (see Section 5.1). Figure 5.3 shows the $p_T$ distributions for the leading tau and the second tau after applying the tau identification requirements. These taus are also required to be in the pseudorapidity region of $|\eta| < 2.1$, due to the high efficiency of the charged-particle track reconstruction in the barrel region (see Figure 3.1), which is not only crucial for the tau reconstruction but also for the tau isolation. This pseudorapidity range is covered by the single tau L1-trigger acceptance. Figure 5.4 shows the $\eta$ distributions for the leading tau and the second tau, after requiring the tau identification criteria; note that the background is spread over all $\eta$-range, while the $\eta$-distribution for the signal is centered at $\eta = 0$.

In order to distinguish the $\tau_h$ candidate from QCD-jets, the tight WP of the MVA-based isolation discriminator was required (byTightIsolationMVArun2v1DBnewDMwLT); this WP ensures a $\tau$ identification efficiency of $\sim$60% and a QCD-jet rejection rate of $\sim$99.8%. For the discrimination against electrons, a loose WP was required (againstElectronMVALooseMVA6) in order to keep a high tau reconstruction efficiency (85%) and a relatively low misidentification rate ($10^{-2}$). Additionally, the $\tau_h$ candidates are required to pass the tight WP of the against-muons cutoff-based discriminator (againstMuonTight3), whose $\tau$ identification efficiency is more than 99%, while the misidentification probability is $1.4 \times 10^{-3}$. The impact on the sensitivity due to the WP of the discriminators against QCD-jets, electrons and muons is presented in Appendix B. The $\tau_h$ identification criteria, used in this analysis, are summarized in Table 5.4.

Figure 5.3: $p_T$ distributions, normalized to unity, for the leading tau (right) and the second tau (left), after all tau requirements. QCD multijet background was estimated via data-driven method (see Section 5.5.1)
Figure 5.4: $\eta$ distributions, normalized to unity, for the leading tau (right) and the second tau (left), after all tau requirements. QCD multijet background was estimated via data-driven method (see Section 5.5.1).

| $p_T$ | $> 70$ GeV |
| $|\eta|$ | $< 2.1$ |
| Tau Decay | 1or3-prongs |
| newDMF |
| Isolation Disc. | byTightIsolationMVArun2v1DBnewDMwLT |
| Anti-electron Disc. | againstElectronMVALooseMVA6 |
| Anti-muon Disc. | againstMuonTight3 |

Table 5.4: Tau Identification Criteria

5.3.2 Topological Selections

Since the $Z'$ boson is expected to be massive, the two oppositely-charged taus would be moving in opposite directions. Therefore, the $\Delta R(\tau_1, \tau_2) > 0.3$ requirement is applied to ensure that the hadronic tau pair is well separated in the $\eta - \phi$ plane.

Additionally, as was mentioned in Section 4.1, due to presence of neutrinos, a missing transverse energy ($E_T$) is expected in the signal events. Figure 5.6 shows the $E_T$ distribution after applying the tau identification criteria. As can be seen in the figure, an $E_T$ selection will reduce considerably the QCD multijet contamination. A requirement of $E_T$ above 30 GeV will reduce the QCD multijet background by 57%, while it keeps 94% of the signal yield for a $Z'$ mass point of 3 TeV. Additionally, this selection allows to reduce the W+Jets contamination by 27%, since no momentum imbalance is expected for this background. As will be discussed in Section 5.5.1, this requirement is also motivated by the fact that the $E_T < 30$ GeV region is highly contaminated by QCD multijet events, which can be exploited to estimate this background.
Figure 5.5: $\Delta R(\tau_1, \tau_2)$ distribution, normalized to unity, for the ditau pair after all tau identification requirements. QCD multijet background was estimated via data-driven method (see Section 5.5.1).

Figure 5.6: $E_T$ distribution, normalized to unity, in linear (left) and log (right) scale after all tau identification requirements. QCD multijet background was estimated via data-driven method (see Section 5.5.1).

The b-jet signature can fake the one produced by a high-$p_T$ tau and, in consequence, a b-jet veto is applied on the events. The veto is applied to any jet tagged as b-jet using the loose WP of the CSVv2 algorithm, which provides a reconstruction efficiency of 83% (see Section 3.5). This veto allows to reduce the $t\bar{t}$ background since, as mentioned in Section 4.2, the top quark most of the times decays into a bottom quark and a W. The b-jet veto reduces by 74% the $t\bar{t}$ background, while the signal efficiency is kept at 90% for a Z' with a mass of 3 TeV. Figure 5.7
shows the distribution of the number of b-jets after requiring the tau identification criteria and the $E_T$ selection.

Figure 5.7: Distribution of number of jets tagged as b-jets, normalized to unity, in linear (left) and log (right) scale, after the tau identification and $E_T$ requirements. QCD multijet background was estimated via data-driven method (see Section 5.5.1).

Since both taus coming from the $Z'$ decay are expected to be back-to-back in the transverse plane, the $\cos \Delta \phi (\tau_1, \tau_2) < -0.95$ selection is required in the events. Figure 5.8 shows the $\cos \Delta \phi (\tau_1, \tau_2)$ distribution after applying the tau identification, the $E_T$, and the b-jet veto selections. Since the production of the two taus is uncorrelated for the QCD multijet and W+jets backgrounds, this requirement reduces those backgrounds by 58\% and 24\%, respectively, while 81\% of the signal events are kept for a $Z'$ mass of 3 TeV.

The $E_T$ of the signal events is expected to be collinear with the direction of the taus. Most of the times, the sum of the transverse momenta of the tau pair will be aligned with the leading tau and, therefore, the $E_T$ would point towards the less energetic tau. As a result, for the signal events, there is a strong correlation between the $E_T$ and the direction of the visible decay products of the tau; in the case of the QCD multijet and W+jets backgrounds, there is not such correlation, since the jet faking the tau signature is produced in a random direction. Consequently, the $\cos \Delta \phi (\tau_{lead}, E_T) < -0.9$ selection, where $\tau_{lead}$ refers to the tau with the greater $p_T$, is required in order to reduce the background contamination of processes such as QCD multijet and W+jets (67\% and 59\%, respectively), while keeping 86\% of the signal events for a $Z'$ mass of 3 TeV. Figure 5.9 shows the $\cos \Delta \phi (\tau_{lead}, E_T)$ distribution after applying the requirements mentioned above (tau identification, $E_T$, b-jet veto and $\cos \Delta \phi (\tau_1, \tau_2)$ selections). In previous analyses, other selections have been applied in order to consider the strong correlation for the signal events between the $E_T$ and the direction of the visible decay products of the tau, such as the so-called pZeta selection, which was used for the previous version of this analysis using 2015 data \cite{[19]}. For the 2016 data, the highest sensitivity is reached with the $\cos \Delta \phi (\tau_{lead}, E_T) < -0.9$ requirement. The study of the sensitivity reached for different selections is described in Appendix C.
Figure 5.8: $\cos \Delta \phi(\tau_1, \tau_2)$ distribution, normalized to unity, in linear (left) and log (right) scale, after requiring the tau identification, $E_T$, and b-jet veto criteria. QCD multijet background was estimated via data-driven method (see Section 5.5.1).

Figure 5.9: $\cos \Delta \phi(\tau_{lead}, E_T)$ distribution, normalized to unity, in linear (left) and log (right) scale, after requiring the tau identification, $E_T$, and b-jet veto criteria. QCD multijet background was estimated via data-driven method (see Section 5.5.1).

5.3.3 Summary
The tau identification criteria and the topological requirements described in the previous sections allow to maintain a high selection efficiency for signal events, providing a strong background suppression, while keeping a reduced influence of systematic effects. The final selection criteria used in this analysis, known as signal region, is presented in Table 5.5.
The relative efficiency, as well as the cumulative efficiency of each topological requirement, for the signal and the backgrounds, are presented in Table 5.6. As can be noted, the ditau trigger is only required on real data, while it is not applied on the simulated samples in which the efficiency of the trigger selection is 100%. Considering that the signal events were generated for two taus in the final state (independently if they decay leptonically or hadronically), no preselection requirements of two hadronic taus were applied on the signal events and, therefore, the values in the table absorb such selection. As can be seen, these topological selections reduce the contamination of backgrounds mainly from \( t\bar{t} \), W+Jets and diboson processes to negligible levels.

<table>
<thead>
<tr>
<th>Trigger Selection</th>
<th>HLT_DoubleMediumIsoPFTau35_Trk1_eta2p1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_{pT} )</td>
<td>( &gt; 70 \text{ GeV} )</td>
</tr>
<tr>
<td>(</td>
<td>\eta</td>
</tr>
<tr>
<td>Tau Decay</td>
<td>( 1\nu3\text{-prongs} )</td>
</tr>
<tr>
<td>Isolation Disc.</td>
<td>byTightIsolationMVArun2v1DBnewDMwLT</td>
</tr>
<tr>
<td>Anti-electron Disc.</td>
<td>( \text{againstElectronMVA_LooseMVA6} )</td>
</tr>
<tr>
<td>Anti-muon Disc.</td>
<td>( \text{againstMuon_Tight3} )</td>
</tr>
</tbody>
</table>

\( \mathcal{E}_T \)

\( > 30 \text{ GeV} \)

\# b-jets (CSVv2 Loose WP) = 0

\( \Delta R(\tau_1, \tau_2) \)

\( > 0.3 \)

\( \cos \Delta \phi(\tau_1, \tau_2) \)

\( < -0.95 \)

\( \cos \Delta \phi(\tau_{\text{end}}, \mathcal{E}_T) \)

\( < -0.90 \)

Table 5.5: Signal Region. The trigger applied only on real data and a proper correction is applied on simulated samples.

<table>
<thead>
<tr>
<th>Out</th>
<th>QCD</th>
<th>DY</th>
<th>W_Jets</th>
<th>( t\bar{t} )</th>
<th>Diffbench</th>
<th>( Z ) [1 TeV]</th>
<th>( Z ) [2 TeV]</th>
<th>( Z ) [3 TeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \epsilon )</td>
<td>( \epsilon_C )</td>
<td>( \epsilon )</td>
<td>( \epsilon_C )</td>
<td>( \epsilon )</td>
<td>( \epsilon_C )</td>
<td>( \epsilon )</td>
<td>( \epsilon_C )</td>
<td>( \epsilon )</td>
</tr>
<tr>
<td>Trigger</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>TauID 1</td>
<td>0.8</td>
<td>0.9</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>TauID 2</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>( \mathcal{E}_T )</td>
<td>47.3</td>
<td>47.3</td>
<td>47.3</td>
<td>47.3</td>
<td>47.3</td>
<td>47.3</td>
<td>47.3</td>
<td>47.3</td>
</tr>
<tr>
<td>( \Delta \Delta \phi(\tau_2, \tau_2) )</td>
<td>88.0</td>
<td>88.0</td>
<td>88.0</td>
<td>88.0</td>
<td>88.0</td>
<td>88.0</td>
<td>88.0</td>
<td>88.0</td>
</tr>
<tr>
<td>( \cos \Delta \phi(\tau_{\text{end}}, \mathcal{E}_T) )</td>
<td>33.4</td>
<td>33.4</td>
<td>33.4</td>
<td>33.4</td>
<td>33.4</td>
<td>33.4</td>
<td>33.4</td>
<td>33.4</td>
</tr>
</tbody>
</table>

Table 5.6: Efficiency (\( \epsilon \)) and cumulative efficiency (\( \epsilon_C \)) for each selection of the signal region.

The ditau trigger is only required on real data, while it is not applied on simulated samples whose the efficiency is 100%. No preselections of hadronic taus are required.

### 5.3.4 Validation Plots for Signal Selections

The performance of the signal selections is validated using (N-1) distributions. An (N-1) distribution corresponds to the case when all the signal region selections are applied with exception of one of them; this allows to enhance the background contribution relevant to the dropped selection in order to probe the level of accuracy between the simulated samples and the data. For instance, the validation of the \( \mathcal{E}_T \) selection is performed checking the data and simulated samples agreement for the \( \mathcal{E}_T \) distribution, after applying all the signal selection criteria with exception of this requirement. Figure 5.10 shows some (N-1) plots for the signal region using the simulated samples described in Table 5.2. As can be seen in the figure, QCD multijet simulated samples have limited statistics with considerable uncertainties, which motivates a data-driven estimation
of this background. Figure 5.11 presents the (N-1) plots using a data-driven method for the QCD multijet background estimation (see section 5.5.1). As expected, the (N-1) distributions are dominated by the QCD multijet background; the good agreement between the background and data validates the QCD multijet data-driven estimation, as well as the simulated samples used.

Figure 5.10: Distribution of variables after all requirements for the signal region, with the exception of the one plotted: $E_T$ (top left), number of b-jets (top right), $\cos \Delta \phi(\tau_1, \tau_2)$ (bottom left), $\cos \Delta \phi(\tau_{lead}, E_T)$ (bottom right). The QCD multijet background estimation was performed using the simulated samples listed in Table 5.2.
Figure 5.11: Distribution of variables after all requirements for the signal region, with the exception of the one plotted: $E_T$ (top left), number of b-jets (top right), $\cos \Delta \phi (\tau_1, \tau_2)$ (bottom left), $\cos \Delta \phi (\tau_{lead}, E_T)$ (bottom right). The QCD multijet background has been estimated with the data-driven method described in section 5.5.1.
5.4 Corrections for Simulated Events

Although the event simulations developed by the particle physics community have been extensively studied and upgraded for years, there are still differences between the simulated events of SM processes and the observed data. Therefore, corrections must be applied on the simulated events in order to minimize such differences and, consequently, in order to improve the accuracy of expected yield, for both, backgrounds and signal simulated samples. The corrections applied are described in this section.

**Pile-Up corrections**

The official simulated samples, certified by the CMS Collaboration, usually are produced at the same time, or before, than the data-taking period. Considering that the pileup interactions depend on the LHC operational conditions, the simulated samples are generated including the effect of the expected pileup interactions during the data-taking period; however, exactly the LHC operational parameters are not necessarily the same used for the event generation. As a result, the simulated samples have different pileup distributions than those of real data. In consequence, the simulated events must be weighted properly in order to reproduce the pileup distribution observed in data.

The method employed by the CMS Collaboration in order to measure the number of pileup interactions in the data consists on multiplying the instantaneous luminosity, for a single bunch crossing, by the total inelastic cross section. This method is reliable since the instantaneous luminosity for a single BX can be computed from the LHC operational parameters, and the total inelastic cross-section, that has been measured. As a result, a certified JSON file is issued, containing the pileup information for the entire data-taking period. Thus, the probability to obtain \( n \) interactions in a data event \( (P_{\text{data}}(n)) \) can be estimated.

The pileup weight applied on a simulated event is given by:

\[
w(n) = \frac{P_{\text{data}}(n)}{P_{\text{MC}}(n)}
\]

where \( P_{\text{MC}}(n) \) is the probability to obtain \( n \) interactions in the simulated event. The pileup weights were applied using the certified JSON file (minimum bias cross-section of 69.2 mb).

**Corrections of tau identification and triggering**

The efficiency of the tau identification algorithm must be taken into account when the expected simulated events are estimated. A comparison between data and SM processes is used in order to obtain the Data/MC tau identification scale factor (TauID. SF) to correct the MC prediction; such comparison is performed using Z and W events, that decay into taus. The TauID SF depends on the isolation working point. Since in this analysis the tight WP was used, the TauID. SF applied for a single hadronic tau is 0.95 \[84\]; in consequence, the overall scale factor, due to the identification of two hadronic taus, that was applied was 0.90 \(\sim\) 0.95 \(\times\) 0.95.

As mentioned in section \[5.1\] the trigger was not applied on simulated samples due to the difficulty of modeling it. Instead, its efficiency was measured from the data, and used in the simulated events as a weight, which is dependent on the tau-\( p_T \). The applied trigger weight, per hadronic tau, is described by the equation \[5.1\].
Figure 5.12: Distribution of variables after applying the signal selection criteria but inverting the $\cos \Delta \phi(\tau_1, \tau_2)$ requirement. QCD multijet background was estimated using the data-driven method, while other backgrounds were estimated directly from simulated samples. The uncertainty is based on statistics. $p_T$ of the leading tau (top left), $p_T$ of the second tau (top right), $E_T$ (middle left), $m(\tau_1, \tau_2)$ (middle right), $\cos \Delta \phi(\tau_1, \tau_2)$ (bottom left) and $\Delta \phi(\tau_{lead}, E_T)$ (bottom right).
Note that the TauID, SF and the trigger weight were estimated per hadronic tau, and their data/MC agreement for ditau final states were validated simultaneously using the Drell-Yan control region (Figure 5.19).

5.5 Background Estimation

The main source of background for the dihadronic tau final state is dominated by the probability of misidentifying QCD-jets as taus, where dijets can fake the $Z' \rightarrow \tau_h \tau_h$ signature; this source accounts for 70% of the total background. As can be noted in Figure 5.10, the QCD multijet simulated samples do not have enough statistics in order to estimate properly the expected SM contamination due to these processes. Additionally, the large uncertainties in treating the processes of fragmentation and hadronization that involved in the modeling of QCD-jet production, as well as the uncertainties of the pile-up effects and the interaction of the jet products with the detector material, make QCD multijet simulated simulation not reliable. Therefore, the QCD multijet background contribution must be determined using data-driven methods, based on the creation of enriched control samples obtained by modifying the signal selection criteria.

Another important source of background comes from Drell-Yan processes, approximately 23%, which have the same topology than the signal. Since these SM processes have been extensively studied, two high-quality hadronic taus are expected to be well modeled by simulation and, therefore, an MC based approach is employed to estimate their contribution to the signal region, instead of a data-driven method. In order to determine the level of accuracy of the simulated backgrounds, an enriched $Z \rightarrow \tau_h \tau_h$ control sample is defined, where no considerable fluctuations between data and backgrounds are expected. This control sample also allows to validate the tau identification criteria used in this analysis.

5.5.1 Background Estimation for QCD

The QCD multijet contamination expected in the signal region is estimated using the so-called “ABCD method”, where the QCD multijet yield ($N_{QCD}^{SR}$), as well as the shape and the normalization in the final distributions, are obtained from data. The basic idea is to estimate $N_{QCD}^{SR}$ from a control sample dominated by QCD multijet events: in this case, the sample is obtained requiring the same signal criteria but selecting $\tau_h \tau_h$ pairs with like-sign (LS) electric charge. Assuming that the electric charges of both QCD-jets are not correlated, the yield of the like-sign QCD multijet events ($N_{QCD}^{LS}$) should be the same as the one of oppositely-charged QCD multijet (OS) events. However, this assumption cannot be true for events where the charges of the quarks, which give rise to the QCD multijets, are correlated. The quark charge can be correlated with the charge of the leading track of the jet; this correlation is especially strong for jets with low multiplicity of high-momentum tracks, which most likely can fake the tau signature. In consequence, if the two quarks are kinematically correlated, for instance through $q\bar{q}$ production, that can result in an oppositely-charged QCD multijet event, that could fake a $\tau^+\tau^-$ event. This leads to an asymmetry between $N_{QCD}^{SR}$ and $N_{QCD}^{SL}$, and, therefore, a normalization factor must be included in the $N_{QCD}^{SR}$ estimation. Thus, the QCD multijet yield in the signal region is given by:

$$N_{QCD}^{SR} = R_{OSLS} \times N_{QCD}^{LS},$$

(5.5)

where $R_{OSLS}$ is the normalization factor, which represents the $N_{OS}^{QCD} / N_{LS}^{QCD}$ asymmetry.
In order to estimate the OS/LS ratio, other enriched QCD control regions are defined. As was mentioned in the previous section (see Figure 5.6), the $E_T < 30$ GeV region is highly contaminated by QCD events and, therefore, it can be used to measure the OS/LS ratio; with this purpose, the $E_T < 30$ GeV sideband is split into two control regions: one composed by oppositely-charged dijet events and, the other one, by like-sign dijet events. Figure 5.13 shows the overview of the data-driven “ABCD method” used to estimate the QCD multijet contribution in the signal region.

Figure 5.13: Overview of the data-driven method to estimate the QCD multijet background.

In summary, the shape of the $m(\tau_h, \tau_h, E_T)$ distribution in the signal region is obtained from the control region C, which consists in like-sign dijet events and $E_T > 30$ GeV. The distribution is normalized to the contribution of oppositely-charged QCD multijet, using the OS/LS ratio. The ratio is measured using control regions B and D ($E_T < 30$ GeV sideband) as follows:

$$R_{OSLS} = \frac{N_{QCD}^B}{N_{QCD}^D}. \quad (5.6)$$

As was mentioned before, control regions B, C and D are dominated by QCD multijet events and simulated samples cannot be used to predict this background. Then, any disagreement between data and the other expected SM processes in these control regions is assigned to QCD multijet events. This means that the QCD multijet estimation is obtained from data by subtracting any contribution coming from non-QCD processes, such as DY, W+Jets, DiBoson and $t\bar{t}$. In consequence:

$$N_{QCD}^B = N_{Data}^B - N_{nonQCD}^B,$$
$$N_{QCD}^C = N_{Data}^C - N_{nonQCD}^C,$$
$$N_{QCD}^D = N_{Data}^D - N_{nonQCD}^D. \quad (5.7)$$

Thus, in order to measure the OS/LS ratio using the equation 5.6, the QCD multijet estimation was performed in control regions B and D, considering the equations 5.7. Table 5.7 shows the data and the SM simulated background yields in these controls regions. Note that the QCD multijet estimation presented in the table corresponds to the $Data - nonQCD$ row. The purity of the QCD multijet samples obtained in the control regions B and D, defined as $(Data - nonQCD)/Data$, is 85% and 99%, respectively. On the other hand, the contamination
from signal in these control regions is less than 0.01% since any \(Z'\) boson events are expected to have \(E_T\) greater than 30 GeV. Since a high QCD multijet purity and low signal contamination is obtained, these control regions are appropriated to estimate the \(OS/LS\) ratio. The \(OS/LS\) ratio resulted from this procedure was 1.34 ± 0.12. Figure 5.14 shows that the measured \(OS/LS\) ratio can be considered as a constant in the effective mass spectrum. A relative systematic uncertainty of 9% is assigned on the QCD multijet estimation due to the measurement of the \(OS/LS\) transfer factor.

The normalization and the determination of the \(m(\tau_h, \tau_h, E_T)\) shape must be tested in order to validate the QCD multijet estimation in the signal region. Nevertheless, since the QCD multijet simulated samples do not have enough statistics, the validation must be performed using real data. Then, a test is performed to check simultaneously the proper normalization (usually known as closure test) and to validate if the \(m(\tau_h, \tau_h, E_T)\) shape in the opposite-sign regions can be modeled correctly by like-sign dihadronic tau events. In order to perform the shape closure/validation test, the QCD multijet background is estimated in the control region B from the \(m(\tau_h, \tau_h, E_T)\) shape obtained in the control region D (like-sign region), and the correct normalization is obtained using the \(OS/LS\) ratio.

<table>
<thead>
<tr>
<th></th>
<th>CR B Yields</th>
<th>CR D Yields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>405.0 ± 20.1</td>
<td>261.0 ± 16.2</td>
</tr>
<tr>
<td>DY</td>
<td>46.9 ± 6.8</td>
<td>0.1 ± 0.1</td>
</tr>
<tr>
<td>WJets</td>
<td>10.8 ± 6.3</td>
<td>1.3 ± 0.9</td>
</tr>
<tr>
<td>DiBoson</td>
<td>0.6 ± 0.4</td>
<td>0.2 ± 0.1</td>
</tr>
<tr>
<td>(tt)</td>
<td>0.6 ± 0.4</td>
<td>0.3 ± 0.3</td>
</tr>
<tr>
<td>nonQCD BKG</td>
<td>58.8 ± 9.3</td>
<td>261.0 ± 16.2</td>
</tr>
<tr>
<td>Data - nonQCD</td>
<td>346.2 ± 22.2</td>
<td>259.1 ± 16.2</td>
</tr>
<tr>
<td>TOTAL BKG</td>
<td>405.0 ± 24.0</td>
<td>261.0 ± 16.2</td>
</tr>
<tr>
<td>(R_{OSLS})</td>
<td></td>
<td>1.34 ± 0.12</td>
</tr>
</tbody>
</table>

Table 5.7: Yields in the controls region B and D used for calculation of \(R_{OSLS}\) ratio.

Figure 5.14: The measured \(OS/LS\) ratio in function of the effective mass.
The agreement between the data and the expected background in the effective mass spectrum determines if the QCD multijet $m(\tau_h, \tau_h, E_T)$-shape in the oppositely-charged taus region can be well modeled from the like-sign region. In case there would be a disagreement, a systematic uncertainty on the shape could be included in the final results. As can be seen in Figure 5.15, there is a good data/background agreement and, therefore, no systematic uncertainties were applied due to the shape of the QCD multijet estimation. In consequence, the only systematic uncertainty considered on the QCD multijet estimation comes from the relative uncertainty of the OS/LS ratio measurement. Additionally, the data/background agreement in the control region B implies that the same method can be used to estimate QCD multijet background in the signal region.

Figure 5.15: $m(\tau_h, \tau_h, E_T)$ distribution in OS tau-pairs with low-$E_T$ (control region B): Normal scale (left), and log scale (right). QCD multijet background was estimated from SS events in the low-$E_T$ sideband (control region D), normalized to QCD multijet estimation in control region B using the $R_{OSLS}$ ratio. Only the statistical uncertainties have been included.

Table 5.8 shows the amount of events coming from each background contamination in the signal region, as well as the yield of data and background in the control region C. The QCD multijet background in the signal region is estimated using the same method described above, which consists of taking the $m(\tau_h, \tau_h, E_T)$ shape from the like-sign dihadronic tau events ($Data^C - nonQCD^C)$ and normalizing it using the OS/LS ratio. The table also shows the expected signal events in the signal region, as well as the contamination due to the signal in the control region C (1.6%). Table 5.9 shows the QCD multijet yield estimated in each control region. The final prediction of this background in the signal region is $382.67 \pm 24.84$ (right-most column of the table), where the uncertainty is fully statistical. The QCD multijet purity in the signal region, defined as $(Data^C - nonQCD^C)/Data^C$, is 85% for effective masses above 600 GeV (see Figure 5.16), where the signal is expected.
Table 5.8: Signal and background yields in signal region and control region C.

The other backgrounds in Table 5.8 were estimated from simulated samples. As was mentioned before, the estimation of Drell-Yan processes, which represents around 23% of the total background in the signal region, is validated using a $Z \rightarrow \tau\tau$ control region (see next section). The remaining processes ($W$+Jets, $t\bar{t}$ and DiBoson), which represent only 8% of the background, are also estimated from simulated samples.

Table 5.9: QCD multijet yields in signal region and in control regions B, C, and D.

Figure 5.16: QCD multijet purity in the mass spectra of the signal region.
Figure 5.17 shows the \( m(\tau_h, \tau_h, E_T) \) distribution in the signal region. This distribution represents the most important plot for this analysis, since if a \( Z' \) exists it would show up as a broad resonance in the high effective mass spectrum, as can be seen in the figure. Then, once the background estimation performed is validated (see Section 5.5.3), the signal selection criteria are applied on real data samples and any excess in the \( m(\tau_h, \tau_h, E_T) \) distribution over the SM expectation could represent the existence of a \( Z' \) boson.

Figure 5.18 shows other distributions obtained after applying the signal selection criteria (where the QCD multijet background was estimated using the data-driven method).

5.5.2 Background Estimation for Drell-Yan events

The Drell-Yan processes represent 23% of the total background and their estimation is performed using simulated samples. These samples are generally reliable since these kind of processes have been studied extensively. A control region dominated by this background is useful to validate the simulated modeling as well as the tau identification used in this analysis. In such control region a good agreement between data and background is expected. However, the definition of an appropriate control region that allows to validate the simulated samples is not straightforward, considering that the dihadronic tau final state is highly contaminated by QCD multijet background, which must be estimated via data-driven methods. However, a semi-clean sample with two high-quality hadronic taus can be obtained requiring a ditau invariant mass less than 100 GeV. Table 5.10 shows the selection criteria used to define the enriched \( Z \to \tau_h \tau_h \) control sample. As can be noted in the table, the selections \( E_T > 30 \text{ GeV}, \cos \Delta \phi(\tau_1, \tau_2) < -0.95, \) and \( \cos \Delta \phi(\tau_{lead}, E_T) < -0.9 \) from the signal region were removed since no events were obtained when the \( m(\tau_1, \tau_1) < 100 \text{ GeV} \) selection was added to the signal selection criteria; in consequence, removing the requirements listed above is equivalent to invert them, which ensures that the Drell-Yan control region is orthogonal to the signal region.
Figure 5.18: Distribution of variables after all requirements for the signal region. The QCD multijet background was estimated from like-sign dihadronic taus events in the nominal-\(E_T\) sideband, normalized with the \(R_{OS/LS}\) ratio. The other backgrounds were estimated directly from simulated samples. The uncertainty is only statistical. The variables plotted are: \(p_T\) of the leading tau (top left), \(p_T\) of the second tau (top right), \(E_T\) (middle left), number of b-jets (middle right), \(\cos \Delta\phi(\tau_1, \tau_2)\) (bottom left) and \(\Delta\phi(\tau_{lead}, E_T)\) (bottom right).
Table 5.10: Selection criteria used for the Drell-Yan control region.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tau-p$_T$</td>
<td>&gt; 70 GeV</td>
</tr>
<tr>
<td>$</td>
<td>\eta</td>
</tr>
<tr>
<td>Tau Decay</td>
<td>1or3-prongs</td>
</tr>
<tr>
<td>Isolation Disc.</td>
<td>newDMF</td>
</tr>
<tr>
<td>Anti-electron Disc.</td>
<td>byTightIsolationMVArun2v1DBnewDMwLT</td>
</tr>
<tr>
<td>Anti-muon Disc.</td>
<td>againstElectronMVAlooseMVA6</td>
</tr>
<tr>
<td># b-jets (CSVv2 Loose WP)</td>
<td>= 0</td>
</tr>
<tr>
<td>$Ch_{\tau_1} \times Ch_{\tau_2}$</td>
<td>&lt; 0</td>
</tr>
<tr>
<td>$\Delta R(\tau_1, \tau_2)$</td>
<td>&gt; 0.3</td>
</tr>
<tr>
<td>$m(\tau_1, \tau_1)$</td>
<td>&lt; 100 GeV</td>
</tr>
</tbody>
</table>

Table 5.11 shows the yields for the $Z \rightarrow \tau_h \tau_h$ validation sample, where the QCD multijet contribution has been determined using the data-driven method discussed in the previous section. Note that a good agreement between the data and the expected overall yield of the SM is obtained since the $Data/MC$ ratio is $1.06 \pm 0.09$. Figure 5.19 shows the dihadronic tau invariant mass distribution, where the shape is consistent with the SM prediction. In conclusion, since the $Data/MC$ ratio and the dihadronic tau invariant mass shape are consistent with the SM expectation, the Drell-Yan processes, as well as the tau identification, are well modeled by the simulated samples. A relative systematic uncertainty of of 8% is considered on the Drell-Yan estimation due to any disagreement in data and the expected background in this control region.

<table>
<thead>
<tr>
<th></th>
<th>DY CR Yields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>431.00 ± 20.76</td>
</tr>
<tr>
<td>DY</td>
<td>346.2 ± 25.1</td>
</tr>
<tr>
<td>WJets</td>
<td>3.3 ± 1.0</td>
</tr>
<tr>
<td>DiBoson</td>
<td>17.4 ± 1.4</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>17.4 ± 1.4</td>
</tr>
<tr>
<td>QCD</td>
<td>37.9 ± 8.5</td>
</tr>
<tr>
<td>TOTAL BKG</td>
<td>406.5 ± 26.5</td>
</tr>
<tr>
<td>Data/MC</td>
<td>1.1 ± 0.1</td>
</tr>
</tbody>
</table>

Table 5.11: Background and data yields in $Z \rightarrow \tau_h \tau_h$ region obtained using the DY selection criteria.

### 5.5.3 Validation

Besides the (N-1) distributions shown in Figure 5.11, other way to validate the background estimation is inverting the $\cos \Delta \phi(\tau_1, \tau_2)$ selection; since in such region the signal contamination is minimum and there is still enough statistics for background estimation. Then, the validation region is defined by:

- $-0.95 < \cos \Delta \phi(\tau_1, \tau_2) < 1.0$
- $E_T > 30$ GeV
- $\Delta \phi(\tau_{lead}, E_T) < -0.9$
- no jets with $p_T$ greater than 30 GeV tagged as b-jets
Figure 5.12 shows distributions of variables after applying the selection criteria listed above. The agreement between the data and the simulated samples confirms the proper estimation of the background; any disagreement between them is covered by the uncertainties.

Figure 5.19: $m(\tau_h, \tau_h)$ distribution for the region obtained using the DY selection criteria. Linear scale (left), log scale (right). Only the statistical uncertainties have been included.

5.6 Systematic Uncertainties

In order to confirm or discard the existence of a $Z'$ boson using the effective visible mass distribution, it is necessary to consider systematic uncertainties. These uncertainties can be caused by the finite resolution of the detector and the simulations involved in the analysis. There are two types of systematic uncertainties: the first type is related with the normalization of the distributions, for instance, the uncertainty in the luminosity measurement; while the second type is related to the shape of the distributions, for instance, the uncertainties on the measurement of the variables needed to determine the $m(\tau_h, \tau_h, E_T)$ distribution.

The different sources of systematic uncertainties considered in this analysis are:

- Luminosity,
- Tau Identification,
- Tau Trigger,
- Tau Energy Scale,
- b-jet Identification,
- Jet Energy Scale,
- Missing Transverse Energy,
- Parton Distribution Functions,
- bin-by-bin statistical uncertainty.
- Background Estimations.

**Luminosity**

Since the luminosity is used to scale the simulated samples (backgrounds and signal), the uncertainty on its measurement must be considered. The CMS Luminosity group has assigned a 2.5% uncertainty on the integrated luminosity of the data collected during 2016 [100].

**Tau Identification**

This uncertainty is associated to the tau identification algorithm. In this analysis, the HPS algorithm was used to identify taus. The efficiency can be affected by several factors such as: the efficiency of the track finding algorithm; the efficiency of identifying the tau decay mode; the probability of tracks associated to the underlying event, falling in the isolation cone; and the probability of pions to fall outside the signal cone. The CMS group dedicated to study the efficiency and reconstruction of the hadronic tau (Tau Particle Object Group, TauPOG) measured the overall systematic uncertainty and assigned to it a value of 5% (see Ref. [101]). Since in this analysis there are two correlated taus in the final state, an overall uncertainty of 10% was assigned on the simulated samples due to the ditau identification. The TauPOG included an additional uncertainty related to the high-$p_T$ tau identification efficiency, which affects the sensitivity of this analysis. This is an asymmetric uncertainty assigned per hadronic tau, which depends on its momentum; the uncertainty on the tau identification goes from $-35\%$ up to $5\%$ of the tau momentum (see Ref. [101]); this result in an uncertainty of the order of $10\%$ for masses around 500 GeV and of the order of $25\%$ for masses around 2 TeV. The uncertainty on the high-$p_T$ tau identification affects the shape of the $m(\tau_h, \tau_h, E_T)$ distribution; this is determined by applying the $p_T$ dependent weights, per tau, and extracting varied mass templates, which are included as shape systematics.

**Tau Trigger**

Another source of uncertainty comes from the modeling of the tau trigger at simulation level (see Section 5.1.1). In order to model the trigger for each hadronic tau, weights were calculated using the tag-and-probe method. Due to the inefficiency observed for taus whose $p_T$ is greater than 70 GeV, a systematic uncertainty of $5\%$ is assigned per hadronic tau. The overall trigger uncertainty, when considering both taus, is $10\%$.

**The Tau Energy Scale**

The tau energy scale is calculated, using simulated data, as the ratio between the energy of the reconstructed tau and the energy of the visible decay products of the generated tau. It represents the detector response to the hadronic tau. The precise measurement of the 4-momentum of the tau decay components is important for this analysis since the effective visible mass is computed from them and from the missing transverse energy. There are systematic uncertainties inherent to the experimental identification of taus, that must be considered in the final result of the
analysis. There are two sources of uncertainty: the first one comes from the measurement of energy clusters and the track reconstruction of the individual tau decay products; the second one comes from the fact that it is not possible to know whether all the tau visible decay products were included or not (for example, neutral pions which can fall outside the signal cone). Studies performed by the TauPOG conclude that no corrections must be applied on the tau energy measurement and that an overall tau energy scale uncertainty of 3\%, independently of the tau decay mode, must be applied \cite{101}. This systematic effect results in a 3\% uncertainty in the signal yield and up to $\sim11\%$ in the background yields.

**b-jet Identification**

Another source of uncertainty comes from the b-tagging identification algorithm used in this analysis. A 30\% uncertainty on the b-jet mis-tagging rate was measured by the CMS b-tagging group (see Ref. \cite{77}). Since in this analysis there was a b-jet veto, this uncertainty must be taken into account. This was done using the equation:

$$\epsilon^N_{b-tag} < 1 = 1 - \sum_{n=1} P(n) \cdot \sum_{m=1}^{n} C(n, m) \cdot f^m \cdot (1 - f)^{n-m},$$

(5.8)

where the second term on the right side represents the efficiency of identifying at least one b-jet in the event, $n$ is the number of jets, $P(n)$ is the probability to obtain $n$ jets in a event, $f$ is the probability that a jet can be misidentified as a b-jet and $C(n, m)$ are the combinatorial factors.

The systematic uncertainty due to the b-jet veto requirement was estimated, using the equation, considering the misidentification rate at the Loose CSVv2 working point (measured by the CMS b-tagging group) together with the probability to obtain at least one b-jet in an event ($\sim10\%$). The resulting uncertainty is $\sim5\%$ on the signal samples. The same uncertainty is assigned to the background samples where b-jets are not expected in the event, such as Drell-Yan and diboson samples. For backgrounds such as $W+$Jets, where at least one real jet is expected, and $t\bar{t}$, where b-jets are expected, this uncertainty is 10\%.

**Jet Energy Scale**

The jet reconstruction and, in consequence, its energy measurement have many sources of uncertainties such as the non-linear calorimeter response, the pileup interactions, the underlying event, the electronic noise, etc. Therefore, corrections are applied on the jet reconstruction, known as Jet Energy Corrections (JEC), to account for all these effects, allowing an improved match between the simulated predictions and the data (see section 3.4). The Jet Energy Scale (JES) corrections are part of the JEC. As recommended by the CMS JetMET POG, a 3\% or 5\% uncertainty (depending on the $\eta$ and $p_T$ of the jet) must be included in order to take into account these systematic effects \cite{102}. This uncertainty makes the signal and simulated based backgrounds fluctuate up to $\sim12\%$.

**Missing Transverse Energy**

Since the $E_T$ is inferred from the 4-momenta of all the reconstructed final states in the event, its systematic uncertainty depends on the topology of the signal process. The uncertainty due to the $E_T$ on the signal acceptance depends mainly on the tau energy scale (TES), the jet energy scale (JES) and the unclustered energy (UCE). The uncertainty on the UCE is 10\%, as recommended by the CMS JetMET POG. Considering the uncertainties coming from UCE, TES and
JES, one gets a systematic uncertainty of 0.5% due to the $E_T$ on the simulated background yields.

**Parton Distribution Functions**

Another source of systematic uncertainty on the simulated samples comes from the imprecise knowledge of the parton distribution functions (PDF) used in the event simulation. These uncertainties were estimated following the Run II recommendations given by the LHC group dedicated to study the PDFs [103]. These uncertainties were estimated with the 68% confidence level using the “PDF4LHC15_mc” sets of PDFs. As a result, the PDF uncertainties for the simulated samples used are smaller than their bin-by-bin statistical uncertainties and, therefore, they are negligible for the background simulated samples. In the case of the signal, the PDF uncertainties varies from 0.7% for the $Z'$ sample with a mass of 500 GeV up to 12% for the $Z'$ sample with a mass of 3 TeV.

**bin-by-bin statistical uncertainty**

The finite statistics of the simulated samples must be included as an additional uncertainty. With this purpose, a per-bin uncertainty is estimated by changing the yield of a given bin, within its statistical uncertainty, while keeping the yields of the other bins unchanged, and evaluating its effect on the shape of the $m(\tau_\mu, \tau_\mu, E_T)$ distribution.

**Background Estimations**

As described in section 5.5.1, the QCD multijet background contamination was estimated using a data-driven method. Since, in order to estimate and subtract this background, simulated samples are used, the systematic effects mentioned above must be considered. With this purpose, these uncertainties are propagated throughout the subtraction.

Besides, a systematic uncertainty is assigned on the QCD multijet estimation due to the measurement of the $OS/LS$ transfer factor used to derive the contamination of this background in the signal region. Since the measured $OS/LS$ ratio is $1.34 \pm 0.12$, a relative uncertainty of 9% is considered on the QCD multijet background estimation. In the case of the Drell-Yan background, an uncertainty of 8% is assigned due to the effect of the ratio between the data and the expected background measured in the Drell-Yan control region (see section 5.5.2). In the case of the diboson background, it is not possible to obtain a semi-clean enriched sample using the dihadronic final state, due to the high QCD multijet contamination; therefore, this sample was defined using the $Z' \rightarrow \tau_e \tau_\mu$ channel. As a result, the data/MC scale factor obtained was $1.01 \pm 0.20$ and, consequently, a 20% systematic uncertainty was assigned on the diboson background [104]. These systematic uncertainties, due to the measurement of data/MC scale factors or the measurement of transfer factors, are known as closure and normalization uncertainties. In the case of W+Jets and $t\bar{t}$ backgrounds, these sources of uncertainty are not considered since no scale factors nor transfer factors were used for their estimation.

**Summary of the applied systematic uncertainties**

Table 5.12 shows the values of systematic uncertainties considered for the signal and backgrounds used in this analysis. In order to estimate these systematic uncertainties, the analysis was performed considering the upward and downward variation for each source of uncertainty, and taking the maximum difference with respect to the nominal value. For instance, in the case of
the 3% uncertainty assigned to the TES, the tau 4-momentum was scaled by a factor of 1.03 and all the variables were recalculated; as a result, the background contribution coming from Drell-Yan, W+Jets and $t\bar{t}$ processes varied up to 11%, as can be noted in the table. The most important source of systematic uncertainty is the high-p$_T$ tau identification which, as mentioned above, corresponds to 25% uncertainty in the case of a reconstructed effective visible mass of 2 TeV, affecting considerably the sensitivity of the search.

<table>
<thead>
<tr>
<th>Source</th>
<th>QCD</th>
<th>DY</th>
<th>W+Jets</th>
<th>$t\bar{t}$</th>
<th>DiBoson</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lumi</td>
<td>−</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>$\tau_h$ Trig</td>
<td>−</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>$\tau_h$ ID</td>
<td>−</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>high-p$_T$ $\tau_h$ ID</td>
<td>−</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
</tr>
<tr>
<td>TES</td>
<td>−</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>b ID</td>
<td>−</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>JES</td>
<td>−</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>pdf</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>(1-12)</td>
</tr>
<tr>
<td>bin-by-bin stat.</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
</tr>
<tr>
<td>Closure+N+Norm.</td>
<td>9</td>
<td>8</td>
<td>−</td>
<td>−</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.12: Summary of systematic uncertainties. Values are given in percent. “s” indicates template variations (“shape” uncertainties). L = 2.5%.

5.7 Summary

All the features involved in the search for $Z'$ bosons in the dihadronic tau final channel were presented in this chapter. The trigger selection for two-hadronic tau events was described in section 5.1. The data collected by the CMS experiment from the pp collisions at $\sqrt{s} = 13$ TeV, delivered by the LHC during 2016, along with the simulated samples used for this search, were presented in section 5.2. The event selection carried out to optimize the signal selection and to reduce any background contribution, minimizing the systematic uncertainties, was outlined in section 5.3. The background estimation in the signal region (Section 5.5), considering the corrections on the simulated samples, was performed and was validated using the N-1 distributions (section 5.3.4) and the inverted cos $\Delta\phi(\tau_h, \tau_h)$ control region (section 5.5.3). Finally, all the systematic uncertainties that affect the signal and background estimations are listed in section 5.6. As a result, the estimations of the backgrounds and of the expected signal are reliable (Table 5.8) and, therefore, any excess observed in the $m(\tau_h, \tau_h, E_T)$ distribution (Figure 5.17) with real data, would represent an evidence of the existence of a $Z'$ boson. The $Z' \rightarrow \tau\tau$ search is model-independent since it is based on the observation of an excess in the high spectrum of the effective visible mass distribution. Consequently, the search for a massive resonance can be sensitive to ditau final states predicted by other theories and, therefore, the results can be reinterpreted according to an specific model. In the next chapter, the complete analysis, assuming the nonexistence of a $Z'$ boson is presented and exclusion limits are estimated.
6 Analysis and Conclusions

This chapter presents the results and conclusions of the search for $Z'$ bosons in the dihadronic tau channel, using proton-proton collisions at a centre-of-mass energy of 13 TeV, with the data collected by the CMS experiment during 2016, which have 35.9 fb$^{-1}$ of integrated luminosity. This analysis is currently unblinded, i.e., the CMS Collaboration has approved to explore with the real data the region where the $Z'\rightarrow \tau\tau$ is expected. In order to avoid any bias in the analysis and to ensure objectivity, the CMS Collaboration does not allow to explore the signal region with real data until all the procedures have been validated; this means that the whole procedure described in Chapter 5, as well as the procedures performed for the other ditau channels ($\tau_{h}\tau_{h}$, $\tau_{h}\tau_{e}$ and $\tau_{e}\tau_{e}$), have already been approved by the collaboration. However, the results using real data are still being reviewed internally and, therefore, they are not included in this document. Instead, expected exclusion limits have been calculated (see Section 6.1), showing that the sensitivity in the dihadronic tau channel has improved over the one obtained with the combination of all four channels, using the data collected by CMS during 2015.

6.1 Analysis

The exclusion limit calculation is based on the assumption that any $Z'$ existence is discarded. This is achieved by generating pseudo-data samples, which are constructed with the background-only hypothesis, and performing the calculation as if they were real data. The quantity of interest in order to set the exclusion limit is $\sigma \cdot B$, where $\sigma$ is the cross-section for the $Z'$ production ($pp\rightarrow Z'$) and $B$ is the branching ratio of this boson decaying into taus ($Z'\rightarrow \tau\tau$); indeed, this quantity is proportional to the expected number of signal events, and it can be compared with the “observed” events (using the pseudo-data). The compatibility of the “observed” data and the signal expectation on the effective visible mass distribution, which provides the best discrimination between signal and backgrounds, is quantified using a modified frequentist approach, known as CL$_s$ method [105,106]. This statistical method, as well as the expected exclusion limits calculation, are presented in this section.

6.1.1 Statistical Method

With the purpose to characterize the non-observation of the signal and to set the exclusion limit, one defines the null hypothesis $H_0$, which describes the signal-plus-background processes. This is tested against an alternative hypothesis $H_1$, which is based on the background-only assumption. In order to quantify the compatibility of the “observed” data with a given hypothesis, there are two approaches: the Bayesian and the frequentist approaches. They differ in the way in which the probability is defined. For the Bayesian approach, the probability can be interpreted as the “degree of belief” of the result for a given experiment. For the frequentist approach, which is used in this analysis, it can be interpreted as the probability to obtain a given measurement after repeating the experiment many times.

In the particle physics community the frequentist approach, based on a binned likelihood ratio, is widely used to establish, or to exclude, the existence of a new phenomenon. In this analysis, the
calculation of the exclusion limit is obtained by using each bin of the \( m(\tau_h, \tau_h, E_T) \) distribution in order to evaluate the likelihood per bin. Their combination results in the upper limit on the cross-section, for a given \( Z' \) mass point, where the level of compatibility of the “observed” data with a given hypothesis is quantified by the confidence level (CL). Since it is a convention, 95\% CL for excluding the signal is used in this analysis.

In the following, the signal and background yields will be denoted as \( s \) and \( b \), respectively. Their predictions are subject to systematic uncertainties (see Section 5.6) that can be included in the exclusion limit calculation as nuisance parameters \( \theta \). Then, the signal and background are functions of the nuisance parameters \( \theta \).

The likelihood function is given by:

\[
L(\theta) = \prod_{i=1}^{N-\text{bins}} L_i(s_i, b_i, n_i; \theta),
\]

(6.1)

where \( \theta \) is the set of nuisance parameters and \( L_i(s_i, b_i, n_i; \theta) \) is the likelihood function for the \( i \)-th bin, which represents the probability function for obtaining an “observed” data yield in the \( i \)-th bin given an hypothesis \( H \). Since the number of independent results is large and the cross-section for a signal event is very low, compared with the cross-section for background events, the probability density is a Poisson distribution:

\[
L_i(s_i, b_i, n_i; \theta) = \frac{(s_i + b_i)^{n_i}}{n_i!} e^{s_i + b_i},
\]

(6.2)

then, \( L_i(s_i, b_i, n_i; \theta) \) is the Poisson probability of observing \( n_i \) events in data in the \( i \)-th bin, given the hypothesis \( H \). As a result, \( L_s+b(\theta') \) is the product of \( N \) Poisson probabilities of observing \( n_i \) events in the data, where \( s_i + b_i \) events are expected (\( N \) is the number of bins). \( L_b(\theta) \) is the product of \( N \) Poisson probabilities of observing \( n_i \) events in data, where only background events are expected.

These likelihood functions are the result of a single measurement of signal and background yields in the \( m(\tau_h, \tau_h, E_T) \) distribution; therefore, \( s_i \) and \( b_i \) are the mean values, in the \( i \)-th bin, of the distribution of all possible \( s_i \) and \( b_i \) values that could have been obtained if the experiment would have been repeated several times. Since repeating the experiment several times is not possible, an ensemble of experiments is simulated. This ensemble is obtained generating pseudo-data samples with the background-only hypothesis; each experiment of this ensemble provides an independent new possible result. As a consequence, a probability distribution is obtained for the \( s_i \) and \( b_i \) yields, allowing to consider any fluctuation in data and, resulting in an improvement on the sensitivity of the statistical method.

The CLs method tests the null hypothesis (signal plus backgrounds) against the background-only hypothesis through the likelihood ratio (LR). The LR is the best discriminator between both hypotheses, and it is given by [105]:

\[
LR = -2 \ln \frac{L_{s+b}(\theta'')}{L_b(\theta')}.
\]

(6.3)

From the equation it is possible to infer that “observed” events with \( LR > 0 \) are more compatible with the background-only hypothesis \( H_0 \) than with the signal-plus-background hypothesis \( H_0 \). Here \( \theta'' \) and \( \theta' \) denote the value of the systematic uncertainties that maximizes the likelihood for \( H_0 \) and \( H_1 \) respectively. This procedure is known as binned maximum likelihood for
the signal-plus-background and background only hypotheses. The dependence on the nuisance parameters is reflected as a broaden likelihood distribution, compared with the one if they were fixed. This represents a loss of sensitivity in the analysis due to the systematic uncertainties [105].

As mentioned in section 5.6, the systematic uncertainties can affect the normalization or the shape of the $m(\tau_h, \tau_h, E_T)$ distribution. They are included in the likelihood calculation through an MC numerical integration method over all the nuisance parameters. In the case when a systematic uncertainty affects the normalization, such as the one of the tau trigger, the nuisance parameters are generated with a logarithmic normal probability density function. In the case when a systematic uncertainty affects the shape, such as the one of the high-$p_T$ tau identification, the nuisance parameters are generated with a Gaussian probability density function for the $m(\tau_h, \tau_h, E_T)$ spectrum uncertainty.

Once the systematic uncertainties are included, the likelihood distribution (equation 6.1) is used to obtain the 95% CL upper limit for the signal cross-section. The confidence level of the upper limit, using the CL$_s$ method is given by [105]:

$$CL_s = \frac{CL_{s+b}}{CL_b},$$

(6.4)

where $CL_{s+b}$ quantifies the compatibility of the “observed” data with the signal-plus-background hypothesis and $CL_b$ quantifies the compatibility of the “observed” data with the background-only hypothesis. In this method the quantity $CL_s$ must be less or equal than 0.05 in order to set the 95% CL in the exclusion limit calculation.

This procedure results in the exclusion limit calculation at a given CL for an specific $Z' \to \tau\tau$ channel. This procedure was performed for the dihadronic tau channel, as well as for the other channels: $\tau_h\tau_\mu$, $\tau_h\tau_e$ and $\tau_\mu\tau_e$. Therefore, in order to set an exclusion limit for the search for $Z'$ bosons in the ditau final state the results must be combined. Even though a high sensitivity is achieved in this analysis, a considerable improvement on the sensitivity is achieved by combining the results of the four ditau channels. This can be done by computing a total binned likelihood:

$$L_{tot} = L(\tau_h, \tau_h) \times L(\tau_h, \tau_\mu) \times L(\tau_h, \tau_e) \times L(\tau_\mu, \tau_e).$$

(6.5)

6.1.2 Exclusion Limit Calculation

An expected exclusion limit at 95% CL is set on $\sigma(pp \to Z') \cdot B(Z' \to \tau\tau)$ using the method described above. In summary, the calculation of the exclusion limit is performed using the $m(\tau_h, \tau_h, E_T)$ distribution to construct the binned maximum likelihood for the signal-plus-background and background-only hypotheses. The binned likelihood is the product of the Poisson probabilities (equation [6.2]). The systematic uncertainties are included in the calculation considering a logarithmic normal probability function for normalization parameters, and a Gaussian probability function for mass-spectrum shape uncertainties. Finally, the exclusion limit calculation is performed with the modified frequentist approach, known as the CL$_s$ method, which uses the likelihood ratio discriminator (equation [6.3]) in order to set the limits on the signal cross-section at the level of accuracy desired (equation [6.4]). Note that the results with real data are not public yet, and instead the limits were calculated by treating pseudo-data, based on background-only hypothesis, as if they were real data.

The statistical procedure described above is performed using the CMS Higgs limit calculation tool Combine [107]. The input of the tool are the so-called data cards, which include the total
yields and the $m(\tau_h, \tau_h, E_T)$ distributions for signal and backgrounds, as well as their systematic uncertainties.

As mentioned above, the exclusion limit calculation was performed based on the background and the signal expectations using the effective visible mass distribution. The final prediction of the backgrounds and the signal (obtained in Chapter 5) in the whole effective visible mass region, as well as the prediction for masses above 600 GeV, are shown in Table 6.1. As can be noted, the QCD multijet processes are the dominant background, which correspond to a 67% of the total SM contamination for the whole mass spectrum, while the Drell-Yan processes represent a 23% of the total background. The next greater background, for the whole mass region, comes from W+Jets events which contribute 7%, while the remaining ones (diboson and $t\bar{t}$) only represent 3% of the total. Moreover, the $m(\tau_h, \tau_h, E_T) > 600$ GeV region, where the $Z'$ events are expected, is highly contaminated by QCD multijet and Drell-Yan events which contribute to the total background a 48% and 40%, respectively; in this region, the remaining backgrounds represent 12% of the total contamination: W+Jets (7%) and diboson (5%).

<table>
<thead>
<tr>
<th></th>
<th>Overall Yield</th>
<th>Yield for $m(\tau_h, \tau_h, E_T) &gt; 600$ GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>DY</td>
<td>131.6 ± 8.6</td>
<td>11.5 ± 1.0</td>
</tr>
<tr>
<td>WJets</td>
<td>41.9 ± 5.2</td>
<td>2.0 ± 0.8</td>
</tr>
<tr>
<td>DiBoson</td>
<td>3.7 ± 1.1</td>
<td>1.4 ± 0.7</td>
</tr>
<tr>
<td>TTbar</td>
<td>4.4 ± 1.2</td>
<td>0.0 ± 0.0</td>
</tr>
<tr>
<td>QCD</td>
<td>382.7 ± 24.8</td>
<td>13.7 ± 4.7</td>
</tr>
<tr>
<td>TOTAL BKG</td>
<td>564.3 ± 26.9</td>
<td>28.6 ± 4.9</td>
</tr>
<tr>
<td>$Z'$ (1 TeV)</td>
<td>465.5 ± 4.4</td>
<td>371.4 ± 3.9</td>
</tr>
<tr>
<td>$Z'$ (2 TeV)</td>
<td>19.6 ± 0.2</td>
<td>19.1 ± 0.2</td>
</tr>
<tr>
<td>$Z'$ (3 TeV)</td>
<td>1.9 ± 0.0</td>
<td>1.9 ± 0.0</td>
</tr>
</tbody>
</table>

Table 6.1: Overall and $m(\tau_h, \tau_h, E_T) > 600$ GeV yields for signal and backgrounds in signal region.

The $Z'$ search is model-independent since it is based on the observation of an excess in the high spectrum of the effective visible mass distribution; this means that the signal acceptance estimation presented in Chapter 5 is independent of the model. If a signal evidence were found, the signal cross-section could point towards a particular BSM scenarios. Similarly, the exclusion limit calculation on the signal cross-section depends on the model. Exclusion limits are calculated for the $Z'_{SSM}$ boson, considering the cross-section presented in Table 5.3 and for the $Z'_{TAT}$ boson, considering that its cross-section is about one third of the one of the SSM model. As a result of the exclusion limit calculation performed for this analysis, we expect to exclude $Z'_{SSM}$ and $Z'_{TAT}$ bosons (decaying into two hadronic taus) for masses below 2.7 TeV and 2.2 TeV, in the case in which no signal is observed.

Figure 6.1 shows the expected limits (dash line) and the leading order theoretical cross-section for $Z'_{SSM}$ and $Z'_{TAT}$ (red and blue lines), in the dihadronic tau channel. The exclusion limits are presented as a function of the effective visible mass. The bands represent the 1σ and 2σ deviations from the expected limits. The bands were obtained using the pseudo-data samples, based on the background-only hypothesis, described in the previous section. Note that the exclusion limit on the $Z'$ mass is determined at the point in which the expected limit on the cross-section exceeds
the theoretical one. For completeness, the expected exclusion limits for the other channels are presented in Figure 6.2. The expected exclusion limits for the $Z'_\text{SSM}$ boson in the $\tau_h\tau_\mu$, $\tau_h\tau_\tau$ and $\tau_\tau\tau_\mu$ channels are 2.69 TeV, 2.6 TeV and 1.75 TeV, respectively [35].

![Figure 6.1](image_url)

Figure 6.1: Expected exclusion limits at the 95% CL on $\sigma(pp \rightarrow Z') \cdot B(Z' \rightarrow \tau\tau)$, for the $\tau_h\tau_h$ channel, as a function of the $Z'$ mass. The result is compared with the leading order theoretical expectations for the SSM and TAT models.

### 6.2 Analysis of Results

As already mentioned, the search for $Z'$ bosons in the dihadronic tau channel, using the data collected by CMS during 2016, has set expected exclusion limits for the $Z'_\text{SSM}$ and the $Z'_\text{TAT}$ for masses below 2.7 TeV and 2.2 TeV, respectively. This section presents a comparison of the results with the ones obtained in the other ditau channels (see Section 6.2.1). Additionally, the results are compared with the most recent searches for $Z'$ bosons performed by CMS and ATLAS (see Sections 6.2.2 and 6.2.3).

#### 6.2.1 Comparison with Other Channels

Figures 6.1 and 6.2 show the expected exclusion limit, calculated for the dihadronic tau channel and for the other ditau final states, using the data collected by CMS during 2016. Note that the expected exclusion limit for the $Z'_\text{SSM}$ mass in the $\tau_h\tau_h$ channel (2.7 TeV) is similar to the ones obtained for the $\tau_h\tau_\mu$ and $\tau_h\tau_\tau$ channels (2.65 TeV and 2.6 TeV). The results are similar since, even though the dihadronic tau channel has the highest branching ratio (42%) among the ditau channels, it also has highest QCD multijet contamination that lowers its sensitivity. Since in the dihadronic case, the tau sample is contaminated by QCD-jets, it makes the $\tau_h\tau_h$ signal less clean than the $\tau_h\tau_\tau$. Another reason that affects the sensitivity of the dihadronic tau channel is the systematic uncertainty associated to the identification of high-$p_T$ taus; this source of systematics represents $\sim$25% uncertainty for a $Z'$ mass of 2 TeV, while it represents $\sim$10% uncertainty on
Figure 6.2: Expected exclusion limits at the 95% CL on $\sigma(pp \rightarrow Z') \cdot B(Z' \rightarrow \tau\tau)$ as a function of the $Z'$ mass. The result is compared with the leading order theoretical expectations for the SSM and TAT models. The figure shows the results for the $\tau_h\tau_h$ (top left), $\tau_h\tau_e$ (top right) and $\tau_e\tau_{\mu}$ (bottom) channels.

the same $Z'$ mass in the case of the $\tau_h\tau_{\ell}$ channels. Note that, although the $\tau_e\tau_{\mu}$ final state is the cleanest channel, due to the low contamination of QCD processes and the high efficiency of the light lepton reconstruction, the search in this channel sets the lowest exclusion limit due to its small branching ratio (6.2%).

6.2.2 Comparison with previous CMS Searches

In Chapter 1 (Table 1.3) the results of previous searches performed by CMS were presented. An exclusion limit of 1.4 TeV for $Z'_{\text{SSM}}$ was set with the data at 7 TeV/8 TeV (Run I). During Run II, at 13 TeV, initial exclusion limits were set for $Z'_{\text{SSM}}$ and $Z'_{\text{TAT}}$ of 2.1 TeV and 1.7 TeV, using the data collected during 2015 (2.2 fb$^{-1}$) and the four ditau channels. In the particular case of the $\tau_h\tau_h$ channel, the exclusion limits observed with this data, were 1.92 TeV for $Z'_{\text{SSM}}$ and 1.51 TeV for $Z'_{\text{TAT}}$. Figure 6.3 shows the observed exclusion limits, for this channel (left) in comparison with the expected limits obtained with the 2016 data (this work). Note that the sensitivity in the dihadronic tau channel has been improved for the 2016 search since tighter exclusion limits are expected (2.7 TeV for the $Z'_{\text{SSM}}$ mass and 2.2 TeV for the $Z'_{\text{TAT}}$ mass).
The main difference between the searches for $Z'$ performed by CMS in 2015 and 2016 is the integrated luminosity: 2.2 fb$^{-1}$ for the 2015 run and 35.9 fb$^{-1}$ for the 2016 run. In the particular case of the $\tau_h \tau_h$ channel, the main differences, besides luminosity, were the exclusion of the so-called $pZeta$ requirement (that was used for the 2015 case and it is defined in Appendix C) and the inclusion of the $\cos \Delta \phi (\tau_{lead}, E_T)$ selection for the 2016 analysis. The $\cos \Delta \phi (\tau_{lead}, E_T)$ requirement improved the significance (see Appendix C), which is reflected in tighter exclusion limits. Figure 6.4 shows the comparison between the exclusion limits observed, combining the four ditau channels, with the 2015 data (left), and the expected exclusion limits obtained, in the dihadronic tau channel, with the 2016 data (right). The figure shows that the sensitivity in the dihadronic tau channel, using the 2016 data, is even higher that the one obtained with the combination of all four ditau channels, using the 2015 data.

![Figure 6.3: Exclusion limits observed in the search for $Z' \rightarrow \tau_h \tau_h$, using the data collected by CMS during 2015 (left) and 2016 (right).](image)

![Figure 6.4: Exclusion limits observed combining the four $Z' \rightarrow \tau\tau$ channels, using the data collected by CMS during 2015 (left). Expected exclusion limits obtained in the search for $Z' \rightarrow \tau_h \tau_h$, using the data collected by CMS during 2016 (right).](image)
6.2.3 Comparison with Results from ATLAS

The search for Z’ bosons decaying into taus performed by the ATLAS Collaboration, using the data collected during 2015 and 2016, with an integrated luminosity 36.1 fb⁻¹, excluded the Z’SSM for masses below 2.42 TeV at 95% CL (see Section 1.3.2). The search was performed for the $\tau_h\tau_h$, $\tau_h\tau_\mu$, and $\tau_h\tau_\tau$ channels [25]. In this analysis, there were degradations at lower masses due to the $\tau_h$-p_T thresholds and at higher masses due to the $\tau_h$ reconstruction and identification efficiency, which decreases at high p_T. Therefore, the results of this work show that if no signal is observed in our analysis in CMS, we would obtain significantly higher exclusion limits (see Figure 6.5).

![Figure 6.5: Exclusion limits observed combining the four $Z' \rightarrow \tau\tau$ channels, using the data collected by ATLAS during 2015 and 2016 (left) [25]. Expected exclusion limits obtained in the search for $Z' \rightarrow \tau_h\tau_h$, using the data collected by CMS during 2016 (right).](image)

6.3 Conclusions

- In this work a search for Z’ bosons decaying into two hadronic taus was performed. The search used the data collected by CMS during 2016 of proton-proton collisions at a centre-of-mass energy of 13 TeV with an integrated luminosity of 35.9 fb⁻¹.

- Since at this stage (end of May of 2018) no real data has been used yet only expected exclusion limits can be quoted.

- The expected exclusion limits obtained in this work are: $m(\tau_h, \tau_h, E_T) > 2.7$ TeV for Z’SSM at 95% CL; $m(\tau_h, \tau_h, E_T) > 2.2$ TeV for Z’TAT at 95% CL.

- The analysis presented in this work has already been approved by the CMS Collaboration and, therefore, real data can be used to search for a signal or set observed exclusion limits. This analysis is being executed and the results should be available for publication very soon (one or two months).
• The expected exclusion limits obtained in this work are significantly higher than those obtained during 2015 data and those obtained by ATLAS during 2015 and 2016 data. Therefore, if no signal is observed, as already announce by the ATLAS Collaboration, this analysis will represent an improvement on the previous ones.

• The higher sensitivity of this analysis was due mostly to improvements in the selection criteria used.

• The Tau identification criteria used in this analysis is slightly different from the one recommended by the CMS TauPOG. However, the results are statistically equivalent.

• The expected exclusion limits quoted in this work corresponds to generic theoretical models (SSM and TAT). However, since the kinetic part of the analysis is model independent, it would suggest that for any $Z'$ masses below 2.2 TeV are most likely excluded.

• This analysis is a general search for massive resonances in the ditau channel. Consequently, it can be sensitive to ditau final states foreseen by other theories and the results can be reinterpreted according to a specific model.
A Trigger Studies

As mentioned in section 5.1 for this analysis the trigger selection was not applied on simulated samples. Instead, it was modeled as a weight, which was obtained from the trigger efficiency measured with data. The TauPOG (the CMS group dedicated to the tau identification criteria) recommends another method to perform the trigger selection. The method consists in selecting the simulated events that fire the HLT_DoubleMediumIsoPFTau35_Trk1_3e2p1 trigger and to apply Data/MC scale factors on them. These scale factor were measured, by the TauPOG, to match the trigger efficiency curves obtained with simulated samples with those obtained with real data. The trigger scale factors depend on the tau-p_T and the decay mode.

The search for $Z' \rightarrow \tau_h \tau_h$ was performed using both trigger selection methods (the one used in the analysis compared with the one recommended by the TauPOG) in order to study their impact on the sensitivity. Table A.1 shows the overall yields obtained with both methods; the difference between them, for the most considerable backgrounds (QCD and DY), is less than 2%, which is within the systematic uncertainty associated with the trigger selection. Figure A.1 shows the significance obtained with both methods as a function of the effective visible mass. As can be noted in the figure, the significances on the high effective visible mass spectrum, where the signal is expected, are pretty similar (the difference is ~3%). This means that, any trigger selection method can be used since there is not a considerable impact on the sensitivity of the search. In order to be consistent with the other di-tau channels, the trigger was not applied on simulated samples.

<table>
<thead>
<tr>
<th>OVERALL YIELDS</th>
<th>No trigger applied</th>
<th>TauPOG recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DY</td>
<td>131.6 ± 8.6</td>
<td>124.1 ± 8.2</td>
</tr>
<tr>
<td>WJets</td>
<td>41.9 ± 5.2</td>
<td>46.9 ± 5.5</td>
</tr>
<tr>
<td>DiBoson</td>
<td>3.7 ± 1.1</td>
<td>4.0 ± 1.1</td>
</tr>
<tr>
<td>tt</td>
<td>4.4 ± 1.2</td>
<td>5.2 ± 1.4</td>
</tr>
<tr>
<td>QCD</td>
<td>382.7 ± 24.8</td>
<td>381.9 ± 25.0</td>
</tr>
<tr>
<td>TOTAL BKG</td>
<td>564.3 ± 26.9</td>
<td>562.1 ± 26.9</td>
</tr>
<tr>
<td>$Z'$ (3 TeV)</td>
<td>1.9 ± 0.0</td>
<td>1.9 ± 0.0</td>
</tr>
<tr>
<td>Significance</td>
<td>0.34</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Table A.1: Signal and background yields obtained with trigger selection method used in this analysis and with the one recommended by the TauPOG. The significance was computed for $m(\tau_h, \tau_h, E_T) > 600$GeV.

†The trigger efficiency curves are described by a Cristal Ball function [108], where the parameters of those fits depends on the tau-p_T and on the tau decay mode [109].
Figure A.1: Significance, as a function of the effective visible mass, obtained with trigger selection method used in this analysis and with the one recommended by the TauPOG.
B Tau Identification Studies

Several studies were performed on the tau identification criteria in order to optimize the signal acceptance and to reduce any background contribution. The $\tau_h$ identification criteria are based on:

- the number of prongs (1or3-prongs or 1or2or3-prongs),
- the MVA-based isolation discriminator (loose, medium or tight),
- the discriminator against electrons (loose or very loose),
- the discriminator against muons (loose and tight),
- the decay mode discriminator (newDMF or oldDMF)

In order to study the impact on the sensitivity due to a particular working point (WP), for a given discriminator, the significance is computed as a function of the effective visible mass. The significance is defined as $s/\sqrt{s+b}$, where the signal yield $s$ and the total background $b$ are estimated using the event selection and the background estimation described in Chapter 5. A comparison among the significances obtained for each WP is performed in order determine the optimal selection for each discriminator. The tau identification criteria recommended by the TauPOG, which differ from those used in this analysis, were considered for these studies.

In this analysis, we select taus with 1or3-prongs, whose decay mode is reconstructed using the newDMF discriminator. For the MVA-based isolation discriminator the tight working point was required. Additionally, taus are required to pass the loose WP of the discriminator against electrons and the tight WP for the discrimination against muons. These selection criteria are summarized in Table B.1. Besides of the sensitivity reached in the $Z' \rightarrow \tau_h \tau_h$ channel, the selection of the working points for the discriminators is based on the consistency with the $\tau_h$ identification criteria used in the others di-tau channels.

<table>
<thead>
<tr>
<th>Tau Decay</th>
<th>1or3-prongs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>newDMF</td>
</tr>
<tr>
<td>Isolation Disc.</td>
<td>$by\text{TightIsolationMVArun2v1DBnewDMwLT}$</td>
</tr>
<tr>
<td>Anti-electron Disc.</td>
<td>$againstElectronMVA\text{LooseMVA6}$</td>
</tr>
<tr>
<td>Anti-muon Disc.</td>
<td>$against\text{MuonTight3}$</td>
</tr>
</tbody>
</table>

Table B.1: Tau Identification Criteria.

B.1 Number of prongs Study

The tau identification algorithm considers an unphysical 2-prong final state, which can result from high-$p_T$ taus in the 3-prong decay mode. In this case, due to the finite spatial resolution of the Tracker System, two of the three tracks can be merged and reconstructed as a single charged
hadron, resulting in an apparent 2-prong final state. The signal and background estimation were performed including this unphysical final state in the tau identification criteria (1or2or3-prongs).

Table B.2 shows the overall yields obtained using both selections, 1or3-prongs and 1or2or3-prongs, for the tau identification. Note, that the QCD contamination increases considerably for 1or2or3-prongs case, which produces a decrease on the sensitivity of the analysis. Figure B.1 shows the significance obtained using both criteria, 1or3-prongs and 1or2or3-prongs. Since a better significance is obtained for \( m(\tau_h, \tau_h, E_T) > 600 \) GeV, where the signal is expected, the 1or3-prongs selection was used in this analysis.

<table>
<thead>
<tr>
<th>OVERALL YIELDS</th>
<th>1or3-prongs</th>
<th>1or2or3-prongs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DY</td>
<td>131.6 ± 8.6</td>
<td>152.0 ± 9.1</td>
</tr>
<tr>
<td>WJets</td>
<td>41.9 ± 5.2</td>
<td>65.2 ± 8.4</td>
</tr>
<tr>
<td>DiBoson</td>
<td>3.7 ± 1.1</td>
<td>4.7 ± 1.2</td>
</tr>
<tr>
<td>tt</td>
<td>4.4 ± 1.2</td>
<td>5.7 ± 1.4</td>
</tr>
<tr>
<td>QCD</td>
<td>382.7 ± 24.8</td>
<td>753.1 ± 34.3</td>
</tr>
<tr>
<td>TOTAL BKG</td>
<td>564.3 ± 26.9</td>
<td>980.7 ± 36.5</td>
</tr>
<tr>
<td>( Z' ) (3 TeV)</td>
<td>1.9 ± 0.0</td>
<td>2.4 ± 0.0</td>
</tr>
</tbody>
</table>

Table B.2: Signal and background yields obtained requiring 1or3-prongs and 1or2or3-prongs in the tau identification. The significance was estimated for \( m(\tau_h, \tau_h, E_T) > 600 \) GeV.

Figure B.1: Significance, as a function of the effective visible mass, obtained with 1or3 prongs (blue) and 1or2or3 prongs (red).
B.2 MVA-based Isolation Discriminator Study

The purpose of the MVA-based isolation discriminator is to reduce the QCD background contamination. As mentioned in section 3.7.2, the energy deposits, excluding those coming from the tau decay products, within the isolation cone define three working points (loose, medium and tight). Less energy deposits represents a better discrimination against QCD background, since a QCD-jet has higher multiplicity of particles with wider energy profile than the hadronic tau decay. Table B.3 shows the overall yields obtained using the three WPs of the isolation discriminator. Figure B.2 shows the significance, as a function of the effective visible mass, for each WP. As expected, the tight WP has the highest significance since it reduces considerably the contamination from QCD processes, which are the dominant background in this analysis. The results are consistent with the TauPOG recommendation (tight WP).

<table>
<thead>
<tr>
<th>OVERALL YIELDS</th>
<th>tight</th>
<th>medium</th>
<th>loose</th>
</tr>
</thead>
<tbody>
<tr>
<td>DY</td>
<td>131.61 ± 8.6</td>
<td>150.7 ± 9.3</td>
<td>182.0 ± 10.8</td>
</tr>
<tr>
<td>WJets</td>
<td>41.91 ± 5.2</td>
<td>72.1 ± 6.8</td>
<td>136.7 ± 10.6</td>
</tr>
<tr>
<td>DiBoson</td>
<td>3.73 ± 1.1</td>
<td>4.6 ± 1.2</td>
<td>5.1 ± 1.2</td>
</tr>
<tr>
<td>tt</td>
<td>4.39 ± 1.2</td>
<td>4.7 ± 1.3</td>
<td>6.5 ± 1.5</td>
</tr>
<tr>
<td>QCD</td>
<td>382.67 ± 24.8</td>
<td>957.9 ± 38.9</td>
<td>2300.7 ± 58.2</td>
</tr>
<tr>
<td>TOTAL BKG</td>
<td>564.30 ± 26.9</td>
<td>1190.1 ± 40.6</td>
<td>2630.9 ± 60.2</td>
</tr>
<tr>
<td>(Z^\prime(3\text{ TeV}))</td>
<td>1.91 ± 0.0</td>
<td>2.3 ± 0.0</td>
<td>2.6 ± 0.0</td>
</tr>
</tbody>
</table>

Table B.3: Signal and background yields obtained using the tight, medium and loose WPs for the MVA-based isolation discriminator. The significance was estimated in the region where the signal is expected \(m(\tau_h, \tau_h, E_T) > 600\text{GeV})

B.3 MVA-based against Electron Discriminator Study

In the case of the MVA-based algorithm developed to discriminate electrons from taus, six WPs are defined according to the tau identification efficiency and the misidentification rate required in each analysis (see Section 3.7.2). In this analysis, the loose WP was used in order to obtain a high tau reconstruction efficiency (~83%), while keeping a relative low misidentification rate \(10^{-2}\). However, the TauPOG recommends to use the very loose WP since a higher reconstruction efficiency is achieved (~85%). Table B.4 shows the overall yields obtained using the loose and very loose WPs of the MVA-based against electron discriminator. As can be noted in the table, although there are a higher background contribution for the very loose WP, the signal yield increases due to the higher tau identification efficiency. As a result, the sensitivity of the analysis is improved using the WP recommended by the TauPOG. Nevertheless, the loose WP was used in this analysis in order to keep consistency with the \(\tau_h\) identification criteria used in the others di-tau channels. Figure B.3 shows the significance obtained using the loose and very loose WPs of the MVA-based against electron discriminator.

B.4 Cutoff-based against Muon Discriminator Study

Due to the low misidentification rate (of the order of \(\sim 1 \times 10^{-3}\)), the working point of the against muon discriminator should not have a considerable impact on the sensitivity of the analysis. The tight WP was used in order to be consistent with the other channels. However,
the TauPOG recommends to use the loose working point since it has a slightly higher tau identification efficiency. Table B.5 shows the overall yields obtained using the tight and loose WPs of the cutoff-based against muon discriminator. The main difference between them is the slightly increment on the signal yield obtained with the loose WP, which is big enough to reach the highest significance. Figure B.4 shows the significance obtained using the tight and loose WPs of the cutoff-based against muon discriminator. Although a higher significance is obtained with the WP recommended by the TauPOG, the tight WP was used in this analysis to be consistent with the \( \tau_h \) identification criteria used in the other di-tau channels.

**B.5 Decay Mode Finding (DMF) Discriminator Study**

As mentioned in section 3.7.1, there are two versions of the HPS algorithm in order to reconstruct the tau decay modes: the oldDMF and the newDMF. The oldDMF reconstructs only the decay modes with 1 or 3-prongs, while the newDMF also considers the unphysical 2-prong final state. The decay mode discriminator was selected according to the agreement between data and background in the Drell-Yan control region described in section 5.5.2. Figure B.5 shows the dihadronic tau invariant mass distribution in the Drell-Yan control region using both decay mode discriminators. As can be noticed, the invariant mass shape are consistent with the SM expectation using both discriminators and, therefore, any of them can be used for the tau identification. The newDMF discriminator was used in this analysis to be consistent with the other di-tau channels; additionally, the newDMF presents a slightly better bin-by-bin Data/MC agreement. Although the newDMF discriminator is not still completely tested by the collaboration, this work showed that there is not a significant impact on the sensitivity of the analysis.
## OVERALL YIELDS

<table>
<thead>
<tr>
<th></th>
<th>loose</th>
<th>very loose</th>
</tr>
</thead>
<tbody>
<tr>
<td>DY</td>
<td>131.6 ± 8.6</td>
<td>142.8 ± 9.1</td>
</tr>
<tr>
<td>WJets</td>
<td>41.9 ± 5.2</td>
<td>47.5 ± 5.5</td>
</tr>
<tr>
<td>DiBoson</td>
<td>3.7 ± 1.1</td>
<td>3.7 ± 1.1</td>
</tr>
<tr>
<td>t\bar{t}</td>
<td>4.4 ± 1.2</td>
<td>5.4 ± 1.4</td>
</tr>
<tr>
<td>QCD</td>
<td>382.7 ± 24.8</td>
<td>400.7 ± 25.4</td>
</tr>
<tr>
<td>TOTAL BKG</td>
<td>564.3 ± 26.9</td>
<td>600.1 ± 27.6</td>
</tr>
<tr>
<td>Z' (3 TeV)</td>
<td>1.9 ± 0.0</td>
<td>2.2 ± 0.0</td>
</tr>
<tr>
<td>Significance</td>
<td>0.34</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table B.4: Signal and background yields obtained using the loose and very loose WPs for the MVA-based against electron discriminator. The significance was estimated for \( m(\tau_h, \tau_h, E_T) > 600 \)GeV.

## OVERALL YIELDS

<table>
<thead>
<tr>
<th></th>
<th>tight</th>
<th>loose</th>
</tr>
</thead>
<tbody>
<tr>
<td>DY</td>
<td>131.61 ± 8.6</td>
<td>132.8 ± 8.7</td>
</tr>
<tr>
<td>WJets</td>
<td>41.91 ± 5.2</td>
<td>41.9 ± 5.2</td>
</tr>
<tr>
<td>DiBoson</td>
<td>3.73 ± 1.1</td>
<td>3.7 ± 1.1</td>
</tr>
<tr>
<td>t\bar{t}</td>
<td>4.39 ± 1.2</td>
<td>4.4 ± 1.2</td>
</tr>
<tr>
<td>QCD</td>
<td>382.67 ± 24.8</td>
<td>392.0 ± 25.1</td>
</tr>
<tr>
<td>TOTAL BKG</td>
<td>564.30 ± 26.9</td>
<td>574.9 ± 27.1</td>
</tr>
<tr>
<td>Z' (3 TeV)</td>
<td>1.91 ± 0.0</td>
<td>2.0 ± 0.0</td>
</tr>
<tr>
<td>Significance</td>
<td>0.34</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Table B.5: Signal and background yields obtained using the tight and loose WPs for the cutoff-based against muon discriminator. The significance was estimated in the region where the signal is expected \( m(\tau_h, \tau_h, E_T) > 600 \)GeV.

---

**Figure B.3:** Significance, as a function of the effective visible mass, obtained with the loose (blue) and the very loose (red) working points of the MVA-based against electron discriminator.
Figure B.4: Significance, as a function of the effective visible mass, obtained with the *tight* (blue) and *loose* (red) working points of the cutoff-based against muon discriminator.

Figure B.5: \( m(\tau_h, \tau_h) \) distribution for the region obtained with the *newDMF* (left) and *oldDMF* (right) discriminators in the DY selection criteria. Only the statistical uncertainties have been included.
C Event Selection

From an hypothetical $Z' \rightarrow \tau_\ell \tau_\ell$ decay, both taus would be oppositely charged, highly boosted and would travel in opposite directions. Due to the presence of neutrinos in the tau decay, a missing transverse energy requirement is applied to reduce the QCD contamination. Additionally, in order to reduce the $t\bar{t}$ contamination, a b-jet veto is required. All these topological signal selections are summarized as:

- $E_T > 30\text{ GeV}$.
- $\#\ b$-jets (CSVv2 Loose WP) = 0.
- $C_{b_{\tau_1}} \times C_{b_{\tau_2}} < 0$.
- $\Delta R(\tau_1, \tau_2)> 0.3$.
- $\cos \Delta \phi(\tau_1, \tau_2) < -0.95$.

However, since the neutrinos coming from the hadronic tau decays are expected to be very energetic and, therefore, collinear to the tau decay products, there is a complex relation between the direction of the hadronic taus and the direction of the missing transverse energy. With the purpose to identify the optimal cut that accounts for such complex relation and, that discriminates better between signal and background, the significance was computed for several topological variables that have been used in previous di-tau final state searches. Table C.1 shows the variables considered for the selection optimization.

<table>
<thead>
<tr>
<th>Label in plot</th>
<th>Color in plot</th>
<th>Selections</th>
</tr>
</thead>
<tbody>
<tr>
<td>CosDphi_LeadingTauAndMET</td>
<td>blue</td>
<td>$\cos \Delta \phi(\tau_{\text{lead}}, E_T) &lt; -0.9$</td>
</tr>
<tr>
<td>pZeta</td>
<td>green</td>
<td>$p_\zeta - 3.1 \times p_\zeta^{\text{vis}} &gt; -50\text{ GeV}$</td>
</tr>
</tbody>
</table>
| Cos_Tau1AndMET_OR_CosTau2AndMET     | purple        | $\cos \Delta \phi(\tau_{\text{lead}}, E_T) < -0.9$ or  
|                                      |               | $\cos \Delta \phi(\tau_2, E_T) < -0.9$           |
| Rate                                | orange        | $-1.05 < \frac{\cos \Delta \phi(\tau_{\text{lead}}, E_T)}{\cos \Delta \phi(\tau_2, E_T)} < -0.95$ |

Table C.1: Selections considered.

The variables of the pZeta cut are defined as:

\[
p_\zeta^{\text{vis}} = \vec{p}_{\tau_1}^{\text{vis}} \zeta + \vec{p}_{\tau_2}^{\text{vis}} \zeta,
\]

\[
p_\zeta = p_\zeta^{\text{vis}} + \vec{E_T} \zeta,
\]
where $\hat{\zeta}$ is the unit vector along the bisector of the visible tau decay products, and $p_{\zeta}^{\text{vis}}$ and $p_{\zeta}$ are two projection variables on the $\hat{\zeta}$ direction. The pZeta selection was used in the search for $Z'$ bosons in the di-hadronic channel performed with the data collected during 2015 [19].

The signal yield and the total background were estimated using the same procedure described in Chapter 5. Figure C.1 shows the comparison among the significances (as a function of the effective visible mass) obtained with each topological cut considered. The significance was computed for several $Z'$ mass points. Note in the figure that the highest significance is obtained with the $\cos \Delta \phi(\tau_{\text{lead}}, E_T) < -0.9$ cut; therefore, this cut was selected for the analysis presented in this work.

**Figure C.1:** Significance, as a function of the effective visible mass, obtained with different selections after requiring $E_T > 30$ GeV, b-jet veto (CSVv2 Loose WP), $Ch_{\tau_1} \times Ch_{\tau_2} < 0$, $\Delta R(\tau_1, \tau_2) > 0.3$ and $\cos \Delta \phi(\tau_1, \tau_2) < -0.95$. Significance computed with a $Z'$ masses of 2000 GeV(top left), 2500 GeV(top right), 3000 GeV(bottom left) and 3500 GeV(bottom right).
Bibliography


[21] V. Khachatryan et al. (CMS). Search for a massive resonance decaying into a Higgs boson and a W or Z boson in hadronic final states in proton–proton collisions at $\sqrt{s} = 8$ TeV. JHEP 02 (2016), 145. [arXiv:1506.01443]


Search for new physics with high-mass tau pairs with 2016 data. CMS-EXO-17-007, AN-17-274 To be Published.


[54] S. Chatrchyan et al. (CMS). The CMS experiment at the CERN LHC. JINST 3 (2008), S08004.


[57] V. Khachatryan et al. (CMS). Reconstruction and identification of $\tau$ lepton decays to hadrons and $\nu_\tau$ at CMS. JINST 11 (2016)(01), P01019. arXiv:1510.07488


