Search for diboson resonances in the all jets final state with CMS at $\sqrt{s} = 13$ TeV and pixel sensors development for HL-LHC

Dissertation zur Erlangung des Doktorgrades an der Fakultät für Mathematik, Informatik und Naturwissenschaften Fachbereich Physik der Universität Hamburg

vorgelegt von
Irene Zoi

Hamburg 2021

Gutachter/innen der Dissertation:	Dr. Andreas Hinzmann Prof. Dr. Erika Garutti
Zusammensetzung der Prüfungskommission:	Prof. Dr. Elisabetta Gallo Prof. Dr. Erika Garutti Dr. Andreas Hinzmann Prof. Dr. Dieter Horns Dr. Kai Schmidt-Hoberg
Vorsitzende/r der Prüfungskommission:	Prof. Dr. Dieter Horns
Datum der Disputation:	21.10.2021
Vorsitzender Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Wolfgang Hansen
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Günter H. W. Sigl
Dekan der Fakultät MIN:	Prof. Dr. Heinrich Graener

Eidesstattliche Versicherung / Declaration on oath

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Die Dissertation wurde in der vorgelegten oder einer ähnlichen Form nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Hamburg, den 15.08.2021

Unterschrift der Doktorandin

"*Due bene,* **tre meglio**, *quattro male, e cinque peggio.*"

Proverbio Toscano

Abstract

This thesis presents the first combined search for new massive resonances decaying to either to two vector bosons or to one Higgs and one vector boson, in the all jets final state. Additionally, it provides a characterization of the spatial resolution before and after irradiation of prototype small pitch sensors for the upgrade of the CMS pixel detector at the high-luminosity LHC.

Many theoretical models addressing the shortcomings of the standard model (SM), the best and yet incomplete description of elementary particles and their interactions, predict the existence of new particles with masses at the TeV scale. This thesis presents a search for such particles, which could be produced at the LHC. The adopted approach allows probing a range of such hypotentical SM extensions, in a single search and with excellent sensitivity. The analysis targets resonances with masses between 1.3 and 6 TeV decaying to pairs of bosons (WW, WZ, ZZ, WH and ZH), where the vector bosons decay hadronically and the Higgs boson decays into a pair of bottom quarks. Multiple signal production modes are investigated and, in the WH and ZH all-jets channels, vector-boson fusion (VBF) is considered for the first time. The analysis is performed on proton-proton (pp) collision data corresponding to an integrated luminosity of 138 fb^{-1} recorded with the CMS experiment at the LHC at a centre-of-mass energy of 13 TeV. Because the considered resonances have large mass, the decay products of each boson are expected to be collimated into one single large radius jet. The signal extraction and data-driven background estimation methods are based on a three-dimensional maximum likelihood fit to the distributions of the reconstructed invariant mass of the two final state jets as well as the mass of the two jets, taking advantage of the signal's resonant behaviour in these three observables. This approach was developed originally in searches for resonances decaying to two vector bosons, and has been extended for the first time in this work to WH and ZH decay modes. The analysis makes use of novel machine-learning-based algorithms to distinguish jets initiated by W/Z bosons or containing b-quark pairs from jets produced in SM background processes. The analysis approach is flexible and sensitivity is demonstrated to 16 signal hypotheses taking into account the various production and decay modes. Moreover, since the resonances can decay into multiple combinations of boson pairs, 10 such combinations are as well considered to achieve the best sensitivity. The analysis establishes a up to an order of magnitude improvement on limits on signal production cross section with respect to previous searches. In a bulk graviton model, spin-0 radions produced through gluon-gluon fusion are excluded with masses of up to 2.7 TeV, while spin-2 gravitons are excluded for masses below 1.4 TeV. When considering production exclusively through VBF, upper limits on the production cross section are set from 3 fb for a resonance mass of 1.3 TeV to 0.1 fb at 6 TeV for the Radion$\rightarrow$ VV. Furthermore, limits on the production cross section of VBF produced $\text{G}_{\text{bulk}}^{\tilde{k}=0.5}$ combining all hadronic WW and ZZ final states are set for the first time, from 4 fb for a resonance mass of 1.3 TeV to 0.2 fb at 6 TeV. In the context of the heavy vector triplet model, spin-1 W' and Z' bosons produced through quark-antiquark annihilation are excluded up to 4.8 TeV, the highest mass exclusion limit to date. In addition, for resonances produced through VBF, limits on the production cross section are set from 7–10 fb at 1.3 TeV to 0.3–0.4 fb at 6 TeV, for the first time in the WH channel, for the first time in the all-jets final states and for the first time combining VV and VH decay modes.

This thesis also shows, through simulation studies, how searches for such particles will greatly benefit from the high-luminosity LHC (HL-LHC) era. A factor of 2

improvement in signal efficiency and cross section limits are expected thanks to the increased granularity of the upgraded tracker that will be used for operation at the HL-LHC. Moreover, the HL-LHC will deliver up to 4000 fb^{-1} of data, enabling a test of the existence of massive resonances with cross sections $\approx$20 times smaller than the ones probed in this thesis.

A finer segmentation of the detectors improves the spatial resolution, the precision of the measurement of a particle position. However, during operation at the HL-LHC, the number of pp interactions in the same bunch crossing will be five times higher than in the last LHC data taking period. The detectors will receive a high radiation dose, which causes a degradation of their performance. The pixel detector, the system closest to the interaction point, will face the most challenging conditions. This thesis presents a detailed study of the spatial resolution of prototype planar silicon pixel sensors for operation at the HL-LHC, comparing the performance before and after irradiation. The prototypes are characterized by a $100\times25\,\mu m^2$ pitch, six times smaller than the one currently in use in the CMS experiment. A dedicated setup composed of three parallel planes of sensors is used to perform precise measurements in the shorter pitch direction. The measurements were performed in the DESY II test beam facility with a 5 GeV electron beam. A sensor irradiated with neutrons to $\phi_{eq} = 3.6 \times 10^{15}\,cm^{-2}$, more than 70% of the full lifetime fluence of the second barrel layer, two sensors irradiated with protons to $\phi_{eq} = 2.1 \times 10^{15}\,cm^{-2}$, corresponding to the full lifetime fluence of the third layer, and several non-irradiated sensors were tested in this study. The thesis presents a review of the different quantities adopted in literature to define the spatial resolution and introduces a new variable. The measurements were repeated for different beam incidence angles to determine the angle providing the best resolution. A spatial resolution of $2.4 \pm 0.1\,\mu m$ ($4.1 \pm 0.1\,\mu m$) was found at the optimal angle for a non-irradiated (proton-irradiated) sensor when correcting for multiple scattering. The results show that the tested sensors are suitable candidates for high precision measurements at the HL-LHC.

Zusammenfassung

Diese Arbeit stellt die erste kombinierte Suche nach neuen schweren Resonanzen, die in zwei Vektorbosonen oder ein Higgs- und ein Vektorboson zerfallen und einen Endzustand mit zwei Jets bilden, vor. Ausserdem wird die Charakterisierung der Ortsauflösung von hochauflösenden Sensoren vor und nach der Bestrahlung präsentiert. Diese Sensoren sind Prototypen für den neuen CMS Pixeldetektors für den HL-LHC. Der HL-LHC ist das Upgrade des LHC mit dem höhere Luminositäten erreicht werden sollen.

Das Standardmodell (SM) ist die beste und dennoch unvollständige Beschreibung der Elementarteilchen und ihrer Wechselwirkungen. Viele theoretische Modelle, die sich mit den ungeklärten Fragen des Standardmodells befassen, sagen die Existenz neuer Teilchen im TeV-Bereich voraus. Die in dieser Arbeit vorgestellte Analyse ist eine Suche nach solchen Teilchen, die am LHC erzeugt werden könnten. Die gewählte Methode erlaubt es, mehrere SM-Erweiterungen in einer einzigen Suche und mit ausgezeichneter Sensitivität zu testen. Es werden Resonanzen mit Massen zwischen 1.3 und 6 TeV untersucht, die in Bosonenpaare (WW, WZ, ZZ, WH und ZH) zerfallen, wobei die Vektorbosonen hadronisch zerfallen und das Higgsboson in ein Bottom-Quark-Paar zerfällt. Es werden mehrere Produktionsprozesse für Resonanzen untersucht und in den hadronischen WH- und ZH-Zerfallskanälen zum ersten Mal die Vektor-Bosonen-Fusion (VBF) betrachtet. Die Analyse wird basiert auf Proton-Proton (pp)-Kollisionsdaten, die einer integrierten Luminosität von 138 fb^{-1} entsprechen und mit dem CMS Experiment am LHC bei einer Schwerpunktsenergie von 13 TeV aufgezeichnet wurden. Da sich die Suche auf schwere Resonanzen konzentriert, wird erwartet, dass die Zerfallsprodukte jedes Bosons in einem einzigen Jet, mit großem Radius, kollimiert sind. Die Methode zur Signalextraktion und datengetriebenen Untergrundabschätzung basiert auf einem drei-dimensionalen Maximum-Likelihood-Fit der Massen der beiden Jets im Endzustand und der invarianten Masse der beiden Jets, wobei das resonante Verhalten des Signals in allen drei Observablen genutzt wird. Diese Methode wurde noch nie auf WH- und ZH-Zerfallskanäle angewendet. Die Analyse nutzt neuartige, auf maschinellem Lernen basierende Algorithmen, um Jets, die durch die Hadronisierung verschiedene Bosonen enstehen, vom SM-Untergrund zu unterscheiden. Dank der flexiblen Methode werden Ausschlussgrenzen für 16 Signalhypothesen ermittelt, wobei die verschiedenen Produktions- und Zerfallskanäle der Resonanzen berücksichtigt werden. Da die Resonanzen unterschiedliche Zerfallskanäle, d.h. unterschiedliche Kombinationen aus Bosonenpaaren, haben, werden ausserdem Ausschlussgrenzen auf die Kombination unterschiedlicher Zerfallskanäle gesetzt und die bestmögliche Sensitivität bei der Suche nach einer bestimmt Resonanz zu erreichen. Die Ausschlussgrenzen auf den Produktionswirkungsquerschnitt werden um bis zu einer Größenordnung im Vergleich zu bestehenden Suchen verbessert. In einem Bulk-Graviton-Modell werden Radionen mit Spin-0, die durch Gluon-Gluon-Fusion erzeugt werden, bis zu einer Masse von 2.7 TeV ausgeschlossen, während Gravitonen mit Spin-2 mit Massen unter 1.4 TeV ausgeschlossen werden. Für den VBF-Produktionskanal werden Obergrenzen für den Produktionswirkungsquerschnitt von 3 fb für eine Resonanzmasse von 1.3 TeV bis 0.1 fb bei 6 TeV für das Radion$\rightarrow$ VV gesetzt. Darüber hinaus werden erstmals Grenzwerte für den Produktionswirkungsquerschnitt auf $G^{\tilde{\kappa}=0.5}_{\text{bulk}}$, welches im VBF Kanal erzeugt wird, gesetzt, in dem alle hadronischen WW- und ZZ-Zerfallskanäle berücksichtigt werden. Die Grenzwerte reichen von 4 fb für eine Resonanzmasse von 1.3 TeV bis 0.2 fb bei

6 TeV. Im Rahmen eines schweren Vektortriplettmodells werden Spin-1 W'- und Z'-Bosonen, die durch Quark-Antiquark-Annihilation erzeugt werden, bis zu 4.8 TeV ausgeschlossen, die bisher beste Massenausschlussgrenze. Darüber hinaus werden für Resonanzen, die durch VBF erzeugt werden, Grenzen für den Produktionsquerschnitt zwischen 7–10 fb bei 1.3 TeV und 0.3–0.4 fb bei 6 TeV gesetzt, zum ersten Mal im WH-Kanal, zum ersten Mal bei All-Jet-Endzuständen und zum ersten Mal bei der Kombination von VV- und VH-Zerfallskanälen.

In dieser Arbeit wird auch anhand von Simulationsstudien gezeigt, wie die Suche nach solchen Teilchen von der Ära des LHC mit hoher Luminosität (HL-LHC) stark profitieren wird. Es wird erwartet, dass die Signaleffizienz und die Grenzwerte für den Produktioinswirkungsquerschnitt dank der feineren Segmentierung des verbesserten Trackers, der für den Betrieb des HL-LHC verwendet wird, doppelt so hoch sein werden. Darüber hinaus wird der HL-LHC bis zu 4000 fb^{-1} an Daten liefern, die es ermöglichen, die Existenz massiver Resonanzen mit Wirkungsquerschnitten zu testen, die $\approx$20 mal kleiner sind als die in dieser Arbeit untersuchten.

Eine feinere Segmentierung der Detektoren verbessert die Ortsauflösung, d. h. die Genauigkeit der Messung einer Teilchenposition. Während des Betriebs am HL-LHC wird jedoch die Anzahl der pp-Interaktionen während einer Strahlkreuzung fünfmal höher sein als in der letzten LHC-Datenerfassungsperiode, und die Detektoren werden einer hoher Strahlungsdosis ausgesetzt sein, was zu einer Beeinträchtigung ihrer Leistungsfähigkeit führt. Der Pixeldetektor, das den Interaktionen am nächsten liegende System, wird den schwierigsten Bedingungen ausgesetzt sein. In dieser Arbeit wird die Ortsauflösung von Prototypen planarer Silizium-Pixelsensoren für den Betrieb am HL-LHC detailliert untersucht und die Leistung vor und nach der Bestrahlung verglichen. Die Prototypen zeichnen sich durch eine Pixelgröße von $100\times25\,\mu\text{m}^2$ aus, die sechsmal kleiner ist als die, die derzeit im CMS Experiment verwendet wird. Es wird ein Aufbau aus drei parallelen Ebenen mit Sensoren benutzt um präzise Messungen in die Richtung des kleineren Pixeldurchmessers durchzuführen. Die Messungen wurden in der DESY II Teststrahlanlage mit einem 5 GeV Elektronenstrahl durchgeführt. Bei den getesteten bestrahlten Sensoren handelt es sich um einen Sensor, der mit Neutronen bis zu $\phi_{\text{eq}} = 3.6 \times 10^{15}\,\text{cm}^{-2}$ bestrahlt wurde, und um zwei Sensoren, die mit Protonen bis zu $\phi_{\text{eq}} = 2.1 \times 10^{15}\,\text{cm}^{-2}$ bestrahlt wurden, was mehr als 70% der Bestrahlung während der erwarteten Lebensdauer in der zweiten Barrelschicht bzw. dem vollen Lebensdauerfluenz der dritten Schicht entspricht. Die Arbeit gibt einen Überblick über die verschiedenen in der Literatur verwendeten Größen zur Definition der Ortsauflösung und führt eine neue Variable ein. Die Messungen wurden für verschiedene Einfallswinkel des Strahls wiederholt, um den Winkel mit der besten Auflösung zu ermitteln. Eine Ortsauflösung von $2.4 \pm 0.1\,\mu\text{m}$ ($4.1 \pm 0.1\,\mu\text{m}$) wurde für einen nicht bestrahlten (protonenbestrahlten) Sensor, unter dem optimalen Winkel und nach Korrektur von Mehrfachstreuungseffekten, gefunden. Die Ergebnisse zeigen, dass die getesteten Sensoren geeignete Kandidaten für hochpräzise Messungen am HL-LHC sind.

Contents

1 Introduction

The Standard Model (SM) of particle physics is the best-to-date description of the known elementary particles and their interactions. Since the early developments of the theory in the '60, the SM has successfully explained experimental observation and proposed precise predictions for phenomena to be observed. Experiments at lepton and hadron colliders validated the theory up to the TeV scale. The discovery of the Higgs boson in 2012 by the ATLAS [1] and CMS [2] Collaborations confirmed the existence of the last particle predicted by the SM. Despite the long list of successes, the theory has shortcomings. It fails to accommodate gravity, one of the four fundamental forces, in its framework. The SM does not include particles able to describe dark matter, nor does it explain the significant different strength of the weak and gravitational interactions. This latter flaw, the *hierarchy problem*, concerns the unexplained fact that the massive bosons of the SM are significantly lighter than the Planck mass, the mass beyond which particles would need a quantum theory of gravity to be described. The existence of this unjustified large scale separation undermines the *natuaraleness* of the theory. Therefore, several SM extensions attempt at reabsorbing the observed hierarchy. Two well known proposed solutions are the bulk scenario of the Randall-Sundrum (RS) model [3–6] and the composite Higgs model [7, 8]. The first introduces a new, warped, extra dimension that swallows up the difference in scales and explains the strength of the gravitational interaction. Two new particles are predicted by this model, the spin-2 bulk graviton and the spin-0 radion. Instead, the second removes the hierarchy by considering the Higgs as a composite particle, whose constituents are kept together by an unknown interaction mediated by new spin-1 resonances, similar to the SM weak gauge bosons. The common characteristic of these scenarios is the existence of massive resonances in the TeV range, whose presence can thus be probed at the Large Hadron Collider (LHC).

This thesis describes a search for such new particles. The method used is easily generalizable to test different possible production mechanisms and decay channels of the resonances. Therefore, with just one framework, many different models can be investigated. The search presented in this thesis focuses on resonances decaying to either two massive vector bosons (W or Z) or to one Higgs and one vector boson, channels that have a high branching fraction in several of the models under study. Since these bosons decay into quark-antiquark pairs in most cases, this analysis considers the all-hadronic final state, where both bosons decay into two quarks, which hadronize in the detector, producing a collimated spray of particles, called *jet*. For resonances with masses in the TeV range, the decay products of the bosons are merged into one single jet, and the analysis is performed in a *boosted regime*. Jet substructure observables are therefore used to distinguish boson originated jets from the large QCD multijet standard model background. The analysis looks for a resonant peak in the boson-jet masses and in the invariant mass of the two jets system, on top of a smoothly falling background dominated by QCD multijet events. The

background estimation and signal extraction procedures use a multi-dimensional maximum likelihood fit, first used in Ref. [9] to search for resonances decaying to two vector bosons. This approach increased the sensitivity of the analysis by up to 30% relative to previous methods. While Ref. [9] proved the method's validity, this analysis applies it to final states with a Higgs and a vector boson for the first time. Furthermore, newly developed algorithms analyzing the jet inner structure to identify boson originated jets and improving the signal identification efficiency have been adopted and validated. In particular, methods for the recognition of the Higgs decay into two b quarks have been introduced. Also, the new resonances can be produced through different mechanisms, illustrated in Fig 1.1. Quark-antiquark annihilation (left) and gluon-gluon fusion (middle) have higher production cross sections than vector boson fusion (right). However, the latter process is still significant and is the only production mode available in some of the theoretical scenarios. This analysis features categories dedicated to identifying resonances produced through vector boson fusion, characterized by two extra jets in the final state. Overall, the search probes sixteen different signal hypotheses with the data collected by the CMS experiment during the 2016–2018 data taking period, corresponding to a total integrated luminosity of 138 fb^{-1}. The existence of resonances with a cross section up to 60% lower than the previous version of this analysis is tested and the sensitivity to such new particles is extended by up to an order of magnitude compared to the previous most recent results.

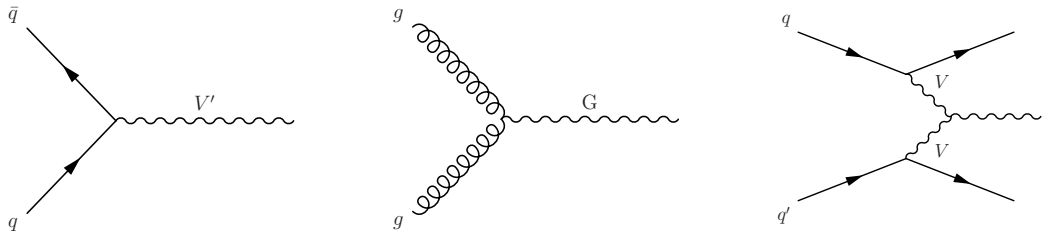


FIGURE 1.1: Feynman diagrams of the new resonances production modes. From left to right: quark-antiquark annihilation, gluon-gluon fusion and vector boson fusion.

The data collected by the CMS Collaboration will more than double by the end of the next data-taking period, and up to 4000 fb^{-1} are expected after the running period of the upgraded LHC machine at the high luminosity phase (HL-LHC) [10], enabling unique research possibilities. At the HL-LHC, resonances with even higher masses and lower production cross sections can be probed thanks to a larger dataset and a center-of-mass energy increased from 13 to 14 TeV. In this latter operation stage, the number of proton-proton interactions in the same bunch crossing (pileup) will be five times higher than in the 2016–2018 data taking period. The majority of these interactions are soft, thus not contributing to the hard process relevant for SM or new physics studies. However, the presence of additional particles can degrade the detectors' performance and consequently the quality of the measurements. Major upgrades of the CMS sub-systems are planned and being implemented to collect high-quality data at the HL-LHC. The most challenging conditions are expected close to the collision point, where the particle density and the radiation level are highest. Here is located the tracker detector that, thanks to its fine segmentation, is used to reconstruct the trajectories and momenta of charged particles. Jet substructure observables and, for instance, techniques employed in identifying b-initiated jets, like the ones used in the analysis described above, rely heavily on the tracker's ability to reconstruct tracks even inside jets. The CMS experiment will feature an

upgraded tracker with increased granularity for operation at the HL-LHC to ensure excellent tracking in a high pileup environment. A simulation study is presented in this thesis showing how the finely segmented tracker will further improve the reconstruction of charged particles with high momenta, increasing the sensitivity of searches like the one described above. The study has been performed in the context of the Technical Design Report of the upgraded tracker detector [11].

To ensure that the foreseen improvements in the tracks reconstruction are met, a study of the performance of prototypes sensors for the future pixel detector upgrade has been carried out. The silicon pixel detector constitutes the innermost component of the CMS tracker system, characterized by a fine 2D segmentation to accurately measure the position of incoming particles in high rate and radiation environments. The size of the detector segmentation determines the precision (*resolution*) of the measurement of the particle's position. The pixel cells of the upgraded detector will be six times smaller than in the current one. The spatial resolution improves with a decreased pixel size, but it degrades with radiation damage. This thesis features a study of the resolution of prototype planar sensors for the upgraded detector, comparing the performance before and after irradiation. A $100\times25\,\mu m^2$ pitch characterizes the prototypes, and the measurements are performed along the shortest dimension. The study has been performed in the DESY II test beam facility with a 5 GeV electron beam. A setup composed of three parallel planes of sensors reduces the material in the beam, compared to the use of external tracking devices, and allows the measurement of irradiated sensors, despite the additional cooling equipment. The quantity defining the sensor's resolution has been investigated and the effect of different track selection requirements, e.g. a different threshold, has been studied. The measurements have been repeated for different beam incidence angles, and the multiple scattering contribution has been assessed.

The thesis consists of two parts. The first describes the search for new heavy resonances performed with the data already collected by the CMS experiment. The standard model, its shortcomings and the considered extensions are illustrated in Chapter 2. The status of diboson resonance searches is also presented. Then, Chapter 3 describes the CMS detector, with a particular focus on the tracking system to compare it with the upgraded one. It covers also the validation of a method used to include effects of radiation damage in the simulation of the current pixel detector. In Chapter 4, the description of proton-proton collision is provided, with details of the event simulation and reconstruction. For the relevance in this thesis, Chapter 5 is dedicated to jets reconstruction and jet substructure techniques. Subsequently, Chapter 6 illustrates the strategy of the analysis presented in this thesis and the optimization of the event selections. The signal and SM background modelling is discussed in Chapter 7, where the multi-dimensional fit procedure is also detailed. Finally, Chapters 8 and 9 present the results of the search and the conclusions, respectively.

The second part presents the studies performed for the upgrade of the CMS detector. In Chapter 10, the motivations for the high luminosity phase are provided in terms of relevance for the physics research and detectors' requirements. Chapters 11 features simulation studies of the reconstruction performance of the upgraded tracker. In Chapter 12, the properties of silicon sensors are illustrated, as well as the effects of radiation damage and the characteristics of the prototypes studied in this thesis. Existing spatial resolution measurements in literature are discussed at the beginning of Chapter 13, followed by the description of the method adopted

in this thesis. The results are presented in Chapter 14. Moreover, Appendix F summarizes preliminary studies on planar and 3D sensor prototypes of a different R&D campaign.

Finally, Chapter 15 provides a summary of all the above contributions.

Part I

Search for diboson resonances in the all jets final state at $\sqrt{s}$= 13 TeV with CMS

2 The Standard Model and Beyond

The Standard Model (SM), developed in the relativistic quantum field theory framework, is a mathematical model that describes all known elementary particles and their interactions through three fundamental forces. The SM explains remarkably experimental data and has made precise predictions since its early formulation in the 1960s. In the first part of this chapter, based on [12–15], the basic concepts of the SM are summarized. Despite being the to-date best description of the constituents of matter and the forces acting among them, the SM still has open questions and shortcomings that will be briefly addressed in the second part of this chapter. Finally, possible extensions of the model, relevant for the searches presented in this thesis, are presented.

2.1 The Standard Model of Particle Physics

The Standard Model classifies the matter particles as *fermions*, spin 1/2 particles governed by the Pauli principle. Particles with integer spin, such as the mediators of the electromagnetic, weak, and strong forces, are called *bosons*.

Fermions are further arranged in *leptons* and *quarks*, each of them additionally divided in three generations with increasing masses. Each generation has two particles. Thus, a total of 12 different kinds of particles is categorized based on their intrinsic properties, the *quantum numbers*: mass m, spin s, electromagnetic charge e, weak isospin T_3 and color charge C. Each particle has an antiparticle with same mass but opposite charges. Quantum numbers define to which interactions a particle is subjected. A particle possessing electromagnetic charge interacts through the electromagnetic force with other particles having electromagnetic charge. A similar behaviour happens for the weak isospin and the color charge, corresponding to the weak and the strong interactions, respectively.

2.1.1 Leptons

Each generation of leptons is composed by a particle with $1e$ electromagnetic charge (electron e, muon μ or tau τ) and a corresponding neutral particle, the neutrino (ν_e, ν_μ or ν_τ). Leptons are color neutral. Leptons with $T_3 = \pm\frac{1}{2}$ are called left-handed and appear in doublets $\begin{pmatrix} \ell \\ \nu_\ell \end{pmatrix}_L$ where to the charged lepton ℓ corresponds $T_3 = +\frac{1}{2}$ and ν_ℓ has $T_3 = -\frac{1}{2}$, with $\ell = e$, μ or τ. If $T_3 = 0$, the lepton is right-handed and appears in singlets. Right-handed neutrinos have not yet been observed. Moreover, neutrinos are the only fermions interacting only through the weak force.

2.1.2 Quarks

Quarks couple to all three interactions described by the SM. They carry one of the three color charges (blue, red or green). Like the leptons, they are also divided in left-handed and right-handed based on their weak isospin. Again, left-handed (right-handed) quarks appear in doublets (singlets). The doublets

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L$$

are composed by an up-type quark (up u, charm c or top t) with electromagnetic charge $+\frac{2}{3}e$ and weak isospin $T_3 = +\frac{1}{2}$ and a down-type quark (down d, strange s or bottom b) with electromagnetic charge $-\frac{1}{3}e$ and weak isospin $T_3 = -\frac{1}{2}$.

2.1.3 Bosons

The mediators of the fundamental forces are bosons with spin = 1. The weak interaction is mediated by the $W^{\pm}$ and the Z^0 bosons with electromagnetic charge ± 1 and 0, respectively. In this thesis the letter V is used to indicate either of the vector bosons W or Z. The W has mass m_W = 80.379 $\pm$ 0.012 GeV while the mass of the Z boson is m_Z = 91.1876 $\pm$ 0.0021 GeV [15]. The photon γ is the mediator of the electromagnetic force and has zero electromagnetic charge. Gluons, mediators of the strong interactions, are electromagnetically neutral. Differently from the other mediators, gluons carry the charge representative of their interaction, color, and can interact with themselves.

The Higgs H boson has spin = 0 and is electromagnetically neutral. It is not a carrier of a fundamental force but through its field the other particles can gain mass. It has a mass m_H = 125.10 $\pm$ 0.14 GeV [15].

A graphical visualization of the SM particles and the interactions among them is presented in Fig. 2.1.

The mathematical formulation of the SM as a quantum field theory requires the SM Lagrangian $\mathcal{L}_{SM}$ to satisfy a set of symmetries. First, $\mathcal{L}_{SM}$ is invariant under global transformation of the Poincaré group (rotations, Lorentz boosts, and space-time translations). Moreover, $\mathcal{L}_{SM}$ is required to be a local gauge invariant under the symmetry group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{2.1}$$

where $SU(3)_C$ is the color gauge group describing the strong interaction, $SU(2)_L$ represents the weak interaction and $U(1)_Y$ refers to the electromagnetic interaction, with Y being the hypercharge.

In the first part of this thesis, natural units such that $c = \hbar = 1$ are adopted.

2.1.4 Electroweak interaction

The $SU(2)_L \otimes U(1)_Y$ subgroup describes the electroweak sector. The isospin T_3 is the charge of $SU(2_L)$ and W_i (i = 1, 2, 3) the gauge bosons. For $U(1)_Y$ the gauge boson is B. The respective couplings are g_W and g_Y.

As a consequence of the spontaneous symmetry breaking, described in Sec. 2.1.6, $SU(2)_L \otimes U(1)_Y$ is broken down to $U(1)_{em}$ and the two electromagnetic and weak forces with their respective mediators arise. Thus, below the unification energy, the two forces behave differently with the weak interaction responsible for the radioactive decays [17] and the electromagnetic interaction creating atomic and molecular

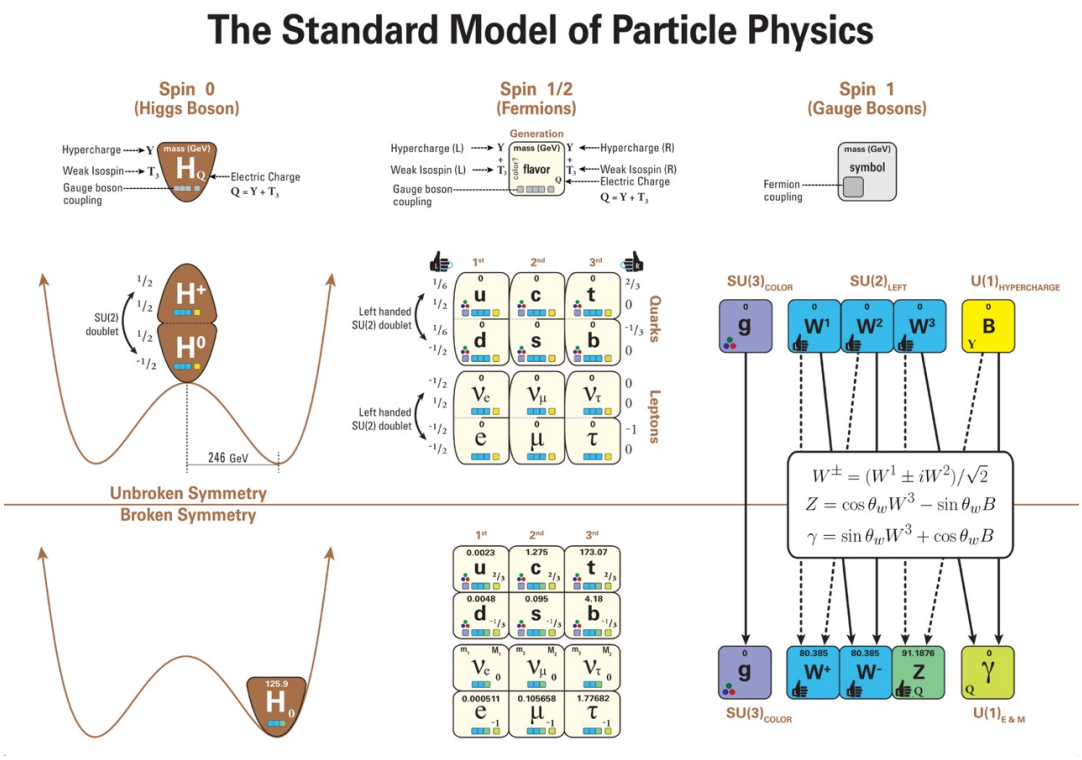


FIGURE 2.1: Graphical visualization of the Standard Model. The (so far discovered) foundamental matter particles, leptons and quarks, are shown in their variations in the center of the figure. Symbols representing the different interactions are also provided. The mediator of the interactions are shown on the right, while the properties of the Higgs boson on the left. The symmetry breaking mechanism described in Sec. 2.1.6 is also illustrated. Antimatter is not shown for semplicity. Taken from Ref. [16].

bindings. Before the symmetry breaking, the two interactions can be described in a unified manner, first proposed in Refs. [18–21].

The physical mediators originated from the symmetry breaking are related to the gauge boson by

$$\begin{aligned} W^{\pm} &= \frac{W_1 \mp iW_2}{\sqrt{2}} \\ Z &= \frac{g_W}{\sqrt{g_W^2 + g_Y^2}} W_3 - \frac{g_Y}{\sqrt{g_W^2 + g_Y^2}} B \\ \gamma &= \frac{g_Y}{\sqrt{g_W^2 + g_Y^2}} W_3 + \frac{g_W}{\sqrt{g_W^2 + g_Y^2}} B. \end{aligned} \tag{2.2}$$

The weak mixing angle θ_W between W_3 and B is $\theta_W := \tan^{-1}(g_Y/g_W)$ while the fundamental unit of electric charge is

$$q := \frac{g_W g_Y}{\sqrt{g_W^2 + g_Y^2}}.$$

The $W^{\pm}$ bosons, responsible for the weak charged current, couple only to particles with $T_3 = \pm 1/2$: left-handed fermions and right-handed anti-fermions. The Z boson is responsible for the weak neutral current and couples to both left- and

right-handed particles but with different coupling strengths. The photon, mediator of the electromagnetic interaction, couples to charged particles.

Each generation of leptons has a lepton number such that L_e is 1 for the e generation and zero for the μ and τ generation. L_μ and L_τ are similarly defined. The weak force is the only interaction that allows the conversion of a charged lepton in its corresponding neutrinos. Mixing among generations is not allowed by the conservation of the lepton number. However, it has been observed that neutrinos oscillate between different generations [22–24].

The interaction of quarks with the $W^\pm$ bosons allows mixing among generations following the Cabibbo- Kobayashi-Maskawa (CKM) matrix [25, 26]:

$$\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97370 \pm 0.00014 & 0.2245 \pm 0.0008 & 0.00382 \pm 0.00024 \\ 0.221 \pm 0.004 & 0.987 \pm 0.011 & 0.0410 \pm 0.0014 \\ 0.0080 \pm 0.0003 & 0.0388 \pm 0.0011 & 1.013 \pm 0.030 \end{bmatrix} \tag{2.3}$$

where the values are obtained from Ref. [15]. The probability of quark i transforming in quark j is given by $|V_{ij}|^2$, showing that generation changing processes are allowed but suppressed.

2.1.5 Quantum Chromodynamics

$SU(3)_C$ is the gauge field theory of the strong interaction, involving coloured quarks and gluons, and is therefore called *Quantum Chromodynamics*. Gluons also have a color charge and can interact among themselves. The fact that the mediators carry the charge characterizing the interaction is a peculiarity of QCD and has important consequences on the *strong coupling* α_s. The coupling strength is higher at large distances, *confining* the quarks in colorless bounded states called *hadrons*. The top quark, the most massive particle of the SM with $m_t = 172.76 \pm 0.30$ GeV [15], has a short lifetime and decays before hadronizing. As a consequence of the hadronization process, quarks and gluons result in a collimated spray of particles in the detectors, called *jet*.

On the contrary, at higher energies/smaller distances, the interaction among quarks is weaker and they become nearly free particles (*asymptotic freedom*).

2.1.6 The Higgs mechanism of symmetry breaking

As it was mentioned, the SM Lagrangian should be invariant after local gauge transformations, requiring the gauge fields to be massless. While this is true for the photons and the gluons, $W^\pm$ and Z bosons are massive. A solution to this inconsistency was proposed by Brout, Englert, and Higgs with the introduction of the spontaneus electroweak symmetry breaking [27, 28].

The *spontaneous symmetry breaking* can happen, for instance, in a quantum system described by a symmetric potential $V(x) = \frac{V_0}{x_0^4}(x^2 - x_0^2)^2$ such as the one in Fig. 2.2, which is invariant for $x \to -x$. This potential is called "double well" because it has two degenerate equal energy minima $\pm x_0$. Randomly choosing to expand around one of the two for the description of the low energy states spontaneously breaks the symmetry.

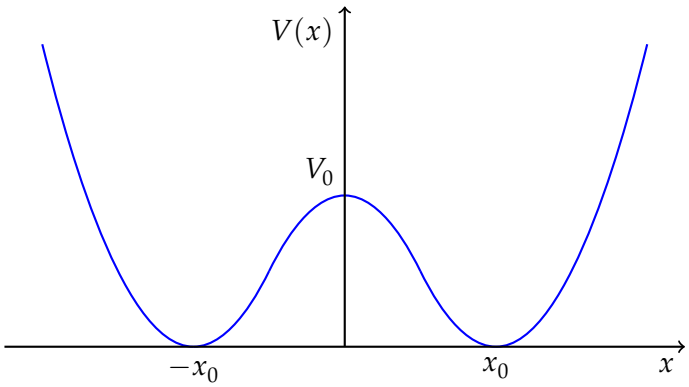


FIGURE 2.2: "Double well" potential in 1 dimension. It presents a symmetry for $x \to -x$ and two degenerate equal energy minima.

The quantum field theory equivalent is given by a complex spin-0 scalar field $\phi(x) = \phi_1(x) + i\phi_2(x)$ with the potential

$$V(\phi) = \lambda \left(|\phi|^2 - \frac{v^2}{2} \right)^2 . \tag{2.4}$$

The potential is 0 when $|\phi| = \frac{v}{\sqrt{2}}$, corresponding to the minimum $\frac{v}{\sqrt{2}}$ also called vev, the *vacuum expectation value*, because it is the state assumed by $|\phi|$ in the vacuum, without injection of energy. To guarantee always non-negative energy, the quartic coupling λ should be positive.

To ensure the breaking of the electroweak symmetry, the scalar field must couple to the electroweak gauge bosons and thus be a weak isospin doublet with hypercharge Y=1. The simplest possibility is

$$\vec{\phi} = \begin{pmatrix} \phi^+ \\ \phi_0 \end{pmatrix}$$

with ϕ^+ and ϕ_0 complex scalar fields, for a total of four degrees of freedom.

For the expansion around the vev, $\vec{\phi}$ can be expressed as

$$\vec{\phi} = \begin{pmatrix} \phi^+ \\ \phi_0 \end{pmatrix} = \begin{pmatrix} \frac{r_1(x)}{\sqrt{2}} e^{i\frac{\theta_1(x)}{\sqrt{2}v}} \\ \left(\frac{v}{\sqrt{2}} + \frac{r_2(x)}{\sqrt{2}} \right) e^{i\frac{\theta_2(x)}{\sqrt{2}v}} \end{pmatrix}$$

where r_1, r_2, θ_1 and θ_2 are scalar fields, with one degree of freedom. Eliminating three of them with an appropriate $SU(2)_L \otimes U(1)_Y$ gauge transformation will give three of the force-carrying bosons a mass. The transformed field is

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{2.5}$$

where $H(x)$ is a real scalar field called the *Higgs boson*.

To determine the masses of the mediator bosons, the field in the form of Eq. 2.5 is inserted in the Lagrangian describing its interaction with the electroweak force obtaining the expressions introduced in Eq. 2.2 for the mediators. The masses of the W and Z bosons are $m_W = g_W v/2$ and $m_Z = \frac{v\sqrt{g_W^2 + g_Y^2}}{2}$, respectively. The photon is

still massless and thus it has a manifest gauge symmetry, ensuring the electric charge conservation. Inserting Eq. 2.5 in Eq. 2.4 the following expression for the potential is obtained:

$$\lambda v^2 H(x)^2 + \lambda v H(x)^3 + \frac{\lambda}{4} H(x)^4. \tag{2.6}$$

The coefficient of the quadratic term describing the Higgs self-interaction gives the Higgs mass $m_H = \sqrt{2\lambda}v$. While the vev can be obtained from other measurements, λ was unconstrained, giving no indication of the size of the Higgs mass. The Higgs boson was discovered in 2012 [1, 2] confirming the Higgs mechanism of symmetry breaking.

As a consequence of the left-handed nature of the weak interaction, gauge invariant masses m_f for the fermions originate from their couplings with the $\vec{\phi}$ field but are not directly predicted by the Higgs mechanism

$$m_{f_i} = \frac{1}{\sqrt{2}} v y_i$$

where y_i is the Yukawa coupling.

2.2 Shortcomings of the model

Experimental data have outstandingly confirmed the prediction of the Standard Model [29]. Nevertheless, there are still open questions and unexplained phenomena whose description is not covered in the SM. The main shortcomings are summarized here.

2.2.1 Gravity

While the SM successfully describes three of the four known fundamental interactions, gravity has not been included yet. This is a consequence of the incompatibility of general relativity with quantum mechanics. In a quantum field theory framework, the gravitational attraction could originate from the exchange of virtual spin-2 *gravitons*. Unfortunately, this approach is non-renormalizable, with divergences preventing meaningful physical predictions.

2.2.2 Dark Matter and Dark Energy

The particles described in the Standard Model only account for ≈5% of the energy in the Universe. The measured orbital velocity of stars within galaxies [30] requires the mass content within the orbital radius to be much higher than the one provided by visible matter. Thus, it has been suggested the existence of so called *dark matter*, weakly interacting with SM particles [31]. Additionally, to explain the accelerated expansion of the Universe [32, 33], a new form of energy with negative pressure must be formulated, *dark energy*. Dark matter and dark energy account for ≈26% and 69% of the content of the Universe. So far the only evidence we have for these contributions is through gravity and the search for the particle nature of dark matter is still on going.

2.2.3 Neutrino masses

In the standard model, neutrinos are predicted to be massless. The observation [22–24] of neutrino flavour oscillations indicates that they have non-zero mass, confirmed by measurements [34], in contrast with the theoretical expectation.

2.2.4 The Baryon Asymmetry

The imbalance between matter and antimatter in the Universe [35–37] is not explained by the SM. The Big Bang model predicts the creation of an equal amount of matter and antimatter. While some SM processes violate the CP symmetry, their magnitude cannot explain the asymmetry observed today.

2.2.5 The Hierarchy Problem

The unexpectedly large difference between the mass of the V bosons and the Planck mass M_{Pl}, $\frac{m_V}{M_{Pl}=\sqrt{\hbar c/G_{Newton}}} \approx 10^{-17}$, is commonly referred as the *hierarchy problem* in the mass scale of the weak and gravitational interactions [38, 39]. This unexplained discrepancy is not a direct threat to the validity of the SM but reduces its solidity.

As illustrated in Sec. 2.1.6, the masses of the V bosons are a consequence of the non-zero expectation value of the Higgs field $\langle|\phi|\rangle = \nu \approx$246 GeV. The Higgs mass, measured to be $\approx$ 125 GeV = $m_H = \sqrt{2\lambda}\nu$ also depends on it and on the quartic coupling λ. The quantum correction to the physical Higgs mass are

$$\frac{\delta m_H^2}{m_H^2} \propto \left(\frac{4m_t^2}{m_H^2} - \frac{2m_W^2}{m_H^2} - \frac{m_Z^2}{m_H^2} - 1\right)\Lambda^2 \propto \left(\frac{\Lambda}{500\ \text{GeV}}\right)^2$$

where Λ is the scale up to which the SM is valid. If Λ is of the order of M_{Pl}, large corrections of the order of 10^{38} GeV must be added to obtain the measured Higgs mass, requiring *fine tuning* cancellation of the diverging terms. This could be avoided if Λ is of the order of the TeV, implying the existence of new particles with a mass of that order of magnitude [40].

Models addressing this issue and describing these potential new phenomena are addressed in the next section.

2.3 Extensions of the model

Several Beyond the Standard Model (BSM) theories propose (partial) solutions to the SM shortcomings described in the previous section. Here two models addressing the hierarchy problem are reviewed, namely the *warped extra-dimension* model and the *composite Higgs* model. These models predict the existence of resonances that could be observed at hadron collider experiments through searches as the one described in this thesis.

2.3.1 Warped extra-dimension model

The Randall-Sundrum model [3, 4] suggests the existence of an additional compact *warped* spatial dimension where gravity is allowed to propagate. This would justify our weakened perception of the gravitational force.

This idea was first developed in the Arkani-Hamed, Dimopoulos and Dvali (ADD) model [41] which supposes the existence of n extra compact spatial dimensions with

radius R. The Planck mass M_{Pl} is then just an effective mass related to the $(4+n)$-dimensional Planck mass $M_{Pl(4+n)}$ by $M_{Pl}^2/8\pi = \overline{M_{Pl}}^2 \sim M_{Pl(4+n)}^{n+2} R^n$. Assuming $M_{Pl(4+n)} \sim$ TeV, for n=1, $R \sim 10^8$ m, as the Earth-Sun distance. This case is thus excluded since the gravitational law is well tested on this scale. If $n \geq 2$, $R \sim 10^{-4}$ m is at the limit of experimental reach [42]. However, while this approach solves the hierarchy problem of the SM, a new one is introduced between $M_{Pl(4+n)}$ and the compactification scale, where features smaller than the the size of the extra dimension can be observed.

The case of n=1 with compactification scale $R \sim 1/\text{TeV}$ has been recovered in the Randall-Sundrum (RS) model [3, 4] with the introduction of a warped geometry in a then 5-dimensional Anti-de Sitter (AdS) spacetime, in which the 4-D Minkowski metric is multiplied by an exponential "warp" factor of the additional dimension. The extra-dimension (*bulk*) is parametrized by $\varphi \in [0, \pi]$ representing a segment between two branes, localized at the end points, the Planck ($\varphi = 0$) and the TeV ($\varphi = \pi$) brane. In the simplest version of the RS model (RS1), the SM particles are localized on the latter, while gravity is relatively stronger on the first and propagates in the bulk. Fields with a fundamental mass m_0 in the bulk have a mass $m = \mathrm{e}^{-kr_c\pi} m_0$ in the TeV brane [3] with k of the order of M_{Pl} and r_c the compactification radius. The assumption that the Higgs field is localized on the TeV brane implies the same suppression relation for the vev $v = \mathrm{e}^{-kr_c\pi} v_0$ with the warp factor solving the hierarchy problem.

New particles originate from the quantum fluctuations of the classical solution to the 5-D metric

$$ds^2 = \mathrm{e}^{-2kT(x)|\varphi|}(\eta_{\mu\nu} + h_{\mu\nu}(x))dx^\mu dx^\nu + T^2(x)d\varphi^2,$$

with $\eta_{\mu\nu}$= diag(1;-1;-1;-1) the Minkowski metric. The zero modes $h_{\mu\nu}(x)$ correspond to the massless gravitons, mediators of the gravitational interaction in the effective field theory. From the tensor fluctuations $T(x)$ of the 4-dimensional part of the metric originate spin-2 excitations, the Kaluza-Klein (KK) graviton. The graviton excitations are close to the TeV brane and could be produced at collider experiments. A spin-0 field, the *radion*, rises from the scalar fluctuations around the 5^{th} extra-dimension.

An extension has been proposed, called the *bulk scenario*, in which the SM particles can propagate in the bulk of the extra-dimension, with the exception of the Higgs boson to avoid the hierarchy problem [5, 6]. The SM fields, in this approach, correspond to the zero modes of their corresponding 5-dimensional fields. Having first and second generation fermion fields close to the Planck brane and the top and V bosons near the TeV one, where the Higgs is confined, would explain the strength of the observed Yukawa couplings.

The free parameters of the model are the mass of the first KK-graviton excitation and the ratio $\tilde{k} = k/\overline{M_{Pl}}$ that governs the total decay width. The decay width is expected to be of the order of percent with respect to the mass and thus the narrow width approximation holds, allowing to probe the resonance properties at the peak.

Since in the bulk model light quarks are close to the Planck brane but gluons have a flat localization in the bulk, the KK-graviton production through gluon-gluon fusion, shown in Fig. 2.3a, is favoured. Similarly, the decays to Higgs or top have the highest branching ratio (BR). The wave function of the gauge bosons is mostly flat in the bulk and falls sharply near the brane. Hence, the branching fraction of decays into a pair of weak bosons is comparable to the one into Higgs and top.

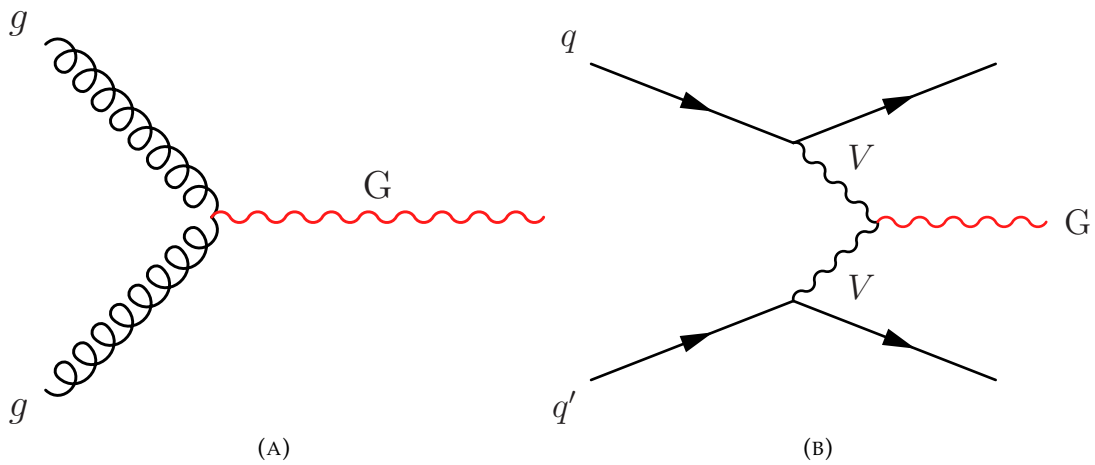


FIGURE 2.3: Feynman diagrams of Graviton and Radion production modes, on the left gluon-gluon fusion and on the right Vector Boson Fusion.

Similarly, Vector Boson Fusion (VBF) production mode is also enabled (Fig. 2.3b). The branching ratios of the different decay modes are showed in Fig. 2.4, while the production cross section σ in Fig. 2.5.

2.3.2 Heavy Vector Triplet

Several theoretical models, such as composite Higgs [7, 8, 48] or sequential extensions of the SM [49, 50], predict the existence of new spin-1 resonances in their attempt to solve the hierarchy problem. In the composite Higgs scenario, for instance, such bosons are the mediators of a new interaction holding the Higgs "partons" together. With the Higgs not being an elementary scalar any more, the hierarchy is removed.

However, the models are characterized by different details and sets of free parameters to which experimental observation are not really sensitive to. Thus, the Heavy Vector Triplet (HVT) [51] framework has been developed introducing an effective Lagrangian to provide a phenomenological description reproducing a large class of these explicit models one. The validity of this simplified model is restricted to on-shell quantities such as the $\sigma\times$BR. Experimental data provide limits on $\sigma\times$BR that can be interpreted in the HVT context or in the explicit models.

As the framework name anticipates, a real vector field V_a ($a = 1, 2, 3$) is introduced in the adjoint representation of $SU(2)_L$ describing three new particles, which have the following charge eigenstate fields

$$V^{\pm} = \frac{V_1 \mp iV_2}{\sqrt{2}}$$
$$V^0 = V_3.$$

The three new heavy bosons, W' and Z', mirror the SM vector bosons and thus are almost degenerate in mass ($M_V \approx M_{\pm,0}$).

The typical strength of a V' boson interaction with the SM W and Z bosons is described by the coupling g_V, which can vary from $g_V \sim 1$ in the weakly coupled case to $g_V \sim 4\pi$ in the strong limit. Two scaling factors, c_H and c_F, parametrize the differences from the typical strength. The first one describes the interaction with the Higgs and the SM vector bosons, especially the decays in bosonic channels, while

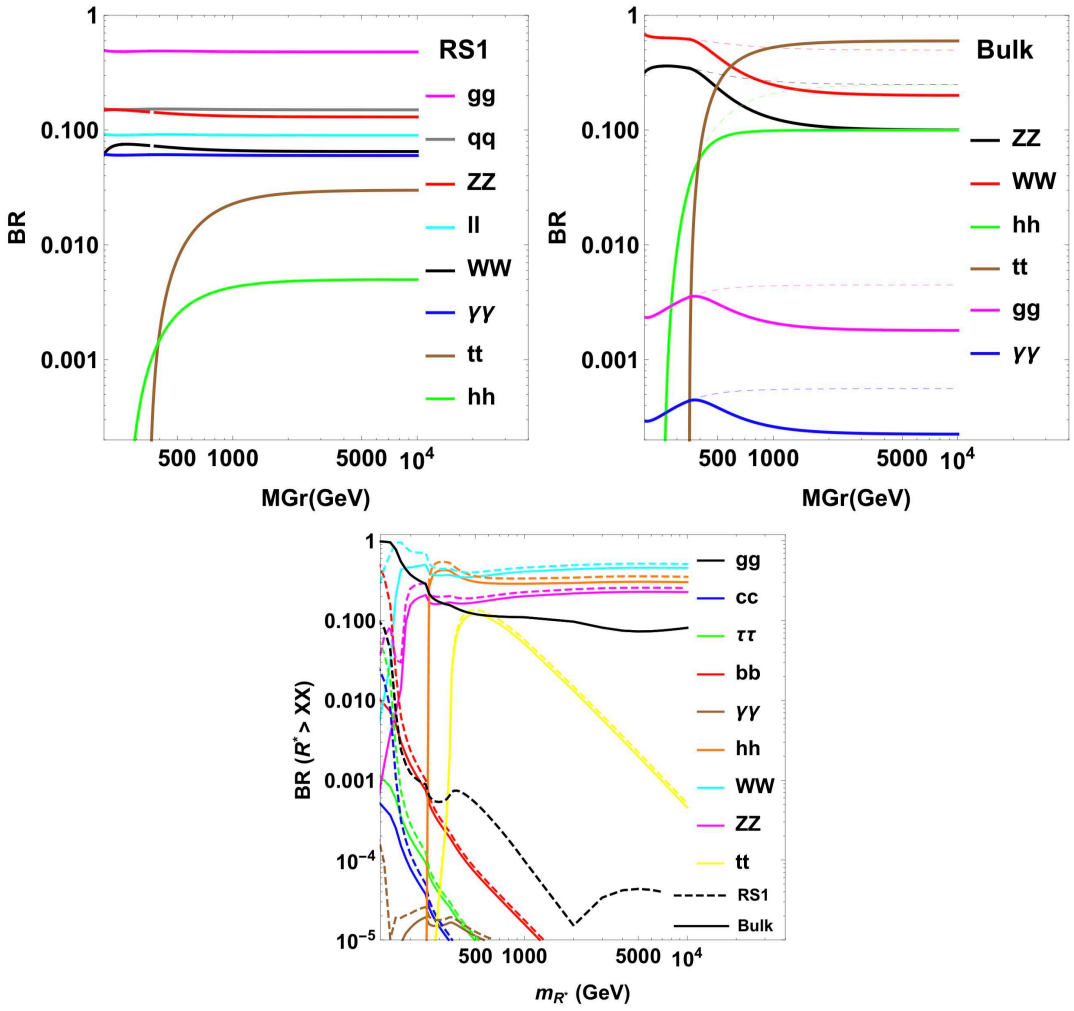


FIGURE 2.4: Top: Graviton branching fractions. Top left: RS1 scenario. The symbol q stands for the sum of light quarks (u, d, s, c, b), while l represents the sum of the three flavors of leptons (e, μ, τ) or neutrinos. Top right: Bulk scenario comparing two hypothesis of fermion embedding. The branching ratios are independent of the k parameter. Bottom: Radion branching fractions to bosons and fermions. The dashed lines stand for RS1 scenario, while the solid ones for the bulk scenario, when $kl = kr_c\pi = 35$. Taken from Ref. [43].

the latter regulates the interaction with fermions and is thus responsible for both the resonance production through Drell-Yan (DY) process (Fig. 2.6) and the fermionic decays. For the interpretation of experimental results, the combinations $g_V c_H$ and $g^2/g_V c_F$ can be treated as fundamental parameters, where g is the is the gauge coupling. In this framework the partial widths are

$$\Gamma_{V^\pm \to f\bar{f}'} \simeq 2\Gamma_{V^0 \to f\bar{f}'} \simeq N_c[f] \left(\frac{g^2}{g_V c_F}\right)^2 \frac{M_V}{48\pi}$$

$$\Gamma_{V^\pm \to W_L^\pm Z_L} \simeq \Gamma_{V^0 \to W_L^+ W_L^-} \simeq \frac{g_V^2 c_H^2 M_V}{192\pi}$$

$$\Gamma_{V^\pm \to W_L^\pm H} \simeq \Gamma_{V^0 \to Z_L H} \simeq \frac{g_V^2 c_H^2 M_V}{192\pi},$$

where $N_c[f]$ is the number of colors and is equal to 3 for the di-quark and to 1 for the di-lepton decays. The first equation describes the decays into two fermions and

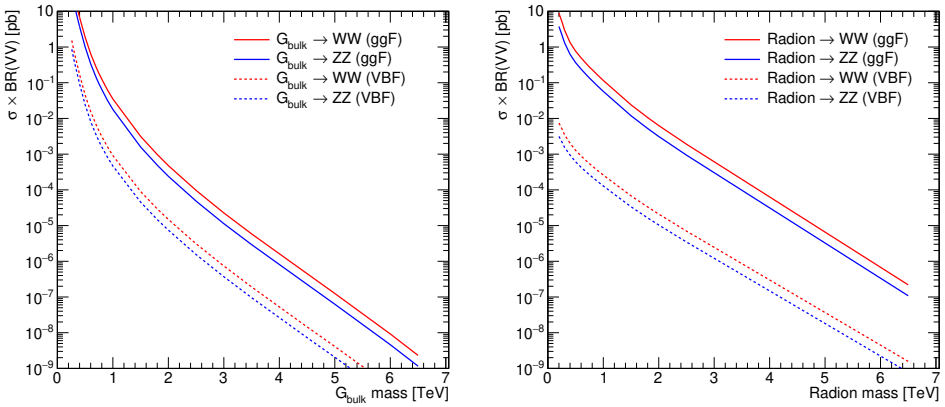


FIGURE 2.5: Product of the theoretical production cross section at a center-of-mass energy of 13 TeV and the branching ratio for the decay into two SM bosons as a function of the resonance mass hypothesis for a bulk graviton (left) and a radion (right) in the gluon fusion (solid lines) and VBF (dash lines) production modes. Obtained from Refs. [44–47].

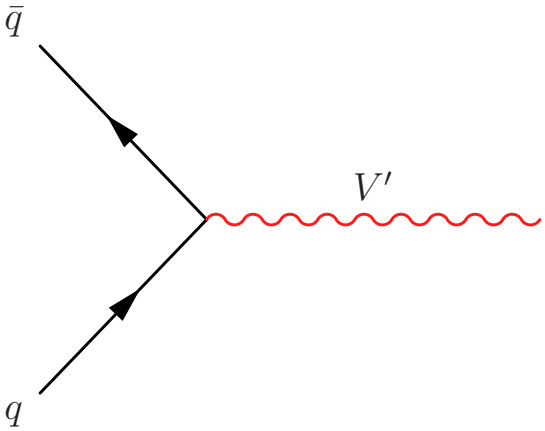


FIGURE 2.6: Feynman diagrams of V' production through quark-quark annihilation, Drell-Yan process.

controls also the DY production rate. The other two equations regulates the decays into a pair of SM vector bosons (with longitudinal polarization) or a SM vector boson and a Higgs boson. Decays like HH, $\gamma\gamma$ or $W\gamma$ are generally suppressed. The production through VBF is also possible, relying on the validity of the *effective W approximation* [52].

Two benchmark models are proposed in Ref. [51]. The *model A* mimics a scenario with an extended gauge symmetry with an additional $SU(2)'$. In this scenario $c_H \sim -g^2/g_V^2$ and $c_F \sim 1$. Hence, the decay into bosons is suppressed for high g_V values.

In the *model B* the possibility for the Higgs to be a light state from an underlying strong dynamics, like in composite Higgs models, is represented with $c_H \sim c_F \sim 1$. In this case the decay into bosons is not suppressed.

An additional scenario in which the resonances are produced exclusively through VBF, the *HVT model C*, is obtained with $c_H = g_V = 1$ and $c_F = 0$.

Figure 2.7 shows the production cross section in the model B and model C scenarios (top) as well as the decay branching ratios for W' and Z' (bottom).

For the narrow width approximation to hold, $g_V \leq 7$ is required, as in the cases considered in this thesis.

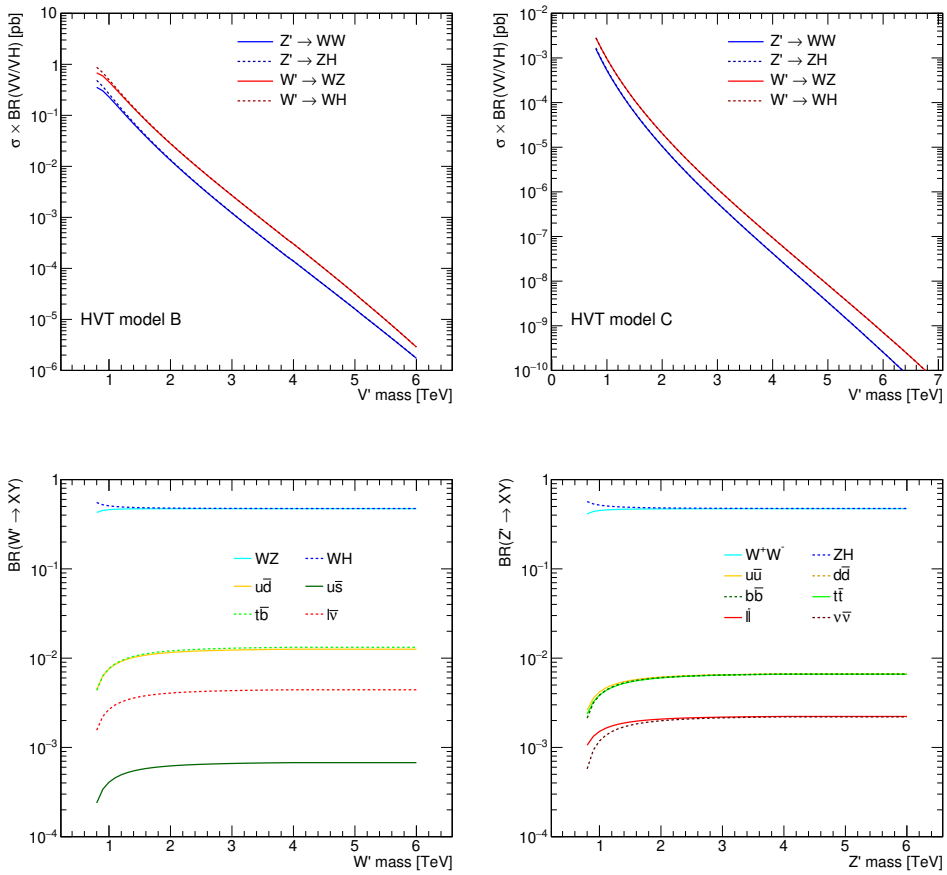


FIGURE 2.7: Top left: Cross sections for the DY production of the neutral and charged resonances as a function of the resonance mass in HVT model B. Top right: Cross sections for the VBF production of the neutral and charged resonances as a function of the resonance mass in HVT model C. The cross sections are calculated for a center-of-mass energy of 13 TeV. Bottom: Branching ratios for the decay of the W' (left) and Z' (right) resonance in two SM bosons or fermions. Obtained from Refs. [53, 54].

2.4 Diboson searches

A large variety of SM extensions, like the ones presented in the previous section, predicts the existence of new spin-0, spin-1 and spin-2 resonances, which could have large couplings to gauge bosons. The ATLAS and CMS collaborations have been performing searches for such particles since the early days of the LHC. They have included many different final states (VV, VH, HH, $\gamma\gamma$, ...) while developing new analysis techniques. The recent reviews in Refs. [55, 56] cover a broad set of these results.

The following contains a summary of the latest results of searches for massive resonances decaying into VV and VH channels, also studied in this thesis. The VV and VH channels, as it was described in the previous section, are the preferred decay modes for the new particles in several SM extensions, providing the best sensitivity. The HH and $\gamma\gamma$ decays are suppressed, for instance, in the HVT model B. As Fig. 2.8 (left) shows, the VV and VH channels set more stringent constraints on the parameters of the HVT models in certain phase space regions, compared to fermionic final states.

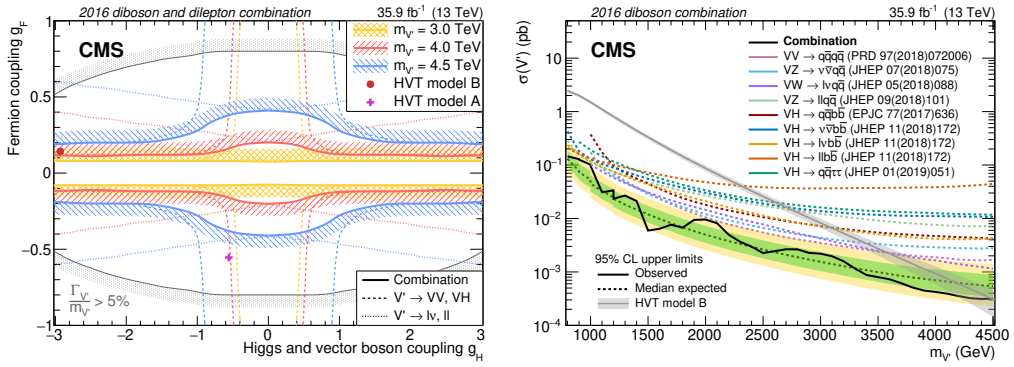


FIGURE 2.8: Left: Observed exclusion limits on the couplings of heavy vector resonances to fermions and SM vector bosons and the Higgs boson for the statistical combination (solid lines) of the dilepton (dotted lines) and diboson channels (dashed lines). Three resonance masses hypotheses (3.0, 4.0, and 4.5 TeV) are considered. The hatched bands indicate the regions excluded. The areas bounded by the thin gray contour lines correspond to regions where the resonance widths ($\Gamma_{V'}$) are predicted to be larger than the average experimental resolution (5%). Right: Observed and expected 95% CL upper limits on cross sections as a function of the HVT triplet mass for the combination of all channels in the HVT model B. The inner green and outer yellow bands represent the $\pm$ 1 and $\pm$ 2 standard deviation variations on the expected limit. The solid curve surrounded by the shaded area shows the cross sections predicted by HVT model B and its uncertainty. Taken from Ref. [53].

At the time of writing, experiments have observed no evidence for such particles. Thus results are presented as exclusion limits on the mass or cross section of the resonances.

The experimental signature of the VV and VH channels consists of two back-to-back, highly energetic bosons.

Hadronic decays of the W and Z bosons into two quarks have the largest branching ratio, while for the Higgs the decay into two b-quarks is favored. Searches conducted in all-hadronic final states benefit from the high branching fraction but have to deal with a significantly high background from QCD multijet events. In Fig. 2.8 (right) the sensitivity of different all-hadronic and semi-leptonic final states can be compared, with the all-hadronic final states being among the most sensitive ones, especially for resonances above $\approx$1.5 TeV. An overview of the most recent VV all hadronic searches performed at a center-of-mass energy of 13 TeV by the CMS collaboration is presented in Ref. [57]. Especially relevant is the result obtained in [9], which is the most stringent limit in the above mentioned channel, despite using almost a half of the statistics used in the latest correspective ATLAS search [58]. The analysis presented in [9] achieves such a high sensitivity thanks to a multi-dimensional fit approach which is adopted also in this thesis. While in Ref. [9] only VV final states were covered, in this work the method is extended to VH hadronic final states and VBF production mode. The latest all-hadronic CMS search for resonances decaying to a vector boson and a Higgs boson is presented in Ref. [59] and is based on 35.9 fb^{-1} collected data, while the analougus ATLAS analysis is performed with 139 fb^{-1} [60].

Searches are carried out also in the semi-leptonic final state, in which only one of the bosons decays into quarks and the other decays leptonically, allowing to reduce the SM background and explore lower regions of the possible resonance mass range. Both collaborations have performed searches in the VV final state, $Z(\nu\bar{\nu})V(q\bar{q})$ [61, 62], $W(\ell\nu)V(q\bar{q})$ [61, 63] and $Z(\ell\ell)V(q\bar{q})$ [61], placing some of the most stringent

limits. CMS has also covered the V(lep)H(had) final state [63, 64] with a 138 fb^{-1} dataset, while ATLAS has a result [65] with 36 fb^{-1}.

In the simplified models of the HVT framework, the masses of the new heavy bosons are nearly degenerate, thus the sensitivity of the single analysis can be improved by simultaneously searching for same mass Z' and W'. Additionally, the detector resolution does not allow to fully resolve the SM W/Z peaks and thus searches targeting different bosonic decay channels can also be combined to increase the sensitivity. The latest combinations from the ATLAS and CMS collaborations with ≈36 fb^{-1} achieved similar results [53, 66]. Figure 2.8 (right) shows the upper limits on the cross section of a Z' or W' as a function of the HVT triplet mass for a combination of VV and VH all-hadronic and semi-leptonic channels.

The most stringent lower limits on the mass of the signal hypotheses described in Sec. 2.3 are summarized in Table 2.1.

Model		Expected limit (TeV)	Observed limit (TeV)	Ref.
Radion gg	WW	3.1	3.1	[63]
Radion VBF	WW	-	-	[61, 63]
Radion gg	ZZ	2.5	3.0	[62]
Radion VBF	ZZ	-	-	[61]
Radion gg	VV	2.9	3.2	[61]
Radion VBF	VV	-	-	[61]
HVT model B W'	WZ / WH	4 / 3.8	4.3 / 4	[61]/[63]
HVT model B Z'	WW / ZH	3.9 / 3.7	4 / 3.7	[63]/[64]
HVT model B V'	VV / VH	3.5 / 3.1	3.8 / 2.50 and 2.76–3.30	[58]/ [59]
HVT model B V'	VV+VH	4.2	4.5	[53]
HVT model C W'	WZ / WH	- /	- /	[61, 63] /
HVT model C Z'	WW / ZH	- / -	- / -	[61, 63] / [64]
HVT model C V'	VV/ VH	/	/	
HVT model C V'	VV+VH			
G_{bulk} $\tilde{\kappa} = 0.5$ gg	WW	1.7	1.8	[63]
G_{bulk} $\tilde{\kappa} = 0.5$ VBF	WW	-	-	[63]
G_{bulk} $\tilde{\kappa} = 0.5$ gg	ZZ	1.1	1.2	[62]
G_{bulk} $\tilde{\kappa} = 0.5$ VBF	ZZ	-	-	[62]
G_{bulk} $\tilde{\kappa} = 0.5$ gg	VV	1.1	0.85	[53]
G_{bulk} $\tilde{\kappa} = 0.5$ VBF	VV			

TABLE 2.1: Summary of the lower limits on the mass of the signal hypotheses described in Sec. 2.3. They are also all considered in this thesis. The symbol - indicates that no limit has been placed despite (at least) one search has been performed; if more than one search has been executed, the most sensitive in terms of cross section are reported. If no analysis has been carried out in one of the channels, the channel is left empty.

3 Experimental setup

Many of the BSM phenomena introduced in Sec. 2.3 could manifest at high energy scales. Reaching the TeV energy range in a controlled manner requires powerful particles' accelerator machines. The data analysed in the first part of this thesis are produced in proton-proton collisions at the Large Hadron Collider (LHC) with a center-of-mass energy of 13 TeV. They have been collected with the Compact Muon Solenoid (CMS) experiment between 2016 and 2018 for a total integrated luminosity of 138 fb^{-1}.

In the first part of this chapter, the LHC complex and working principle are introduced, then the CMS experiment is described.

In Sec. 3.3.3 studies performed to validate the introduction of radiation damage effects in the detector simulation are presented. This work has been performed under the supervision of Dr. Tatjana Šuša and Dr. Tamás Almos Vámi. The radiation damage contributions have been estimated by Prof. Morris Swartz and the integration in the simulation framework has been performed by Dr. Paul Schütze. The author of this thesis produced different simulation samples to verify the correct implementation and functioning of the framework, performed comparison with data and monitored the status of the collected data.

3.1 The Large Hadron Collider

The LHC [67, 68] is a hadron collider composed of superconducting accelerators, covering a total circumference of 26.7 km. It is located $\approx$ 100 m underground at the Conseil Européen pour la Recherche Nucléaire (CERN), between France and Switzerland. It can accelerate and collide protons and heavy-ions; only the first mode is relevant for this thesis and discussed in the following.

Powerful machines are required to accelerate such heavy particles and the LHC is, in fact, only the final acceleration stage. Protons are extracted from hydrogen atoms and inserted in a chain of linear and circular accelerators. They have an energy of 450 GeV at the injection in the LHC. Here protons circulate in opposite directions in two beampipes, kept in ultra-high vacuum conditions (10^{-10} to 10^{-11} mbar) to prevent collisions with air molecules. The LHC is composed of alternating arcs and straight sections. The arcs host the 1232 superconducting dipole magnets, which bend the protons into the circular trajectory. They are made of niobium-titanium, cooled at a temperature of 2 K to operate in superconducting mode, and reach a 8.33 T magnetic field. Quadrupole magnets, focusing and squeezing the beam, and other higher-order magnets, are also located in the arcs. The radio-frequency (RF) cavities, whose electromagnetic fields oscillate with a frequency of 400 MHz, accelerate the protons up to 6.5 TeV per beam, for a total center-of-mass energy of 13 TeV. It is designed to ultimately reach 7 TeV per beam and, thus, a center-of-mass energy of 14 TeV. The LHC has four collision points where the main experiments are

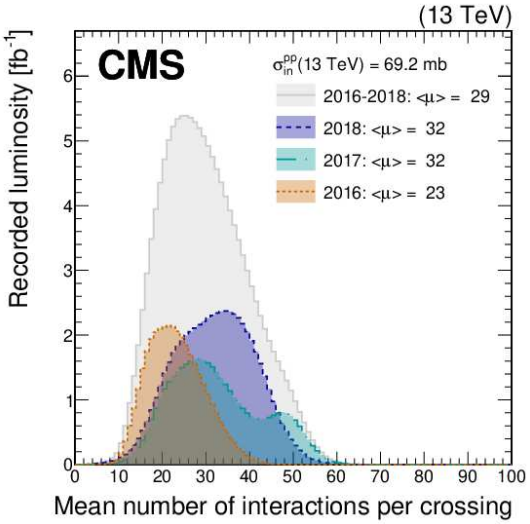


FIGURE 3.1: Distribution of the mean number of inelastic interactions per crossing (pileup) in data for pp collisions in 2016 (dotted orange line), 2017 (dotted dashed light blue line), 2018 (dashed navy blue line), and integrated over 2016-2018 (solid grey line). A total inelastic pp collision cross section of 69.2 mb is chosen. The mean number of inelastic interactions per bunch crossing is provided in the legend for each year. Taken from Ref. [81].

located: ATLAS (A Toroidal LHC ApparatuS) [69] and CMS (the Compact Muon Solenoid) [70] are two general-purpose detectors used for a large variety of studies, ALICE (A Large Ion Collider Experiment) [71] is specialized in heavy-ions physics and LHCb (LHC beauty) [72] performs precision measurements of CP violation and rare decays of B hadrons. Moreover, three other smaller experiments are also located at the LHC: LHCf [73], MoEDAL [74] and TOTEM [75].

Together with the beam energy and the type of accelerated particle, the instantaneous luminosity $\mathcal{L}_{inst}$ is one of the most important characteristic of a collider. It describes the frequency of the interactions among the bunches in the beams: considering two Gaussian beams colliding head-on

$$\mathcal{L}_{inst} = \frac{N_1 N_2}{4\pi\sigma_x\sigma_y} f N_b$$

where N_1, N_2 are the numbers of particles in each beam, f the revolution frequency, N_b the number of bunches and σ_x (σ_y) the standard deviation of the beams Gaussian profile in the x (y) transversal direction. Beam crossing angle and offset can then be taken into account introducing corrective factors [76]. At the LHC, each bunch has $\approx 10^{11}$ protons and the maximum number of bunches per fill is 2808. The bunch diameter is reduced down to 16 µm by the focusing magnets while the collisions have a frequency of 40 MHz, i.e. they happen every 25 ns. The LHC instantaneous design luminosity of 10^{34} s^{-1} cm^{-2} was exceeded in 2018. The mean number of interactions per bunch crossing (pileup) is presented in Fig. 3.1.

If the cross section for a particular type of event is σ, the number of events N occurring in a certain time t is given by $N = \sigma \int_0^t \mathcal{L}_{inst} dt$ [77], where $\int \mathcal{L}_{inst} dt$, the integrated luminosity, is the amount of collected data. CMS collected 36 fb^{-1} [78], 42 fb^{-1} [79] and 60 fb^{-1} [80] of data in 2016, 2017 and 2018, respectively, for a total of 138 fb^{-1}.

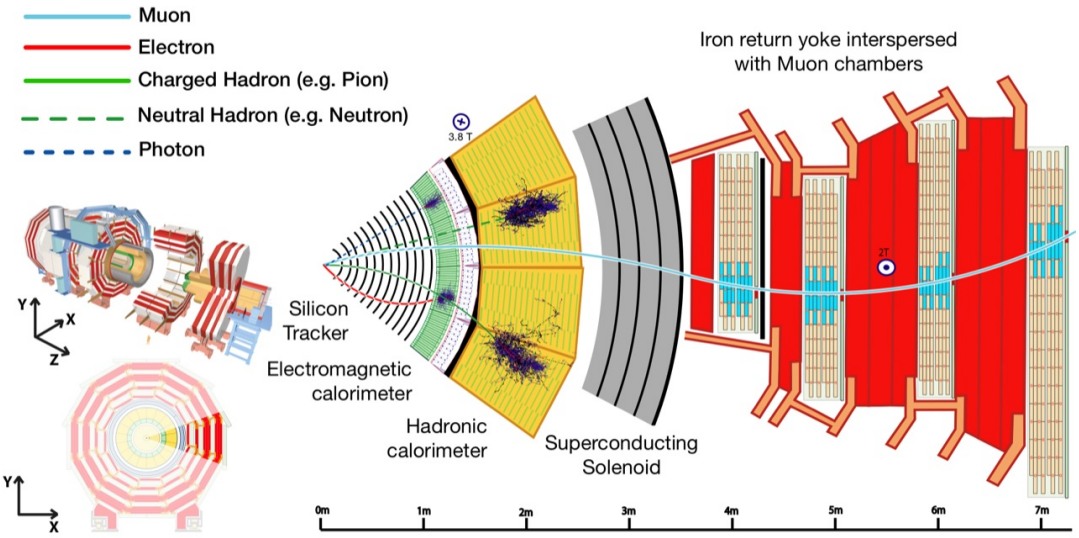


FIGURE 3.2: Schematic views of the CMS experiment. Left: 3-dimensional expanded view of the CMS experiment and front view of the experiment. The highlighted slice is enlarged on the right. Right: Slice of the CMS apparatus with the main sub-systems. The different particle interactions in the sub-detectors are also drawn. Adapted from Ref. [82].

3.2 The CMS experiment

The CMS detector is a multi-purpose apparatus [70] equipped with several sub-systems to measure different particle types and their properties. The sub-systems are arranged in an onion-like structure around the beam-pipe, as shown in Fig. 3.2. The innermost part is occupied by the pixel and strip silicon trackers, surrounded by the electromagnetic and hadronic calorimeters, all enclosed in a superconducting solenoid. The muon chambers, the outermost component, are placed in the return yoke of the magnet.

While a more detailed description of the tracking system is provided in Sec. 3.3, below the key characteristics of the other sub-components are summarized.

In the following, the same coordinate system adopted by CMS is used. The origin is centred at the nominal collision point inside the experiment, the x-axis points radially inward toward the LHC center, the y-axis points vertically upward, and the z-axis is defined along the beam direction toward the Jura mountains. The polar angle θ is measured from the z-axis and the pseudorapidity is defined as

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{3.1}$$

The azimuthal angle ϕ is measured from the x-axis in the x-y plane, and r denotes the radial coordinate in this plane. The momentum p_{T} and energy E_T transverse to the beam direction are computed from the x and y components.

3.2.1 Electromagnetic colorimeter

The Electromagnetic Calorimeter (ECAL) [83] measures the energy deposition from electromagnetic showers of photons and electrons. It is a hermetic and homogeneous calorimeter made of lead tungstate ($PbWO_4$) crystals with coverage in pseudorapidity up to $|\eta| < 1.479$ in the barrel region and $1.479 < |\eta| < 3$ in the end-caps. The crystals' granularity is $\approx 0.0174 \times 0.0174$ in η-ϕ. The scintillation light

is detected by Silicon Avalanche Photodiodes (APDs) in the barrel region and Vacuum PhotoTriodes (VPTs) in the endcap region. The detector thickness in radiation lengths is greater than $25X_0$. One of the driving criteria in the ECAL design was the capability to detect the $H \to \gamma\gamma$ decay; thus a preshower system is installed in front of the endcap for $\pi^0 \to \gamma\gamma$ rejection.

The energy resolution is given by

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

where S is the stochastic term from the shower development, N is the noise contribution and C a constant term depending on the calibration. The reported values refer to electrons' energy resolution.

In 2017, the ECAL endcap region $2.5 < |\eta| < 3$ was affected by high noise levels induced by radiation damage and too low thresholds, mitigated at analysis level.

3.2.2 Hadronic colorimeter

The Hadron Calorimeter (HCAL) [84] measures the energy deposition of showers initiated by hadrons and is thus a crucial component for studying processes involving hadronic jets, as well as neutrinos or other signatures resulting in apparent missing transverse energy. It is a brass/scintillator sampling calorimeter. The scintillation light is converted by WaveLength-Shifting (WLS) fibres embedded in the scintillator tiles and then channeled to photodetectors. This light is detected by Hybrid PhotoDiodes (HPDs) that can provide gain and operate in high axial magnetic fields. The barrel and endcap regions cover up to $|\eta| < 1.3$ and $1.3 < |\eta| < 3$, respectively. An iron/quartz-fibre forward calorimeter extends the coverage up to a pseudorapidity of 5.2. The Cherenkov light emitted in the quartz fibres is detected by photomultipliers. The forward calorimeter ensures full geometric coverage for the measurement of the transverse energy in the event. The granularity in η–ϕ is 0.087×0.087 and $\approx 0.17 \times 0.17$ for $|\eta| < 1.6$ and $|\eta| \geq 1.6$, respectively. The thickness in interaction lengths varies from 7 to 11 λ_I depending on η. The energy resolution of the hadronic calorimeter is

$$\left(\frac{\sigma}{E}\right)^2 \approx \left(\frac{a}{\sqrt{E}}\right)^2 + b^2,$$

where b = 5% and a depends on the region: 65% in the barrel, 85% in the endcaps, and 100% in the forward region.

The HCAL endcap was upgraded in 2018 [85], substituting the HPDs with Silicon PhotoMultipliers (SiPMs) and increasing the number of readout channels.

3.2.3 Superconducting solenoid

The magnet [86] provides the large bending power needed for a precise measurement of the momentum of high-energy charged particles and thus played an important role in the design of the detector layout. In CMS, a 3.8 T magnetic field is achieved by a superconducting niobium-titanium solenoid with a length of 12.9 m and an inner diameter of 5.9 m to accommodate the tracker and the calorimeters inside. The magnetic field lines are closed by the return yoke located in the outer part of CMS. The yoke is composed of 6 endcap disks and 5 barrel wheels for a total of several ktons of magnetized iron blocks alternated to the muon chambers.

3.2.4 Muon chambers

The Muon Spectrometer [87] is used to identify muons and assist the tracker in measuring the momentum and charge of high-p_T muons. At high momenta ($\approx$ 1 TeV), both detector parts together yield a momentum resolution of about 5%. It is composed of three types of gaseous particle detectors covering different η ranges characterized by a different strength of the magnetic field and muon flux:

- Drift Tube chambers (DTs) cover the barrel region $|\eta| < 1.2$ where a smaller number of muons is expected and the magnetic field is less strong. They provide a spatial resolution of 100 µm in the r-ϕ plane and 150 µm in the longitudinal z coordinate. The particle direction is measured with a precision of 1 mrad.
- Cathode Strip Chambers (CSCs) are placed in the endcap region $0.9 < |\eta| < 2.4$, where the muon flux is larger. They are characterized by a resolution of 75–150 µm in the r-ϕ plane.
- Resistive Plate Chambers (RPCs) are placed in both the barrel and the endcap regiioins, at $|\eta| < 1.6$. With their fast response of 2–3 ns and a 1 ns time resolution, the RPCs are used as an additional triggering system.

3.2.5 Trigger system

The trigger system is used to reduce the large amount of data ($\approx$ 40 TB/s) produced by the pp-collisions happening in the LHC, which cannot be entirely stored. A total rate reduction of a factor $\approx 10^6$ is achieved in two steps, the Level-1 (L1) Trigger [88] and the High-Level Trigger (HLT) [89]. The L1 Trigger consists of custom-designed, programmable electronics and uses coarsely segmented data from the calorimeters and the muon system. The HLT is a software system with access to the read-out data and can therefore perform complex calculations.

3.3 The Tracking system

The tracking system [90, 91] is used to precisely measure the charged particles' trajectories and reconstruct the primary interaction and the secondary vertices. Charged hadrons constitute roughly two-thirds of the particles produced in pp-collisions; thus their reconstruction plays a fundamental role in the physics program of the experiment. Furthermore, the identification of secondary vertices is essential, for instance, in the reconstruction and tagging of particle sprays originated from the hadronization of the b-quark. The tracking system is the innermost detector, placed around the interaction point where the intense particle flux causes severe radiation damage. The tracking system must feature good position and momentum resolution and high reconstruction efficiency to fulfil its purposes. Additionally, it must have a high granularity to provide low occupancy per bunch crossing, the number of particles coming through a detector cell for each event, and has to be radiation hard. Due to these requirements, the tracking system is completely made of silicon based detectors with decreasing granularity from inside out, pixel sensors in the inner part and strip sensors in the outer one. The two sub-detectors are divided into a barrel part ($|\eta| < 1.3$) and an endcap part ($1.3 < |\eta| < 2.5$).

Promptly produced, isolated muons of $p_T > 0.9$ GeV are reconstructed with $\approx$ 100% efficiency. In the central region, muons of p_T = 100 GeV have resolutions

of approximately 2.8% in p_T, and 10 and 30 µm in transverse and longitudinal impact parameters, respectively. Additionally, tracks with $1 < p_T < 10$ GeV have a resolution in p_T of $\approx 1.5\%$. The resolution in their transverse (longitudinal) impact parameters improves from 90 µm (150 µm) at $p_T = 1$ GeV to 25 µm (45 µm) at $p_T = 10$ GeV. For vertices with many tracks, characteristic of interesting events, the achieved vertex position resolution is 10–12 µm in each of the three spatial dimensions [92].

The properties of silicon detectors, relevant for the second part of this thesis, are described in Chapter 12. In this section, the characteristics of the detectors employed for the collection of the data analyzed in the first part of this thesis are summarized.

3.3.1 The strip detector

The silicon strip tracker has a surface of $\approx 198\,\mathrm{m}^2$, counting 15′148 modules for a total of 24′244 sensors and 9.3 million strips. Due to the large extension, the working conditions vary considerably over the strip detector volume, and different sensor designs are required. The Tracker Inner Barrel (TIB) and Tracker Inner Disks (TID) extend in radius between 20 and 55 cm. The former is composed of 4 layers, whilst the latter has 3 disks at each end. The Tracker Outer Barrel (TOB) surrounds the Inner Tracker and extends radially between 55 and 116 cm. It consists of 6 layers. Finally, the Tracker Endcap (TEC) extends radially between 22.5 and 113.5 cm. Each TEC is composed of 9 disks, carrying up to 7 rings of silicon microstrip detectors.

The strips are p^+ implants in a n substrate, oriented along the z direction in the barrel and along the r coordinate in the endcaps. In TIB and TID the sensors are 320 µm thick and 10 cm long. The strip pitches range from 80 to 120 µm. In the outer sub-detectors, the strips are 25 cm long and have 97–183 µm pitch. The sensor thickness is 320 µm and 500 µm for TID and TEC, respectively.

The microstrip detector provides a single hit spatial resolution measurement of about 20–50 µm in the transverse plane and about 200–500 µm along the longitudinal coordinate, depending on the number of fired strips and their pitch [92].

3.3.2 The pixel detector

The silicon pixel detector is the innermost part of the tracking system. It is located around the beam pipe and, therefore, close to the interaction point. It consists of a barrel part (BPIX) and two endcaps in the forward region (FPIX).

The pixels sensors are made in n-in-n technology consisting of n-implants in an n-substrate, while the backside of the sensor is p-doped. They are 285 µm thick. The pixel cell dimensions are $100 \times 150\,\mu\mathrm{m}^2$, to achieve similar track resolution in both r–ϕ and z directions.

The original pixel detector (Phase-0), conceived to cope with the LHC design luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, was installed already at the beginning of the first data taking in 2010. In 2016, the ageing of the detector caused by radiation damage resulted in a reduced hit efficiency. Thus, an upgraded version of the pixel detector (Phase-1) [93, 94] was installed during the Extended Year-End Technical Stop (EYETS) between the 2016 and 2017 data taking periods [95]. While the original detector had three layers in the barrel and two in the endcaps, the upgraded system features one additional pixel layer in the barrel region and one additional endcap disk per side. The innermost layer and disks of the Phase-1 detector are located closer to the beam pipe to improve the measurement of the interaction vertices. The performance is additionally enhanced by the material budget reduction and the larger bandwidth

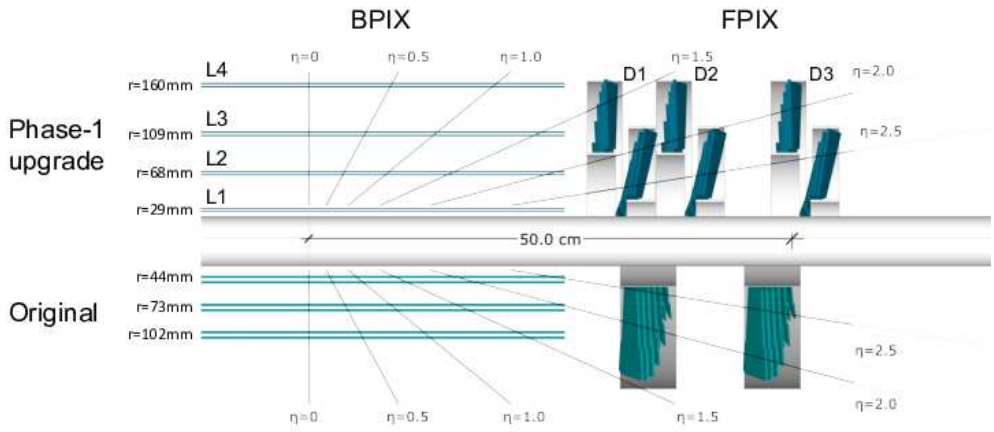


FIGURE 3.3: Layout of the Phase-1 pixel detector installed during the Extended Year-End Technical Stop between the 2016 and 2017 data taking periods compared to the original detector layout used till the 2016 data-taking period included, in longitudinal view. Taken from Ref. [98].

of the readout chips [96, 97]. A comparison between the original and upgraded versions of the pixel detector can be found in Fig. 3.3.

Together with the efficiency, the position resolution is the most important performance parameter of the detector. The measured value of the position resolution depends on the track angle and the radial position of the layer. In the Phase-0 (Phase-1) BPIX detector, the fitted width of the residual distribution, i.e. the distribution of the residual difference between the measured and the reconstracted track position, is 13.27 (9.5) µm in the r–ϕ direction and 34.08 (22.2) µm in the z direction [99] ([98]).

3.3.3 Pixel radiation damage and simulation

During the 2016–2018 operation period, the LHC has exceeded its design luminosity. On the one hand, this has allowed the experiments to collect more precious data. On the other hand, it has already anticipated some of the challenges that the detectors will face in the HL-LHC era (Sec. 10.3.1).

The distribution of the average pixel charge as a function of the production depth of the charge carriers in the silicon substrate can be used to monitor the performance of the detector. For a non-irradiated and depleted detector, the profile is expected to be flat as the detector is fully efficient and all the created charge is collected, as shown in Fig. 3.4 (left), featuring a newly installed pixel detector. Severe radiation damage in the Phase-1 pixel detector alters its response, as presented in Fig. 3.4 (right). Here the detector has been operated for several months and the profile has a slope, caused by a reduced charge collection for the charge carriers created away from the collecting electrodes. This is due to the formation of trapping sites in the silicon substrate, as explained in Sec. 12.5, where the effects of radiation damage are illustrated.

In all the physics analyses performed by the CMS collaboration, data are carefully compared with simulation describing both the SM (or BSM) processes and the detector behaviour. Thus, a realistic description of the detectors' performance should include the effects of radiation damage. However, such effects, described in Sec. 12.5 for silicon sensors, are non-trivial, and the implementation must consider the computing time needed to produce the simulation. In the CMS simulation framework (see Sec. 4.2.2), the effects of radiation damage in the pixel detector have been

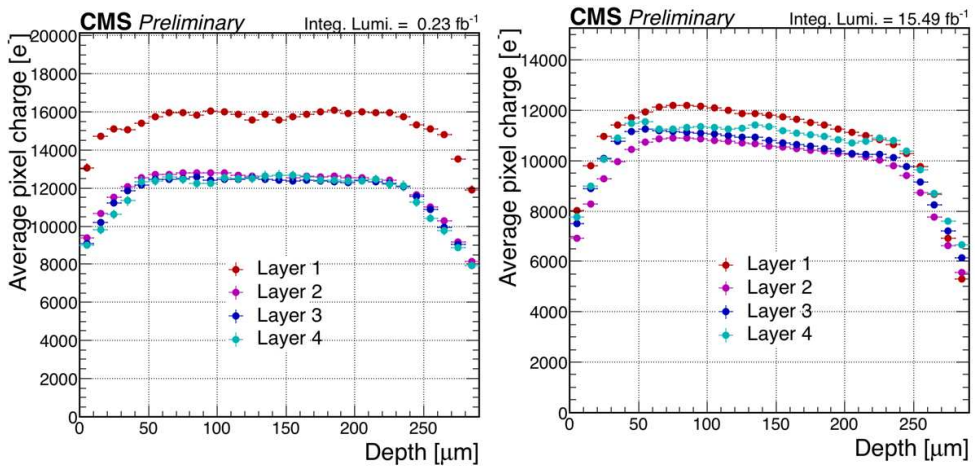


FIGURE 3.4: Average pixel charge as a function of the depth of the charge carriers creation in the silicon substrate for the four BPIX Phase-1 layers. On the left, the almost newly installed pixel detector presents a flat profile as the (fully depleted) detector is fully efficient, while on the right, the detector, after several months of operation, shows a reduction of the charge collection for the charge carriers created away from the collecting electrodes. Taken from Ref. [100].

recently taken into account, employing the *cluster charge reweighing* (CCR) method. This approach applies *templates* of previously simulated responses of a sensor to a traversing particle under various conditions [101] to the charges generated in the absence of radiation damage in the CMS simulation framework. Each template is the average of multiple simulations of a particle traversing the sensor with a certain angle and position within a pixel. The simulation is produced with the PIXELAV software [102], which randomizes the charge carriers creation, their drift, and the occurrence of delta electrons. A commercial semiconductor simulation code is used to generate a full three dimensional electric field map and to take into account the charge carrier trapping due to radiation damage, then provided to PIXELAV. The initially induced charge q_{in} of each pixel in a cluster is thus scaled with a weight obtained from the template pixel charges at the same position inside the cluster. The weight is calculated from the ratio of the template of an undamaged idealized sensor with a uniform electric field (t_0) and the template (t_ϕ) of a physical sensor with an irradiation dose ϕ. For any pixel, specified by its pixel coordinate (i,j) the re-weighted charge q_{out} is determined as [103, 104]

$$q_{\text{out}}(i,j) = \frac{t_\phi(i,j)}{t_0(i,j)} \cdot q_{\text{in}}(i,j).$$

The procedure has been successfully validated in this work and included in the legacy simulation of the data-taking period 2016–2018. Reference distributions obtained from tracks with a minimum p_{T} of 3 GeV forming large clusters in the pixel detector are shown in Fig. 3.5.

However, the templates have been derived for the average detector performance in a certain data-taking year and do not reflect the changes of performance developed from the beginning to the end of the data-taking period, as shown in Fig. 3.6 for the year 2018. An effort is currently ongoing in the collaboration to take such variations into account. Nevertheless, the results presented in Ref. [105] constitute proof that the CCR approach can include radiation damage in the CMS simulation without significantly affecting the computing time and that the templates can be

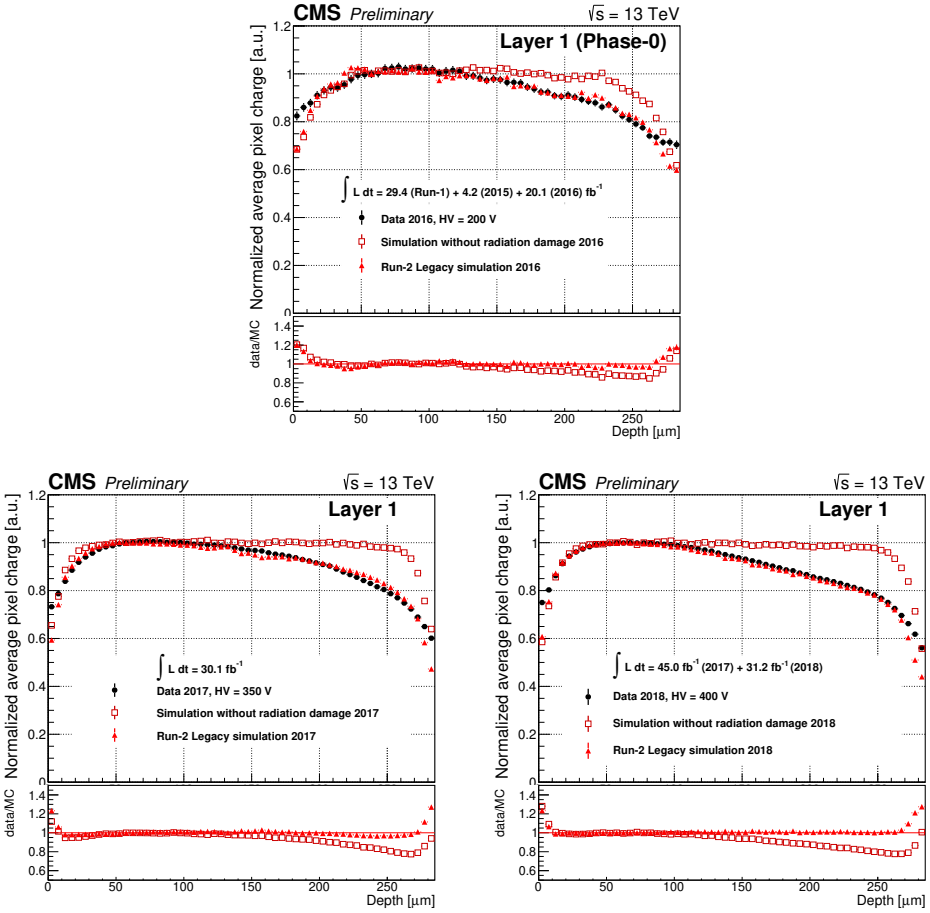


FIGURE 3.5: The normalised average pixel charge as a function of the production depth in the silicon substrate is shown for Layer 1 of the Barrel Pixel detector. It is compared to an ideal simulation where no radiation damage is implemented and to a realistic simulation where the pixel charge is re-weighted to take into account the radiation damage. On the top, the Phase-0 detector is shown in 2016 after 29.4 fb^{-1}, 4.2 fb^{-1}, and 20.12 fb^{-1} of data is collected in Run-1, 2015, and 2016, respectively. On the bottom left (right), the Phase-1 detector is shown in 2017 (2018) after 30.1 fb^{-1} of data is collected (after 45.0 fb^{-1} of data is collected in 2017 and 31.2 fb^{-1} of data is collected in 2018). Published in Ref. [105].

tuned to reflect the behaviour observed in the detector. These improvements will be especially relevant in the HL-LHC period, where the radiation damage is expected to be more significant. Furthermore, templates validated with data obtained with highly irradiated sensor from beam tests could be used to predict the performance of the detector.

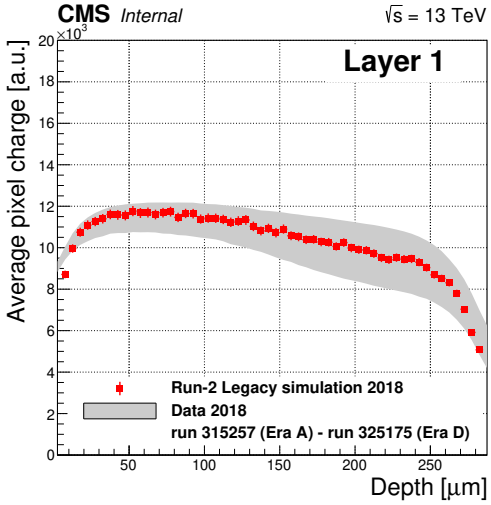


FIGURE 3.6: Normalised average pixel charge as a function of the production depth in the silicon substrate for Layer 1 of the Barrel Pixel detector in 2018. The gray band, representing the data over the full 2018 data-taking period, is compared to the legacy 2018 simulation tuned on the average behaviour.

4 Event and object reconstruction and simulation

To perform precise measurements of the SM properties, or to probe BSM theories, data are carefully compared with theoretical predictions. The description of SM (or BSM) processes is simulated from the probability distributions to obtain specific outcomes from the pp-collisions. To properly compare data and theoretical models, the detector effects must be taken into account. Also, particle signatures recorded in the detectors must be interpreted and classified. This chapter first describes the physics of proton-proton collisions (Sec. 4.1), then the event simulation is presented (Sec. 4.2) and successively the event reconstruction is explained (Sec. 4.3). Last, techniques used to mitigate pileup are reported. The analysis contained in this thesis studies all-jets final states; thus, Chapter 5 is dedicated to jet reconstruction and methods to exploit jet substructure.

4.1 Physics of proton-proton collisions

A description of the proton composed of *partons* is necessary to characterise high-energy pp-collisions. In the parton model [106], the *parton distribution function* (PDF) is the probability distribution $f_i(x, \mu_F^2)$ that the parton i, either a quark or a gluon, carries the fraction x of the proton's momentum. The PDF depends on the factorization scale μ_F^2 at which a hadron is being probed. Protons are composed of the three *valence quarks* uud, the gluons (mediators of the strong force), and the *sea quarks* (quark-antiquark pairs with smaller x).

The pp-interaction cross section $\sigma(\mathrm{pp} \to X)$, for energies significantly larger than the binding energy of the partons in the proton, can be calculated with the *factorization theorem* [107] from the partonic cross sections $\hat{\sigma}_{ij\to X}$ as

$$\sigma(\mathrm{pp} \to X) = \sum_{i,j} \int dx_1 dx_2 f_{i,\mathrm{p}}(x_1, \mu_F^2) f_{j,\mathrm{p}}(x_2, \mu_F^2) \hat{\sigma}_{ij\to X}(\hat{s}, \mu_R^2, \mu_F^2, \alpha_s).$$

The sum runs over all possible initial-state partons i, j, with longitudinal momentum fractions $x_{1,2}$, giving rise to a final state X at a centre-of-mass energy of $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$. The partonic cross sections can be calculated in perturbation theory. The renormalization scale μ_R^2 and the factorization scale μ_F^2 appear in the expressions of the $\hat{\sigma}_{ij\to X}$ calculated from truncated expansions in the strong coupling constant α_s. On the contrary, PDFs cannot be predicted by perturbative QCD as they also contain the low energy information of the scattering, which is non-perturbative. Proton's PDFs are obtained from fits to deep inelastic scattering data, $\mathrm{p\bar{p}}$- and pp-interactions, for

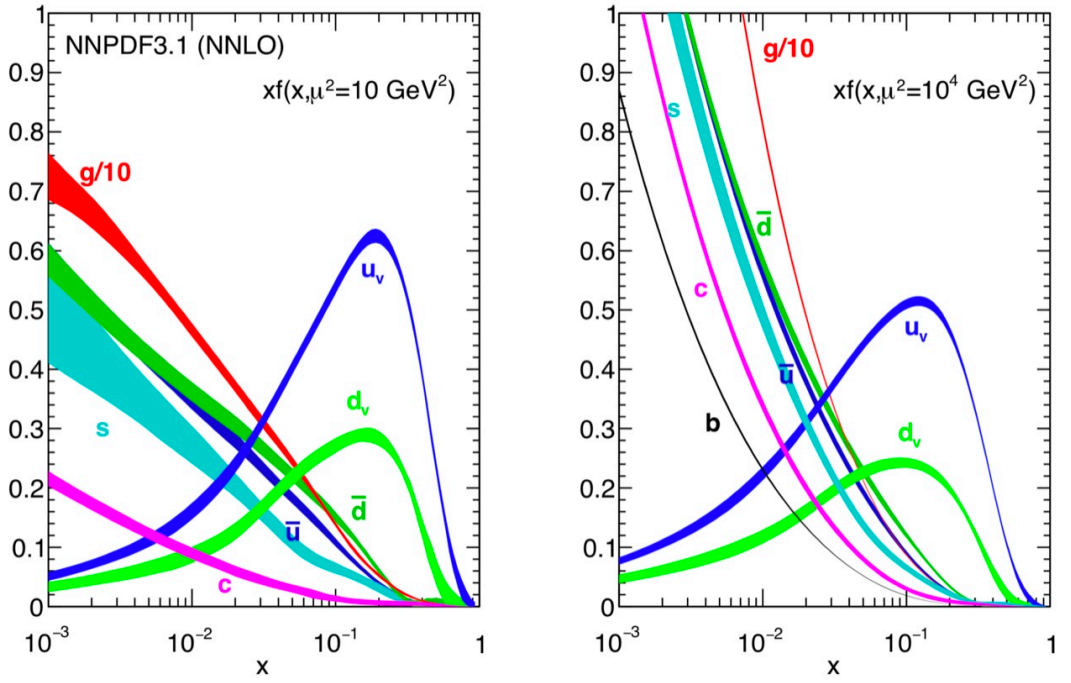


FIGURE 4.1: PDFs $f(x, \mu_F^2)$ multiplied by x at $\mu_F^2 = 10\ \text{GeV}^2$ (left) and $\mu_F^2 = 10^4\ \text{GeV}^2$ (right), obtained in NNLO NNPDF3.1 global analysis [111].

a certain momentum fraction x_i at an energy scale μ_F^2. They are extrapolated at different scales using the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) equations [108–110]. For the simulation used in this analysis the NNPDF3.1 PDFs [111], shown in Fig. 4.1, and the previous version NNPDF3.0 [112] are used.

4.2 Event simulation

Due to the composite nature of the protons, the description of complete final states at hadron colliders involves multi-particle calculations. The simulation of pseudo-events in such a high-dimensional phase space can be achieved through Monte-Carlo generators. It can be divided into several steps corresponding to the different sub-processes (*factorization*), as shown in the schematic representation in Fig. 4.2.

The *hard process* indicates the parton interaction with the highest momentum transfer. The hard cross section $\hat{\sigma}_{ij\to X}$ is determined in perturbation theory using numerical methods. The calculation can be performed at tree-level, referred to as leading order (LO) in the strong coupling α_s, or incorporating further gluons or quarks emissions, referred to as next-to-leading (NLO) or next-to-NLO (NNLO) corrections.

Higher-order terms are approximated using *parton showers*, programs designed to simulate the emission of additional soft and collinear partons. For a hard configuration with a parton of flavour i and cross section σ_0, the cross section that the hard configuration is accompanied by a parton j with momentum fraction z is [113]:

$$\mathrm{d}\sigma \approx \sigma_0 \sum_{\text{partons},i} \frac{\alpha_s}{2\pi} \frac{\mathrm{d}\theta^2}{\theta^2} \mathrm{d}z P_{ij}(z), \tag{4.1}$$

where θ is the opening angle between the partons and $P_{ij}(z)$ is a set of functions describing the parton emission probability in the collinear approximation. Parton

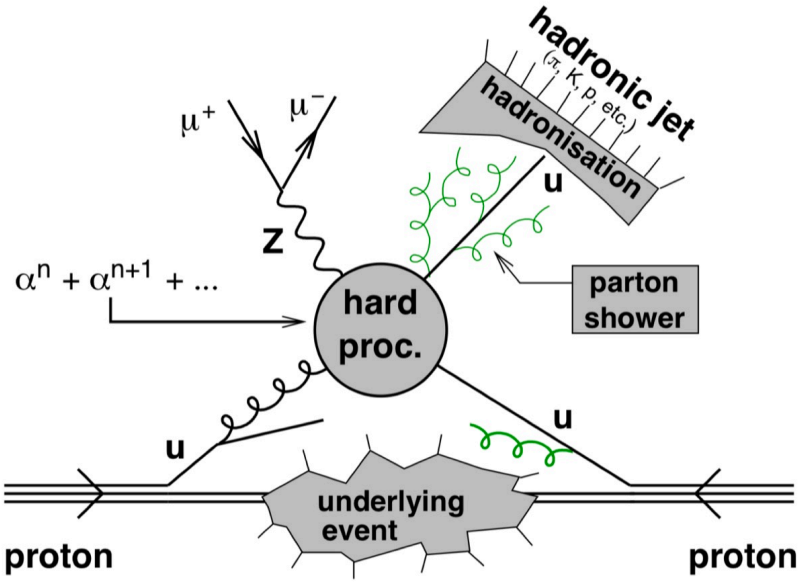


FIGURE 4.2: Schematic representation of a proton-proton collision, involving a quark-gluon scattering that leads to a final state consisting of two muons originated from a Z boson and a hard jet. Taken from Ref. [107].

shower algorithms repeat iteratively the parton splitting described in Eq. 4.1. The cascade is stopped once the energy scale reaches the cut-off scale ($\approx$ 1 GeV), where the confinement of partons into hadrons begins. Soft and collinear divergences are also taken into account. The first is due to the generation of many particles with very low energies. However, such particles cannot be measured below a certain value in the experiment. The collinear divergences are caused by collinear particles that cannot be distinguished with the experimental finite resolution. Thus, introducing a resolution parameter representing the scale below which the splitting is not resolved removes the divergences.

A *matching* procedure is applied to avoid double counting of partons when interfacing the hard scatter matrix element with the parton showering algorithm [114].

Partons that do not directly concur in the hard process produce soft, non-perturbative, secondary interactions, called *underlying event* (UE) and are simulated using phenomenological models tuned to the data [115].

At the scale where quarks and gluons cannot be considered free any more, the *hadronization* process, where partons form color-neutral hadrons, is reproduced through non-perturbative phenomenological models. The two main hadronization approaches are the string [116, 117] and cluster [118] ones. The former assumes that quark-antiquark pairs are subject to a linear potential (*string*) and when the energy stored in the string is larger than that of a quark-antiquark pair, a new pair is produced and the string fragments. Thus, the string algorithm transforms partonic systems directly into hadrons, while the other approach employs an intermediate stage of cluster objects. Both methods have free parameters that can be tuned to better reproduce the data.

Finally, extra pp-inelastic interactions are added to take into account the additional particle production due to pileup.

4.2.1 MC generators

Several MC generators are available for the matrix element calculation, such as MADGRAPH5_aMC@NLO [119] and POWHEG [120]. Others are multi-purpose generators that can be interfaced with the event generators for the showering and hadronization processes or used for the full chain of processes, as, for instance, PYTHIA [121] or HERWIG++ [122]. MADGRAPH5_aMC@NLO can be used to calculate LO matrix elements and NLO QCD corrections of SM processes, in association with parton-shower matching. Interfaced with additional packages, it provides also BSM processes generation. POWHEG provides the hardest process with NLO accuracy, while softer ones are calculated at LO precision. PYTHIA generates the matrix element at LO accuracy. Thus, it is often interfaced with NLO generators to provide the (p_{T}-ordered) parton showering step, the hadronization (with string model) and the underlying event's modelling. HERWIG++ uses a different (angular) ordering in the parton showering step and the hadronization is described with the cluster model.

Appendix A provides a complete description of simulated and data samples used for the analysis described in Chapters 6–8. More details on the corrections applied to LO simulations to take into account NLO effects can also be found in Appendix A. Here the generators used for the main samples are listed:

- **BSM signals** are generated at LO of MADGRAPH5_aMC@NLO versions 2.4.2 and 2.6.0 matched with PYTHIA versions 8.205 and 8.230, for 2016 and 2017-2018 detector conditions, respectively. The NNPDF 3.0 and 3.1 LO PDFs are used together with the CUETP8M1 [123] and CP5 [124] underlying event tunes in PYTHIA for 2016 and 2017-2018 conditions, respectively. The cross section of the signal samples is computed at NLO with MADGRAPH5_aMC@NLO with PDF4LHC15_100 [112, 125–129].

- **QCD multijet production** is generated with three different configurations
  - PYTHIA standalone
  - the LO mode of MADGRAPH5_aMC@NLO [130] matched and showered with PYTHIA
  - HERWIG++ 2.7.1 [122] with the CUETHS1 tune [123].

- **Top quark pair production**, as well as **single top** and **boson pair production**, are modelled at NLO with POWHEG [131] and showered with PYTHIA. To calculate systematic uncertainties related to the vector boson tagging efficiency, two additional simulated samples of top quark pair production at LO are used: one generated with MADGRAPH5_aMC@NLO and interfaced with PYTHIA, and the second one generated and showered with PYTHIA.

- **W+jets and Z+jets production** are simulated with the LO mode of MADGRAPH5_aMC@NLO matched and showered with PYTHIA. The NNPDF 3.1 [111] PDFs are employed for simulated V+jets events with the 2017-2018 data taking conditions for the 2016-2018 data analysis.

The background samples use the same underlying event tunes listed for the signal samples. Two corrections dependent on the transverse momentum (p_{T}) [132, 133] are applied to the V+jets backgrounds to correct the p_{T}-distribution of the vector bosons computed at LO in QCD to the one predicted at NLO in QCD, and to account for electroweak effects at high p_{T}.

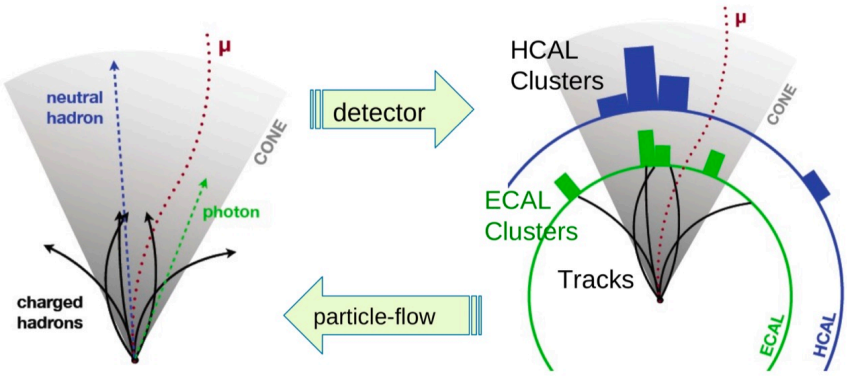


FIGURE 4.3: Schematic view of the particle-flow concept: Signals from different sub-detectors are combined to construct objects that represent individual particles with defined types. Taken from Ref. [137].

4.2.2 CMS detector simulation

Finally, MC generators are interfaced with the GEANT4 toolkit [134, 135] to carry out a detailed simulation of the CMS detector. The GEANT4 software takes as input the particles simulated by the MC generators, calculates their trajectories and the interaction with the different CMS sub-systems, whose accurate description is also given to the software. The description includes the detectors' geometries and materials and information on the status of the sub-systems.

Furthermore, a weight is applied to simulated events such that their pileup distribution matches the data distribution, estimated from the total inelastic cross section and the instantaneous luminosity.

4.3 Particle-Flow event reconstruction

The event reconstruction consists of identifying the final state particles originating in the pp-collision from their characteristic signatures in the different sub-detectors. The CMS Collaboration adopts the *Particle-Flow* (PF) [136] reconstruction approach. Figure 4.3 schematically illustrates the PF concept, where information obtained from different elements is combined to classify the final state particles and objects, and precisely measure their kinematic properties. Each particle produces several *PF elements* in the various CMS sub-detectors, which are then connected by a *link algorithm* to form *PF blocks*. Finally, particles are reconstructed in sequence from the list of blocks, starting from muons which can be more easily identified. Then electrons, photons, and charged and neutral hadrons are classified. At last, jets are formed from *PF candidates*.

Below, the event reconstruction's steps are described, more details are provided for the elements relevant for the studies presented in this thesis.

4.3.1 PF elements

4.3.1.1 Reconstruction of tracks and vertices

Tracks are reconstructed from hits of charged particles in the pixel and strip tracker [92]. Hits are obtained clustering signals above specified thresholds (*local reconstruction*). To improve the position estimation of the hits in the pixel detector, the observed distribution of the cluster charge is compared to template distributions taking into

account the radiation effects, estimated with the approach introduced in Sec. 3.3.3. The subsequent step of the track reconstruction procedure consists of the translation between the local coordinate system of the hits and the global coordinate system of the track, taking into account discrepancies between the assumed and actual location and surface deformation of detector elements as found through the alignment process [138, 139].

To obtain a high track-finding efficiency while limiting the reconstruction of fake tracks, an iterative tracking procedure is employed. The CMS tracking algorithm [140], called Combinatorial Track Finder (CTF), is an adaptation of the Kalman filter [141] to allow pattern recognition and track fitting to occur in the same framework. Track finding is performed in multiple stages, repeated at each iteration:

- Seed generation from 2 or 3 hits compatible with a charged-particle trajectory
- Kalman filter based track finding to extrapolate the seed trajectories along the expected flight path of a charged particle and to search for additional hits
- Track-fitting module to provide the best possible estimate of the parameters of each trajectory
- Track selection setting quality flags and discarding tracks that fail specific requirements.

The first iterations target high-p_{T} tracks, which are easier to reconstruct and can then be removed to reduce the combinatorial of the following steps, that features more complex and time-consuming seeding, filtering, and tracking algorithms. The average track-reconstruction efficiency for promptly-produced charged particles with p_{T} > 0.9 GeV is 94% (85%) for $|\eta| < 0.9$ ($0.9 < |\eta| < 2.5$). The inefficiency is caused mainly by hadrons that undergo nuclear interactions in the tracker material. For isolated muons, the corresponding efficiencies are $\approx 100\%$ [92].

Figure 4.4 shows the reconstruction efficiency and misreconstruction rate for charged hadrons in a sample of simulated QCD multijet events as a function of the reconstructed track p_{T}. Charged hadrons that are not reconstructed in the tracker can be detected as neutral hadron by the calorimeters, with degraded energy resolution and biased direction due to the bending of their trajectory in the magnetic field. Charged hadrons contribute two-thirds of a jet's energy; thus, increasing the tracking efficiency is essential. While iterative tracking largely improves the performance of the track reconstruction algorithm, the tracking efficiency at high-p_{T} remains limited. Nevertheless, good jet energy and angular resolutions are guaranteed by the calorimeters. Charged hadrons reconstruction at high-p_{T} will be considerably improved in the Phase-2 Upgrade of the CMS tracker, as described in the studies presented in Chapter 11.

Reconstructed tracks are used for the reconstruction of the primary vertices [142] of all the pp-interactions in one event, identifying the hard scatter and the pileup collisions. First tracks consistent with being produced in the primary interaction region are identified. Then tracks are clustered depending on their z-coordinates at their point of closest approach to the centre of the beam spot, allowing the reconstruction of any number of pp-interactions in the same LHC bunch crossing. The coordinates of candidate vertices containing at least two tracks are then obtained using an *adaptive vertex fitter* [143]. The primary vertex reconstruction efficiency is extremely high independently of the number of tracks associated with the vertex, $\approx 100\%$ and 98% for vertices with more than two tracks and with only two tracks, respectively. Instead, the resolution in the vertex coordinates strongly depends on

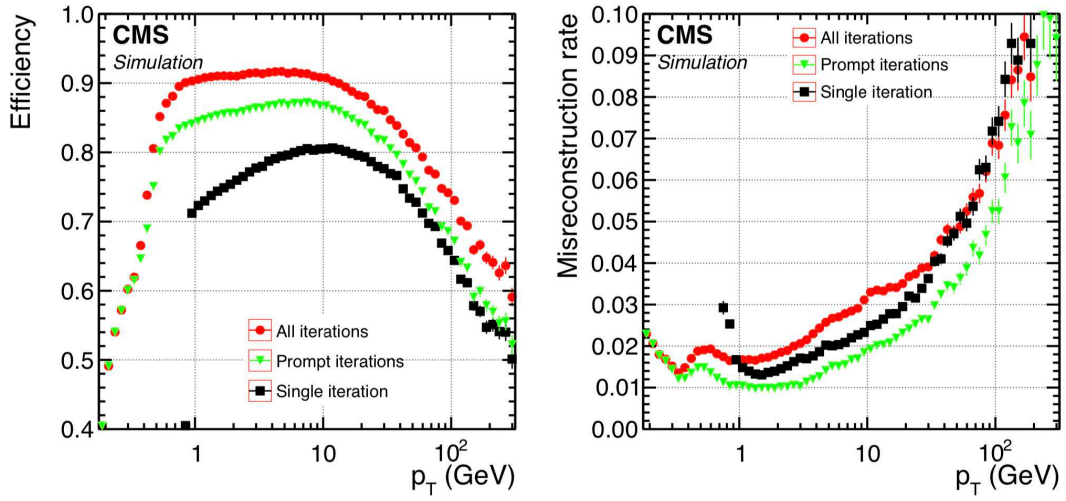


FIGURE 4.4: Efficiency (left) and misreconstruction rate (right) of the global combinatorial track finder (black squares); and of the iterative tracking method (green triangles: prompt iterations based on seeds with at least one hit in the pixel detector; red circles: all iterations, including those with displaced seeds), as a function of the track p_T, for charged hadrons in multijet events without pileup interactions. Only tracks with $|\eta| < 2.5$ are considered in the efficiency and misreconstruction rate determination. The efficiency is displayed for tracks originating from within 3.5 cm of the beam axis and ± 30 cm of the nominal centre of the CMS detector along the beam axis. Taken from Ref. [136].

the number of associated tracks and on the tracks' average p_T. For vertices with many tracks, a position resolution between 10–20 µm is achieved [92]. The reconstructed vertex with the largest value of summed physics-object p_T^2 is taken to be the primary pp interaction vertex (leading vertex, LV, in the following). The physics objects are the track jets, clustered using the anti-k_T algorithm [144, 145] jet finding algorithm (see Sec. 5.1) with the tracks assigned to the vertex as inputs, and the associated missing transverse momentum, taken as the negative vector sum of the p_T of those track jets. The other vertices are referred to as pileup (PU) vertices.

4.3.1.2 Energy reconstruction

The energy reconstruction in the ECAL and HCAL calorimeters is performed through clustering algorithms serving multiple purposes:

- Detection and measurement of the energy and the direction of stable neutral particles (photons and neutral hadrons)
- Distinguish the neutral particles' energy deposits from the charged hadrons ones
- Reconstruct and identify electrons and all accompanying bremsstrahlung photons
- Assist the energy measurement of charged hadrons with low-quality or high-p_T tracks.

Clustering is performed separately in each sub-detector. Cells with energy deposition above a specified threshold and higher than the depositions in neighbouring cells are considered as clusters seeds. *Topological clusters* are then formed by aggregating surrounding cells with significant energy depositions to the cluster seed.

Clusters within a topological cluster are identified with an expectation-maximization algorithm based on the *Gaussian-mixture model*. The Gaussian-mixture model affirms that the energy deposits in the individual cells of a topological cluster are generated by Gaussian energy deposits from a number of particles equal to the number of seeds. With the iterative expectation-maximization algorithm, the positions and energies of the Gaussian functions are taken as cluster parameters.

Finally, calibrations derived from high statistic simulations are applied to the clusters to compensate the response to hadrons and energy losses due, for instance, to non-active detector material. The ECAL calibration is based on photons, while for HCAL neutral kaons are used. Both calibrations are probed with studies with particle data and their combination is used to model energy and η-dependent non-linearities in HCAL.

4.3.2 Link algorithm

Once the PF elements from different sub-detectors are reconstructed, the link algorithm pairs them to form PF blocks. The linking procedure is restricted to the nearest neighbours in the (η,ϕ) plane to limit the computing time.

Tracks are extrapolated from their last measured hit in the tracker to the calorimeters. A link is created if the extrapolated position is within the cluster area. If more links can be formed, only the one with the smallest distance between the extrapolated track and the cluster position is kept. Similarly, links can also be established between ECAL and HCAL clusters or between ECAL and preshower clusters. A common secondary vertex can be used to connect charged-particle tracks for nuclear-interaction reconstruction. Lastly, a track in the tracker can be connected with the reconstructed elements in the muon detector.

4.3.3 Particle reconstruction

The PF blocks are then used to identify particle candidates.

Muon candidates are identified and reconstructed first, and their corresponding PF elements are removed from the PF blocks list. Different criteria are enforced for standalone and non-isolated muons.

Next, electrons and isolated photons are processed. For the first also the energy of bremsstrahlung photons is collected. Different selections are applied for electron candidates found in the barrel or in the endcaps. Photons are reconstructed with the ECAL clusters only and photons' conversion in e^+e^- pairs in the tracker material is also considered. Then, the used tracks and clusters are masked for the subsequent processing step.

Last, hadrons from jet fragmentation and hadronization are identified. Together with neutral and charged hadrons, also non-isolated photons from the π^0 decay or, more rarely, muons from early charged hadrons decays can be reconstructed in this step. Neutral hadrons and photons are identified from calorimeter clusters not linked with any tracks. The classification of the clusters in one or the other particle type depends if they are located in the tracker acceptance region or not, to obtain the best jet energy reconstruction:

- For $|\eta| < 2.5$ (tracker acceptance region) ECAL clusters are turned into photons and HCAL clusters are turned into neutral hadrons, since in hadronic jets 25% of the energy is carried by photons and neutral hadrons deposit only 3% of the jet energy in the ECAL.

- For $|\eta| > 2.5$ charged and neutral hadrons cannot be distinguished. Since both hadron types leave $\approx 25\%$ of their energy in the ECAL, ECAL clusters linked to a given HCAL cluster are assumed to arise from the same hadron shower, while only ECAL clusters without such a link are classified as photons.

The energy of each identified particle is also estimated accordingly with the above classification. The remaining HCAL clusters of the PF block are linked to the remaining tracks, which can be linked to still unused ECAL clusters. Then, the calibrated calorimetric energy is determined. Finally, the particle content is determined by comparing the sum of the track momenta and the calibrated calorimetric energy:

- If the calibrated calorimetric energy is larger than the sum of the track momenta, the excess may be interpreted as the presence of photons and neutral hadrons.
- If the calibrated calorimetric energy is compatible with the sum of the track momenta, no neutral particle is identified, and the charged-hadron momenta are redefined combining the tracker and the calorimeters information. This aggregation allows the precise energy determination when the track parameters are measured with degraded resolutions and ensures a smooth transition between the low- and the high-energy regime, dominated by the tracker measurements and the calorimetric measurements, respectively.
- If the calibrated calorimetric energy is significantly smaller than the sum of the track momenta, a search for muons is performed, with looser identification criteria.

Jets are then reconstructed starting by the PF candidates in the event. The jet reconstruction and the identification of different jet types is especially relevant for the analyses presented in this thesis and is thus described in the next Chapter (Chapter 5). The following section introduces different pileup mitigation techniques that can be applied before the jet reconstruction is performed.

4.4 Pileup mitigation techniques

As already introduced, the multiple pp-interactions happening at each bunch crossing are referred to as pileup (PU). In particular, this is the *in-time* pileup, indicating the additional pp-collisions occurring within the same bunch-crossing as the primary hard interaction and thus producing extra tracks and energy deposits. The pp-collisions happening in the previous or subsequent bunch crossing can also contribute energy deposits to the same time window as the primary hard interaction because of the finite signal decay time of the ECAL and HCAL detectors (*out-of-time* pileup). These additional deposits can contaminate the reconstructed jets, degrading the jet observables evaluation and affecting the jet inner structure. The identification of highly boosted heavy resonances, performed in the analysis presented in this thesis, relies significantly on jets and jets' substructure observables. Advanced techniques to mitigate the effect of pileup are thus necessary.

The CMS Collaboration has been developing pileup removal algorithms well before the beginning of Run-2 [146] and is constantly improving the techniques to adapt them to the continuously increasing LHC luminosity.

This section describes the Charged-Hadron Subtraction (CHS) and the PileUp Per Particle Identification (PUPPI) algorithms, the two most commonly used pileup suppression approaches adopted in the CMS Collaboration, both employed in the performed analyses.

4.4.1 Charged-hadron subtraction

The charged-hadron subtraction (CHS) [136, 146] method, as the name suggests, excludes the charged hadrons associated with PU vertices from the jet clustering procedure. The removal is possible thanks to the tracking information associated to charged hadrons. However, neutral particles and all reconstructed particles outside the tracker acceptance cannot be associated with one of the reconstructed primary vertices with this technique. Nonetheless, thanks to the uniformity of the p_{T} density of pileup interactions in the (η, ϕ) plane, the average p_{T} contributions expected from pileup can be subtracted, applying event-by-event jet-area-based correction [147–149] to the jet four-momenta to remove the remaining energy due to neutral particles originating from PU vertices.

4.4.2 Pileup per particle identification

The pileup per particle identification (PUPPI) [81, 150, 151] algorithm is used to remove pileup, identified through global information from the event and local information at the particle level. It associates a weight to each particle. Such weight is used to rescale the four-momentum of the particle, correcting at the same time the jet kinematic variables and the jet shape in an observable-independent way.

For charged particles, the PUPPI weight is assigned based on tracking information. The weights range from 0 (particles most likely coming from PU vertices) to 1 (particles coming from the hard scatter). A weight of 1 is assigned also to charged particles not associated with any vertex if their distance of closest approach to the LV along the z axis is smaller than 0.3 cm; a weight of 0 is applied in all other scenarios. A discriminating variable (α) is used to assign weights to the neutral particles. Within the CMS Collaboration, α is defined as follows [81]:

$$\alpha_i = \log \sum_{j \neq i,\, \Delta R_{ij} < R_0} \left(\frac{p_{\mathrm{T}j}}{\Delta R_{ij}} \right)^2 \begin{cases} \text{for } |\eta_i| < 2.5, & j \text{ are charged particles from LV,} \\ \text{for } |\eta_i| > 2.5, & j \text{ are all kinds of reconstructed particles,} \end{cases}$$

where i refers to the particle in question, j are other particles, $p_{\mathrm{T}j}$ is the transverse momentum of particle j in GeV, and $\Delta R_{ij} = \sqrt{(\Delta\eta_{ij})^2 + (\Delta\phi_{ij})^2}$ is the distance between the particles i and j in the η–ϕ plane. The summation runs over the particles j in the cone of particle i with a radius of $R_0 = 0.4$. The α variable is designed to be large for particles close to other particles originating from the LV or highly energetic particles (outside the tracker acceptance region).

The expected PU distribution in an event is estimated from charged particles assigned to PU vertices. In this way, the α_i of each neutral particle can be compared, with a signed χ^2 approximation, with the median and RMS of the α distribution of the charged PU particles to compute each neutral particle's probability to arise from the LV or a PU vertex. Finally, the p_{T} weight w_i of the i-th neutral particle is given by the cumulative distribution function of the χ^2 distribution evaluated at the χ^2_i value.

4.4.3 Performance of the CHS and PUPPI algorithms

PUPPI achieves a stronger PU reduction, especially for events with more than 30 interactions, approximately the mean number of inelastic interactions per bunch crossing in the 2016–2018 data taking periods (Fig. 3.1 in Chapter 3).

In general, the PUPPI algorithm provides the best performance for jet mass and substructure observables, removing the PU dependence. Concerning jet-momentum resolution and PU jet rejection, the preferred algorithm depends on the physics process under study: the PUPPI algorithm provides a better jet momentum resolution for jets with $p_{\mathrm{T}} < 100$ GeV, whereas CHS does so for $p_{\mathrm{T}} > 100$ GeV [81].

The analyses presented in this thesis take advantage of the different performance of the presented PU suppression algorithms. Both are employed, depending on the jet observable being studied or the physics processes from which the jets originate.

5 Jet reconstruction and jet substructure

Jets are collimated streams of particles produced from the hadronization of quarks and gluons. Within the CMS Collaboration, they are reconstructed from the PF candidates in the event using specific clustering algorithms described in Sec. 5.1. The energy calibration procedure that jets undergo before being used in the analyses is described in Sec. 5.2. The inner structure of jets provides valuable information on the origin of the jets. Different substructure observables used to identify (*tag*) the decay products of heavy objects are presented in Sec. 5.3.

In Sec. 5.3.4.1 jet substructure studies specific for the analysis described in Chapters 6–8 are presented. The decorrelation method applied to the jet identification algorithms has been developed by Dr. Anna Benecke and the data to simulation corrections of the algorithm identification efficiency have been estimated by the algorithm developer Dr. Loukas Gouskos. The author of this thesis, under the supervision of Dr. Andreas Hinzmann and in collaboration with Dr. Clemens Lange and Dr. Jennifer Ngadiuba, has validated the decorrelation method in the context of the analysis described in the next chapters.

5.1 Jet clustering

For the jets reconstruction, the studies performed in this thesis use sequential clustering algorithms, designed to be infrared and collinear (IRC) safe, i.e. insensitive to the emission of soft or collinear gluons. In the analyses presented in this thesis, the anti-k_{T} algorithm [144, 145] has been adopted. The algorithm starts from a list of entities (particles or combinations of particles). It iteratively introduces two (p_{T}-dependent) distances to classify a list element as a jet or recombine it. The distances d_{ij} between all possible elements pairs i and j and d_{iB} between entity i and the beam B are defined as

$$d_{ij} = \min(p_{Ti}^{2n}, p_{Tj}^{2n}) \frac{\Delta R_{ij}^2}{R^2}$$
$$d_{iB} = p_{Ti}^{2n},$$

where ΔR_{ij}^2 is the distance between i and j in the η-ϕ plane, and R is the jet *radius* parameter. The value of the integer n distinguishes the different sequential clustering algorithms. The anti-k_{T} is obtained for n= -1. Instead, the Cambridge-Aachen (CA) algorithm [152] has n = 0.

At each iteration, the smallest d_{ij} and d_{iB} are compared. If $d_{ij} < d_{iB}$, the i and j entities are combined in a new entity, that is added to the input list while i and j are removed. Instead, if $d_{iB} < d_{ij}$, i is classified as jet and removed from the input

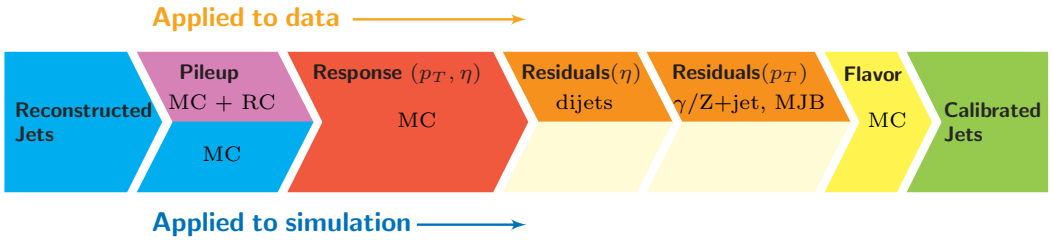


FIGURE 5.1: Consecutive stages of JEC, for data and MC simulation. All corrections marked with MC are derived from simulation studies, RC stands for random cone, and MJB refers to the analysis of multijet events. Taken from Ref. [149].

list. The procedure is repeated until the list is empty. Given how the distances are defined, hard particles are clustered first, and the jet area in the η–ϕ plane does not depend on the soft particles.

The FASTJET package [145] is used to perform the anti-k_T clustering described above.

In this thesis, narrow and large jets are reconstructed with the standard CMS radii $R = 0.4$ (referred to as *AK4*) and $R = 0.8$ (referred to as *AK8*), respectively.

5.2 Jet energy calibration

Jets are experimentally reconstructed objects that must be calibrated to correct their energy scale. Jet energy corrections (JEC) are derived following the procedure described in Ref. [149] and are used for the calibration of the jet energy scale (JES). Figure 5.1 shows the factorized approach adopted by the CMS Collaboration, where subsequent corrections accounting for different effects are applied in a fixed order. The jets four-momenta are rescaled at each stage.

1. The *pileup offset corrections* are determined from simulation of di-jet events with and without pileup on top. They are parametrized as a function of the jet area, η, p_T and average p_T density per unit area. Furthermore, corrections for residual differences between data and simulation are derived with the random cone (RC) method [153] from data collected using a random trigger in the presence of a beam crossing (*zero bias*).

2. The *simulated response corrections* depend on the η and p_T of the jet to take into account the non-uniform detector response. Their estimation is based on the measurement of the discrepancies between the reconstructed p_T of the jet and the true p_T before simulating the interaction of the hadrons from the parton shower with the detector (*generator level* p_T).

3. *Residual* corrections account for minor differences in the jet response in data and simulation. They are estimated from Z/γ+jet events and multi-jet events as a function of η and p_T and applied to the simulation.

4. An optional *flavour*-dependent correction, not used in this analysis, is extracted from MC simulations.

The discrepancies between data and simulation regarding the jet p_T resolution are also considered applying additional smearing in simulation, referred as jet energy resolution (JER) corrections.

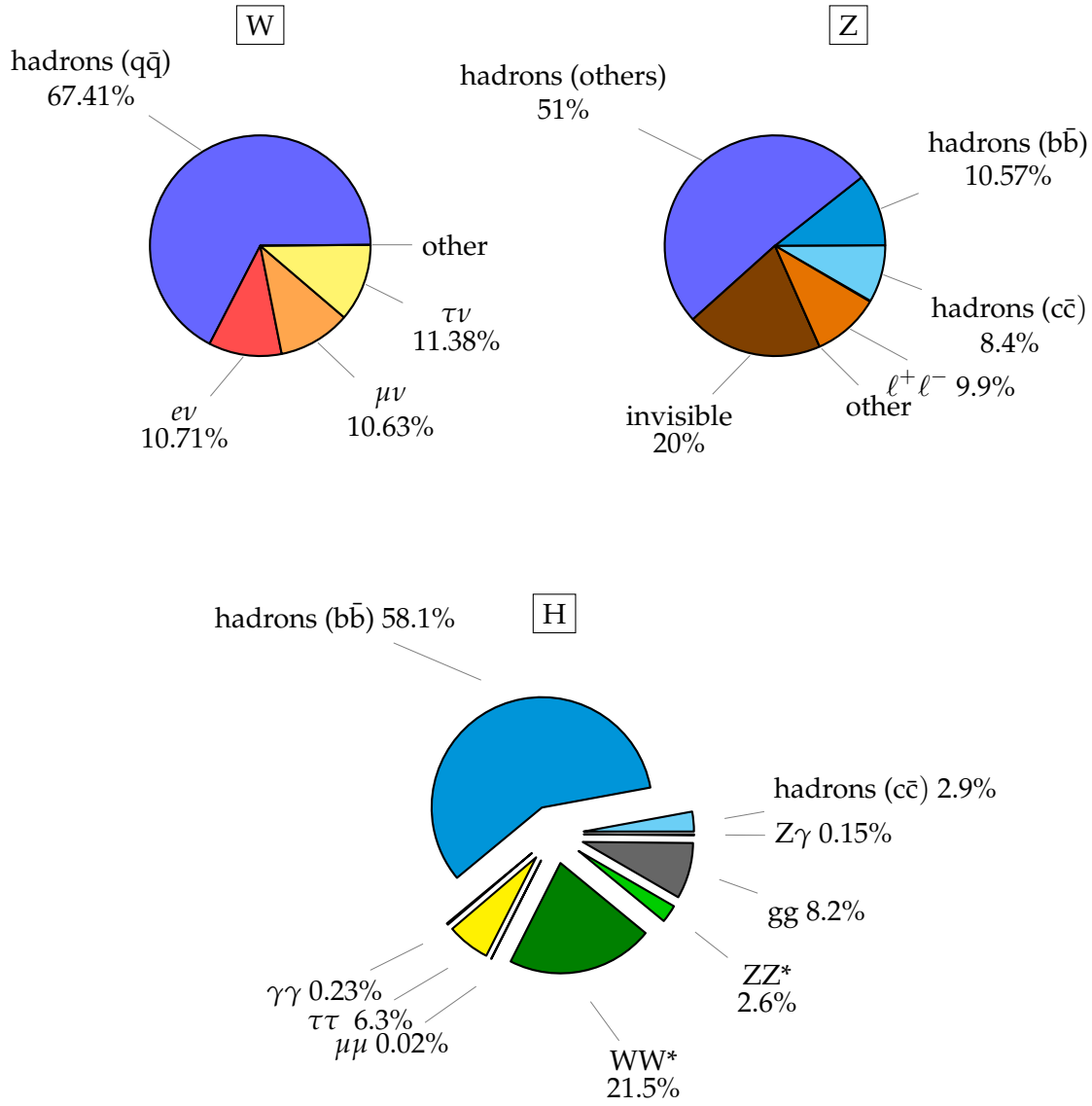


FIGURE 5.2: Top left (right): W (Z) boson decay branching fractions, obtained from Ref. [15]. Bottom: Higgs decay branching fractions as predicted in the SM, adapted from Ref. [155].

JES and JER performances with the data collected by the CMS experiment in 2016–2018 can be found in Ref. [154].

5.3 Boosted objects identification with jet substructure techniques

W and Z bosons (also referred as V bosons) decay mostly into pairs of fermions, as shown in Fig. 5.2 (top), favouring more than two-thirds of the times the hadronic decay into a quark-antiquark pair (67.4% and 69.9%, respectively [15]). The Higgs boson decay branching fractions are presented in Fig. 5.2 (bottom): the preferred channel is $H \to b\bar{b}$. Noticeably, the Z boson also decays to $b\bar{b}$ in 15% of the hadronic decays.

The angular distance $\Delta R_{q\bar{q}} = \sqrt{\Delta^2\phi_{q\bar{q}} + \Delta^2\eta_{q\bar{q}}}$ between the two quarks resulting from the decay of V or H bosons can be estimated as

$$\Delta R_{q\bar{q}} \approx \frac{2m}{p_T}, \tag{5.1}$$

where $m \approx \mathcal{O}(100\,\text{GeV})$ and p_T are the boson mass and transverse momentum, respectively. For low p_T bosons, the two jets obtained from each of the final state quarks are well separated. Instead, if a boson carries a high momentum (*boosted* regime), the two jets are close to each other, merged into one single jet. For instance, already for a W boson with p_T of 400 GeV, the angular distance between quarks is ≈ 0.4 and the corresponding jets cannot be resolved with a single AK4 jet. This latter case happens, for instance, for bosons originating from heavy resonances, like the ones studied in this thesis.

Therefore, jet substructure observables and dedicated algorithms are necessary to identify the two subjets of boosted W-, Z- or H-jets and distinguish such jets from quark or gluon initiated jets of QCD multijet background.

Large (AK8) jets are adopted in this thesis to reconstruct high p_T jets. With such a large area, contributions from underlying events or pileup could degrade or bias the measurement of the substructure observables. The latter can be reduced with the mitigation techniques described in Sec. 4.4. Grooming techniques are instead used to reduce soft and wide-angle radiation and improve the jet mass resolution.

The techniques and variables employed in this thesis are discussed below. Variables, algorithms and sets of requirements on observables used to identify an object are commonly called *taggers*.

5.3.1 Soft drop algorithm

The *soft drop* (SD) algorithm [156] is primarily used to reduce the jet contamination from initial state radiation, underlying event and pileup. Before applying the algorithm, the anti-k_T jets are re-clustered with the CA approach, which has an angular-ordered structure. The last step of the CA clustering is then reverted by splitting the jet j in the j_1 and j_2 subjets.

The soft drop condition is defined as

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R}\right)^{\beta}. \tag{5.2}$$

It depends on the soft drop threshold z_{cut} and the exponent β. If the subjets satisfy the condition, j is the final soft drop jet. Otherwise, the procedure is repeated on the subjet with the highest p_T, while the other is removed. In this thesis, the parameters are $z_{\text{cut}} = 0.1$ and $\beta = 0$. The latter condition ensures soft and collinear safety when evaluating the mass of the final SD jet [157]. The soft drop mass m_{SD} is the invariant mass of all constituents of the final SD jet.

The advantages of using the SD corrected jet mass instead of the raw jet mass are shown in Fig. 5.3. Before (solid line) SD is applied, the W peak (blue) is large and shifted to higher values than the expected boson mass. After SD is performed (dotted line), the peak is narrower and centred at the W mass. Furthermore, the separation between the W-jets and the q/g-jets is increased.

In this thesis the expressions jet mass, M_{SD}, m_{SD}, m_{jet} and $\text{m}_{\text{SD}}^{\text{jet}}$ are used interchangeably to refer to the jet soft drop mass.

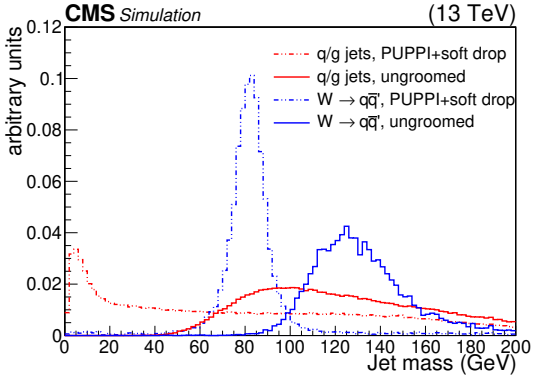


FIGURE 5.3: The jet mass spectra for quark/gluon-jets (red) and W-jets (blue) before (solid line) and after (dotted line) softdrop has been applied. Taken from Ref. [158].

5.3.2 *N*-subjettiness

As described above, jets originated from boosted V or H bosons consist of two merged prongs, while q/g-jets are not expected to have such an inner structure. The N-subjettiness τ_N variable [159] has been introduced to quantify how likely it is for a reconstructed jet to be composed of N subjets. After N subjets are identified, τ_N is calculated as

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \cdots, \Delta R_{N,k}\}, \tag{5.3}$$

where k runs over the constituents of the jet and the distances $\Delta R_{i,k}$ are calculated relative to the axis of the i-th subjet. The normalization factor d_0, for a characteristic jet radius R, is defined as

$$d_0 = \sum_k p_{T,k} R. \tag{5.4}$$

If all the constituents are aligned with the identified N subjets, $\tau_N \approx 0$ and the initial jet is likely to have N (or fewer) subjets. If $\tau_N >> 0$, a large energy fraction is not aligned with the supposed N subjets and, thus, the jet has at least $N+1$ subjets. For the algorithm to properly work, the subjets must be properly chosen. While the ideal choice would be to determine τ_N by minimizing all possible candidate subjet directions, the subjet axis is instead chosen to force the exclusive-k_T algorithm [160, 161] to cluster N subjets to limit the computing time.

For jets with two prongs (for instance, originated from a W boson), τ_2 is close to 0, while τ_1 approximately 1. On the contrary, for a q/g initiated jet $\tau_1 \approx 0$. Thus, the ratio $\tau_{21} = \tau_2/\tau_1$ is commonly used to discriminate the discussed jet types. This variable is used in the study presented in Chapter 11 and Fig. 11.3a shows an application of τ_{21} to discriminate a signal from a heavy resonance against the QCD multijet background.

5.3.3 Jets originating from b quarks

Jets originating from b quarks (b-jets) have specific characteristics that distinguish them from jets initiated from lighter quarks or gluons. The b-jets identification is relevant for many applications, from searches for new physics to many SM studies. For instance, the top quark decays to a b quark and W boson with ≈100% BR. Furthermore, it allowed the observation of the $H \to b\bar{b}$ decay [162, 163].

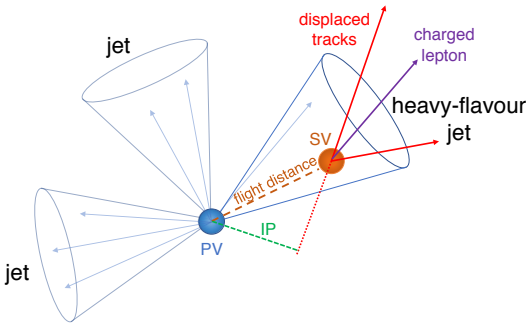


FIGURE 5.4: Illustration of a heavy-flavour jet with a secondary vertex (SV) from the decay of a b or c hadron resulting in charged-particle tracks (including possibly a soft lepton) that are displaced with respect to the primary interaction vertex (PV), and hence with a large impact parameter (IP) value. Taken from Ref. [164].

B quark initiated jets contain B hadrons, which are characterized by a relatively long lifetime of 1.5 ps. If the B hadron has a momentum sufficiently higher than its mass (≈ 5 GeV), it can travel a distance of the order of few millimetres, from where it is created in the primary vertex (PV) to its decay. This generates a secondary vertex (SV) displaced from the PV. Figure 5.4 schematically illustrates the decay of a heavy flavour jet, such as a b-jet. The (large) impact parameter (IP), the closest distance from the primary vertex to a given track, is also shown. High granularity trackers, as the one in use in the CMS experiment, are able to reconstruct the secondary vertex. The information can then be used to develop dedicated techniques to identify b quark initiated jets [164].

5.3.4 Jet tagging with machine-learning techniques

Efficient classification of heavy objects is essential for many SM and BSM analyses. New taggers based on machine-learning (ML) techniques developed by the CMS Collaboration [165] can considerably reduce the background rate without degrading the signal efficiency compared to non-ML techniques. In the analysis presented in Chapters 6–8, the *DeepAK8* tagger [165] has been adopted to improve the signal-to-background discrimination.

The DeepAK8 tagger is a multi-classifier that can differentiate hadronically decaying particles into five categories ($W/Z/H/\mathrm{t}/\mathrm{other}$) subdivided into the particles' different decay channels (e.g. $Z \to \mathrm{q\bar{q}}$, where q is a light quark, and $Z \to \mathrm{b\bar{b}}$). The inputs to the algorithm are low-level variables (e.g. particle-flow candidates) to benefit from the information made available by the high granularity of the detectors. They are organized in two lists per jet:

- The *particle list* contains 100 jet constituent particles, ordered by decreasing p_T. For a better determination of the jet substructure, the measured properties of the particles (p_T, energy deposit, charge, the angular separation between the particle and the jet axis or the subjet axes, etc.) are also provided to the algorithm. For charged particles also track related quantities can be included. Overall, 42 variables are considered for each particle.

- The *secondary vertex list* can feature up to 7 SVs, each described by 15 variables (SV kinematics observables, displacement, quality criteria, etc.). The two-dimensional impact parameter significance is used to sort the list elements. This list assists the heavy-flavour jet tagging.

The DeepAK8 features a customized Deep Neural Network (DNN) architecture, developed to handle the high number of inputs and their correlations. The DNN has two steps. In the first step, each of the lists is given to a corresponding Convolutional Neural Network (CNN) based on the ResNet model [166]. The CNNs have 14 and 10 layers for the particle list and the SV list, respectively. Each layer can provide 32–128 output channels. The two CNNs work in parallel to transform the inputs and extract the helpful characteristics. The jets classification happens in the second step, where a simple, fully connected network combines and processes the output of the two CNNs. The second step consists of a single layer with 512 units. It is followed by a ReLU activation function and a Dropout [167] layer of 20% drop rate. The training is performed with a 50 million jets sample. The 80% of the sample is used for the training itself, while the remaining 20% is reserved for the validation. The signal and background samples are reweighted to obtain a uniform distribution in p_{T} to avoid biases in the training process.

The features extracted by the CNNs can be highly correlated with the jet mass. As a consequence, the jet mass distribution of background jets selected with the DeepAK8 tagger can be modified and look similar to the signal. Decorrelating the tagger from the jet mass results in many advantages [168] and is of crucial importance for the analysis described in Chapters 6–8. The DeepAK8-MD (Mass Decorrelated) has been developed to provide a mass decorrelated tagger that preserves (as possible) the discriminating power of the DeepAK8. It is based on an adversarial training approach [169], and the signal and background samples are reweighted to obtain flat p_{T} and m_{SD} distributions. The DeepAK8-MD is equipped with an additional mass prediction network to derive the jet mass from the features extracted by the CNNs. This extra network has three fully connected layers of 256 units each and a SELU activation function [170]. It is trained to predict the m_{SD} of the background jets to the closest 10 GeV value in the range 30–250 GeV by minimizing the cross-entropy loss. An accurate mass prediction shows that the features obtained by the CNNs are strongly correlated with the jet mass. The accuracy of the mass prediction is then given as a penalty to the tagger algorithm. The mass prediction network is trained for ten steps for each of the CNNs training stages to handle the updating output of the CNNs.

The final output of the DNN is a score for each (sub)-category, used to select specific decay channels.

In the analysis presented in Chapters 6–8 the W, Z and H decays of interest are selected with the DeepAK8 scores listed below:

$$\mathrm{ZHbbvsQCD} = \frac{\mathrm{score}_{Z\to \mathrm{b\bar{b}}} + \mathrm{score}_{H\to \mathrm{b\bar{b}}}}{\mathrm{score}_{Z\to \mathrm{b\bar{b}}} + \mathrm{score}_{H\to \mathrm{b\bar{b}}} + \sum \mathrm{score}_{\mathrm{QCD}}} \tag{5.5}$$

and

$$\mathrm{WvsQCD} = \frac{\mathrm{score}_{W\to \mathrm{q\bar{q}}} + \mathrm{score}_{W\to \mathrm{q\bar{c}}}}{\mathrm{score}_{W\to \mathrm{q\bar{q}}} + \mathrm{score}_{W\to \mathrm{q\bar{c}}} + \sum \mathrm{score}_{\mathrm{QCD}}}, \tag{5.6}$$

where

$$\sum \mathrm{score}_{\mathrm{QCD}} = \mathrm{score}_{\mathrm{QCD}\to \mathrm{b\bar{b}}} + \mathrm{score}_{\mathrm{QCD}\to \mathrm{c\bar{c}}} + \mathrm{score}_{\mathrm{QCD}\to \mathrm{b}} + \mathrm{score}_{\mathrm{QCD}\to \mathrm{c}} + \mathrm{score}_{\mathrm{QCD}\to \mathrm{other}}.$$

The WvsQCD final score is used to select candidate W or Z bosons decaying to two quarks while the ZHbbvsQCD combines the scores of $Z \to \mathrm{b\bar{b}}$ and $H \to \mathrm{b\bar{b}}$ to increase the Z tagging efficiency.

In Fig. 5.5 the performance of the DeepAK8 taggers are shown for classes relevant for this thesis. The Receiver Operating Characteristic (ROC) curves presenting the

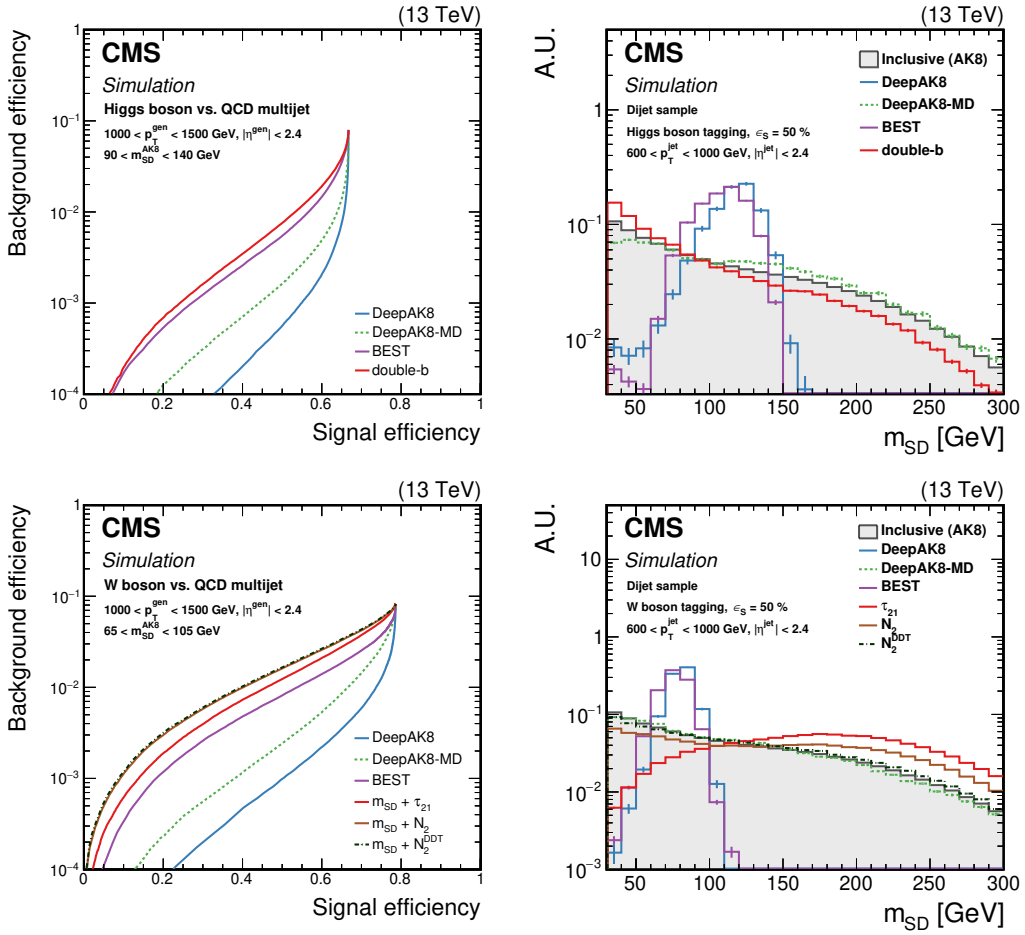


FIGURE 5.5: Top (bottom) left: Comparison of the identification algorithms for hadronically decaying H (W) boson in terms of ROC curves. Top (bottom) right: Normalized m_{SD} distribution for QCD multijets background, inclusively and after selection by each H (W) tagging algorithm. The working point chosen corresponds to a signal efficiency (ϵ_S) of 50%. The error bars represent the statistical uncertainty in each specific bin, which is related to the limited number of simulated events. Taken from Ref. [81].

misidentification rate as a function of the signal efficiency show how the DeepAK8-MD, despite a loss in discriminating power with respect to the nominal version, still has a superior behaviour compared to the other non-ML approaches shown, e.g. τ_{21}. Furthermore, the *sculpting* of the m_{SD} distribution is considerably reduced with respect to the DeepAK8.

5.3.4.1 2D decorrelation of the DeepAK8 taggers

The m_{SD} spectrum of the QCD-like background in Fig. 5.3 is smoothly falling (above $\approx$ 20 GeV), while for the *resonant* signal the m_{SD} distribution presents a peak. This feature has been exploited to perform a *bump hunt* (also) in the jet mass distribution in the analysis presented in Ref. [9], that pioneered the approach followed in this thesis. In Ref. [9], a decorrelated version of the τ_{21} variable (τ_{21}^{DDT}) was adopted to eliminate sculpting features that could be introduced in the m_{SD} spectra from the tagging requirements.

As it was illustrated in the previous section (Sec. 5.3.4), the DeepAK8 tagger

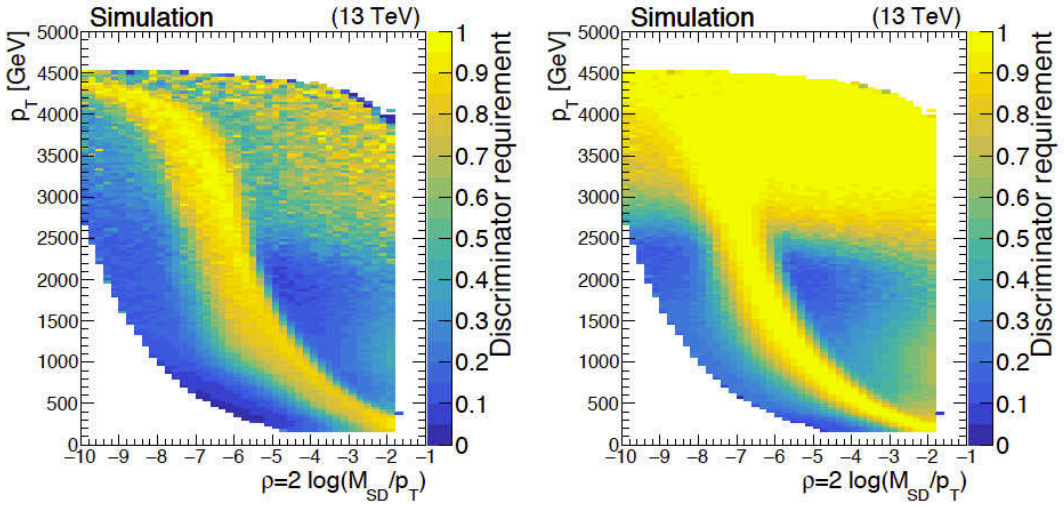


FIGURE 5.6: 2D map derived for (left) the DeepAK8 ZHbbvsQCD tagger and (right) the DeepAK8 WvsQCD tagger for a 5% misidentification rate for QCD multijet simulation. Taken from Ref. [151].

has superior performance in terms of signal identification compared to non-ML approaches, including τ_{21}^{DDT}. However, it introduces a significant shaping in the m_{SD} distribution. The DeepAK8-MD also introduces (small) features if compared to the inclusive distribution.

A 2D-decorrelation method for the DeepAK8 developed in Ref. [151] makes such taggers employable for the analysis presented in this thesis. In the 2D-decorrelation, first introduced in Ref. [171] for an energy correlation function (N_2^{DDT}), a three-dimensional histogram of the variables $\rho = \log(m_{SD}^2/p_T^2)$, p_T and tagger discriminator is filled for the leading-p_T jet. Afterwards, a *map* is generated showing the discriminant requirement for a certain misidentification rate X% for each (ρ, p_T) bin in the three-dimensional histogram. Thus, the maps describe the requirement on the discriminator for a X% misidentification rate.

The method has been applied on the DeepAK8 scores introduced in Eq. 5.5 and 5.6 for the ZHbbvsQCD and the WvsQCD tagger, respectively. Maps for a 5% misidentification rate for QCD multijet simulation for such scores are shown in Fig. 5.6.

The decorrelation performance (*mass sculpting*) is evaluated in the difference of the soft drop mass distribution of the leading-p_T jet with and without a requirement of the tagger. Fig. 5.7 (upper part) shows the mass sculpting of QCD multijet simulation for the ZHbbvsQCD tagger for the DeepAK8-MD (left) and the DeepAK8 with 2D (map) decorrelation (right). The former method introduces a peak and a dip in the m_{SD} spectrum for certain requirements on the discriminator, while the latter does not introduce such features, resulting in mass sculpting of less than 10% over the whole mass range [151].

The maps in Fig. 5.6 present steep variations of the discriminant value for small variations of jet ρ and p_T. Therefore, discrepancies in the jet p_T and/or jet mass between data and the MC simulation used to derive the maps may induce a significant change in the discriminant value and hence considerably shape the distributions. The assessment of the potential impact of this feature has been performed in a data control region (CR) by comparing the differences in the jet mass distribution in two different cases. The CR consists of events where one jet passes a tight requirement on

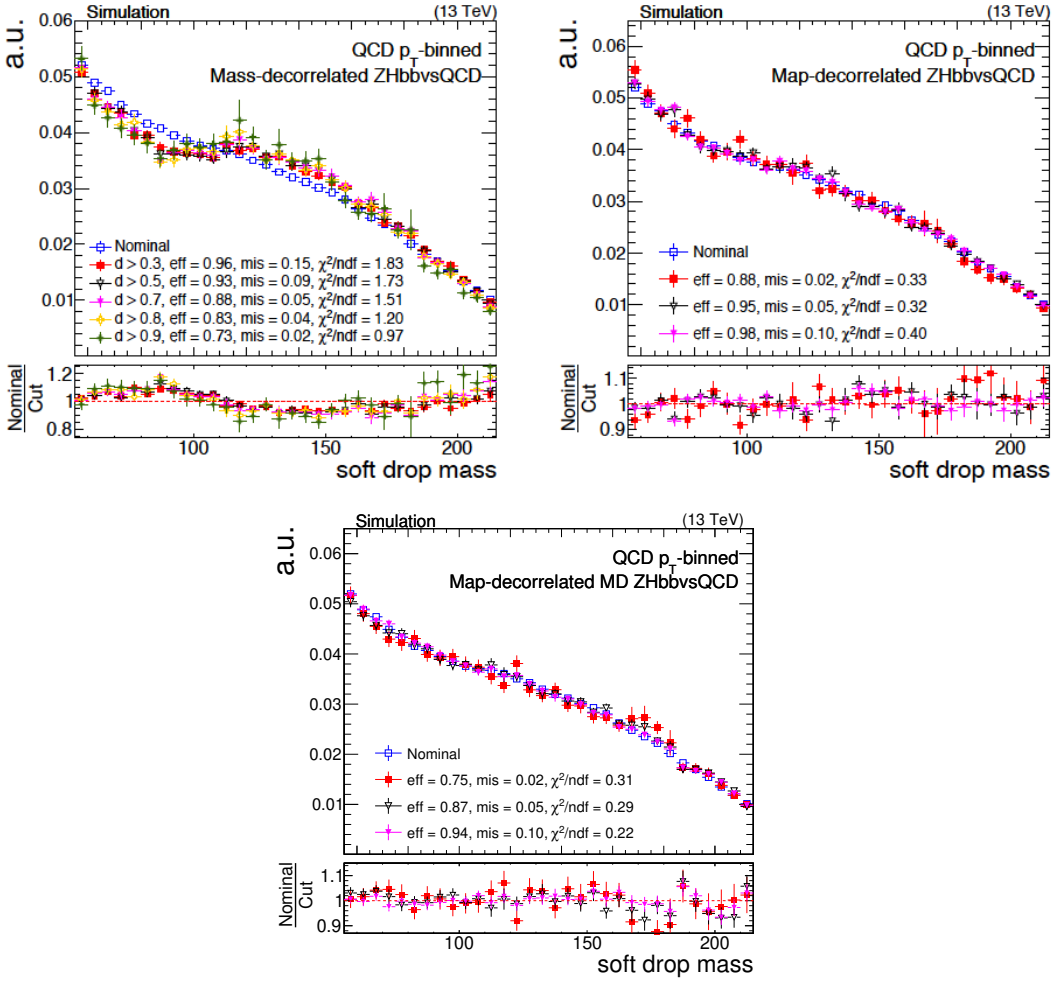


FIGURE 5.7: Soft drop mass of the leading-p_T AK8 jet in QCD multijet simulation for DeepAK8-MD ZHbbvsQCD (upper left), DeepAK8 ZHbbvsQCD 2D-decorrelated (upper right) and DeepAK8-MD ZHbbvsQCD 2D-decorrelated (bottom). Different requirements on the discriminator corresponding to different misidentification rates (mis) and signal efficiencies (eff) are shown and compared to the nominal (without any requirement). The χ^2/ndf is calculated for each spectrum with respect to the nominal spectrum. The lower panel shows the ratio between the nominal spectrum and one of the spectra with a requirement on the discriminator. Taken from Ref. [151] (upper part) and Ref. [172] (lower part).

the WvsQCD tagger and the other fails a loose one[1]. For the comparison, jets passing the same misidentification rate requirements are considered, in one case reading the 2D map with the nominal jet p_T and mass value and in the other case shifting the value of the jet p_T or mass of 1–2% from the nominal.

The shift in the jet mass value resulted in significant differences in the m_{jet} distribution in the data CR, as shown in Fig. 5.8 (left). Such differences are, instead, considerably reduced if the maps are derived from the DeepAK8-MD tagger (Fig. 5.8, right). Thus, for the analysis presented in this thesis, the 2D-decorrelated DeepAK8-MD tagger is used. Maps obtained from the DeepAK8-MD tagger for a 5% misidentification rate for QCD multijet simulation for the scores introduced above are shown in Fig. 5.9. The variation of the tagging requirements in neighbouring bins are now

[1]The control region is defined in the CMS internal Ref. [173]

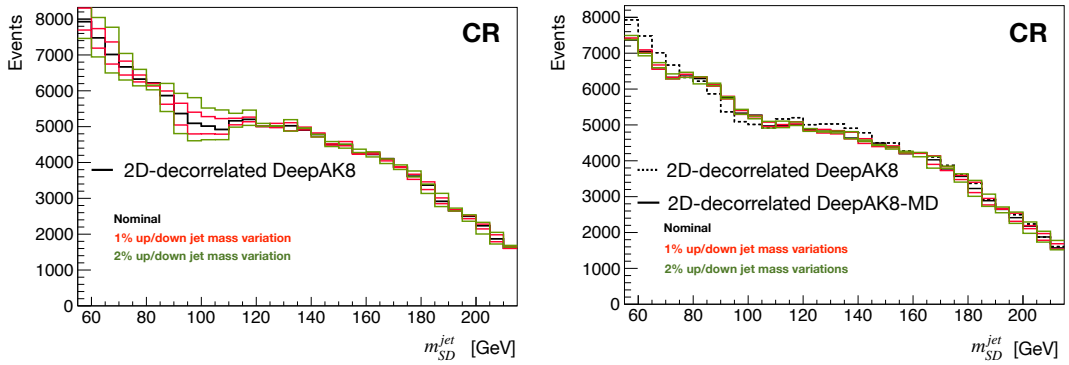


FIGURE 5.8: Jet mass distributions in a data control region consisting of events where one jet passes a tight requirement on the WvsQCD tagger and the other fails a loose one after the 2D-decorrelated WvsQCD discriminant derived from QCD MC is applied (black) and after the cut value is chosen from the map for 1 (red) and 2% (green) shifted values of the jet mass. The left and right plots are for the DeepAK8 and DeepAK8-MD, respectively.

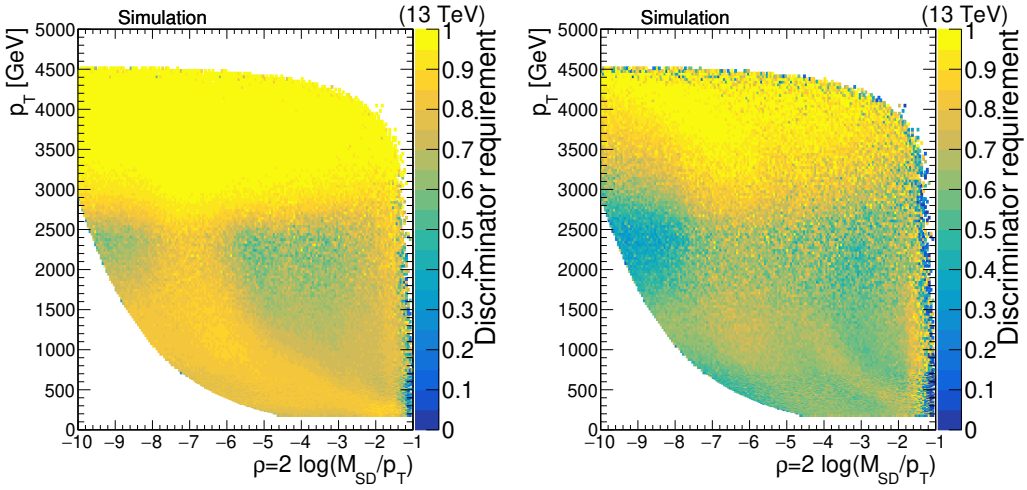


FIGURE 5.9: 2D map derived for (left) the DeepAK8-MD ZHbbvsQCD tagger and (right) the DeepAK8-MD WvsQCD tagger for a 5% misidentification rate for QCD multijet simulation.

smoother than in Fig. 5.6. For completeness, Fig. 5.7 (lower part) shows the mass sculpting of QCD multijet simulation for the ZHbbvsQCD tagger for the DeepAK8-MD 2D-decorrelated.

The performance in terms of misidentification rate and signal efficiency is shown in Fig. 5.10. The 2D decorrelation shows a loss of signal efficiency with respect to the algorithm used to derive the maps. However, the loss is smaller than between the DeepAK8 and its MD version. Furthermore, the adoption of the 2D-decorrelated DeepAK8-MD is still advantageous compared to non-ML approaches.

In the context of the search described in the following chapters targeting resonances decaying to two boson where each of the bosons produce a large jet, jet tagging requirements must be applied to both final state jets. Thus, it was verified that the mass sculpting is still negligible even when applying a cut on the discriminator of both leading-p_T jets. In fact, the results presented in Fig. 5.7 have been derived only for the first leading-p_T jet in the event. In particular, Fig. 5.11 shows a comparison of the dijet invariant mass and softdrop jet mass distribution with and

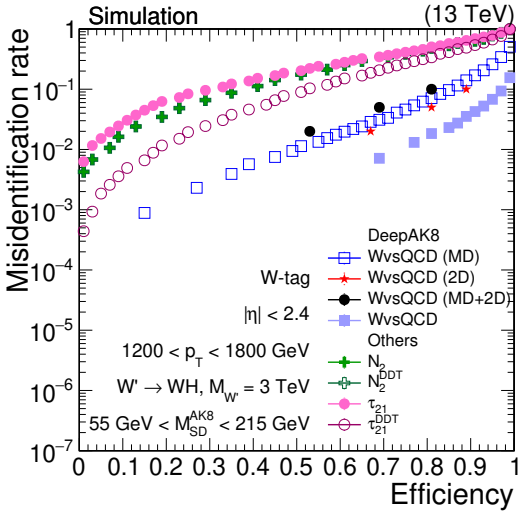


FIGURE 5.10: ROC curves evaluated with 2017 QCD background and signal simulations for different taggers available to identify a W boson with p_T in the range 1200–1800 GeV. Also shown are the performance of the 2D DDT maps at fixed mistag rates of 2, 5, and 10% when using the nominal (red stars) and MD (black dots) version of the DeepAK8 tagger to derive the map.

without applying cuts on the discriminants for the QCD multijet PYTHIA8 simulation sample. Even after applying tight requirements on the discriminant of both jets the spectrum of these variables agrees with the inclusive spectrum within statistical uncertainties. As it will be described in Chapter 7, this allows to derive QCD templates once without any tagger discriminant applied, i.e. once for all event categories (see Sec. 6.3), with the advantage of better stability as the full MC statistics is used, rather than the limited statistics after application of tight tagger requirements. Furthermore, it has been verified that the maps generated with the PYTHIA8 simulation maintain the same performance when applied to QCD multijet simulation generated with MADGRAPH5_aMC@NLO interfaced with PYTHIA or HERWIG++.

5.3.4.2 Measurement of the data to simulation corrections

Possible differences in tagging efficiencies between data and simulation are taken into account by the jet tagging efficiency scale factors (SFs), applied as a correction to the the simulated samples. The SFs are determined as the ratio of the efficiency in data over the efficiency in simulation. The 2D-decorrelated DeepAK8-MD SFs have been estimated together with the systematic uncertainties on the measurement (e.g. the used parton shower algorithm or the size of the simulated sample).

For different misidentification rates of the WvsQCD tagger defined in Eq. 5.6, the $W \to q\bar{q}$ SFs are estimated using the same procedure described in Ref. [165]. The efficiencies are measured with the tag-and-probe method illustrated in Ref. [174] in a data sample enriched in semi-leptonic $t\bar{t}$, where one of the t-quarks decays leptonically in the muon channel (*tag*), and the other provides a hadronic decay (*probe*). The scale factor for misidentification of top quark jets are also measured. A maximum likelihood fit to the m_{SD} is performed simultaneously on the events that *pass* and *fail* the requirement on the WvsQCD. In Fig. 5.12 an example of the mass distributions used to derive the SFs is shown.

For the ZHbbvsQCD, the SFs are derived in data in a sample enriched in $g \to b\bar{b}$ as described in Ref. [176]. With the current available luminosity, it is not possible to

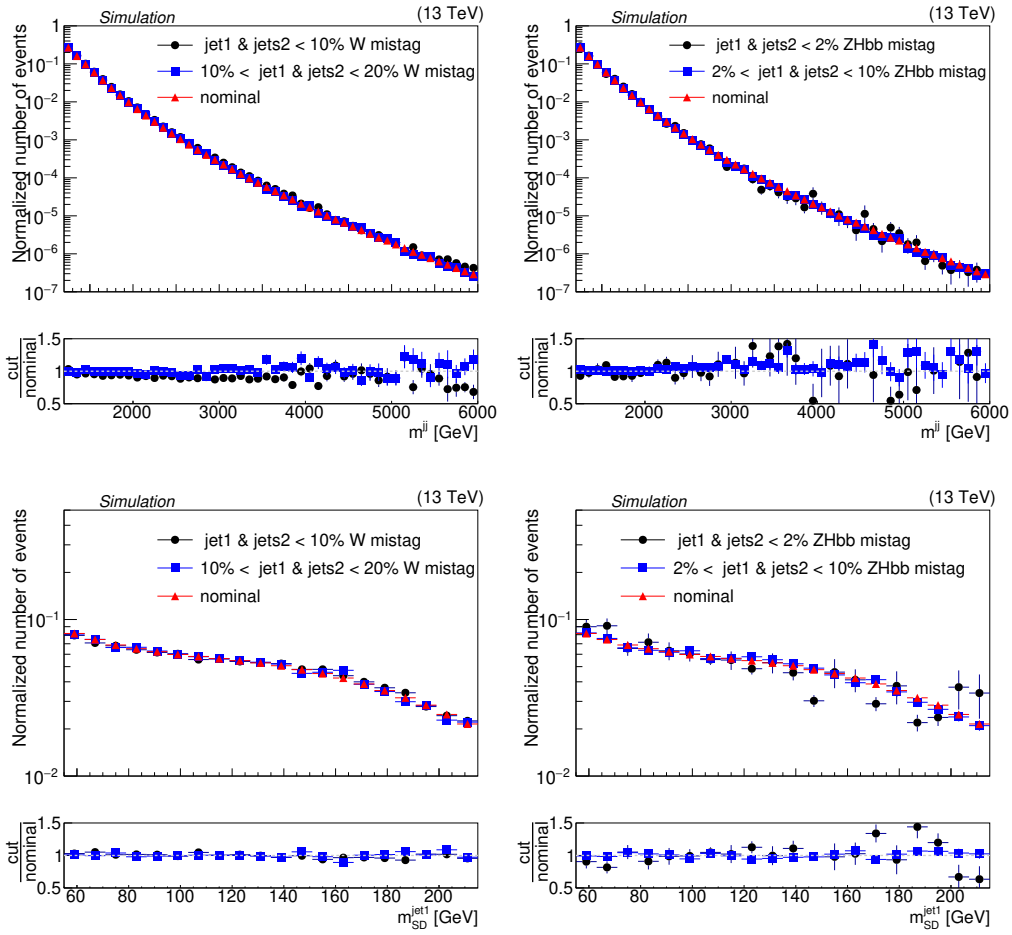


FIGURE 5.11: Dijet invariant mass (top) and softdrop jet mass of the first leading jet (bottom) distributions with and without applying cuts on the map-decorrelated DeepAK8-MD WvsQCD (left) and ZHbbvsQCD (right) discriminants.

obtain a pure Z or H sample. Thus the calibration relies on the use of proxy jets that have characteristics similar to the signal, e.g. by requiring a secondary vertex. The approach ultimately relies on a template fit method of the CSV discriminator [164] distribution to extract the fractions of different flavour jets ($b\bar{b}$, $c\bar{c}$ or udsg). The fit is performed simultaneously on the *pass* and *fail* regions. In Fig. 5.13 an example of the distributions used to derive the SFs is shown.

The applied SFs are summarized in Chapter 6, where the chosen taggers' working points are described.

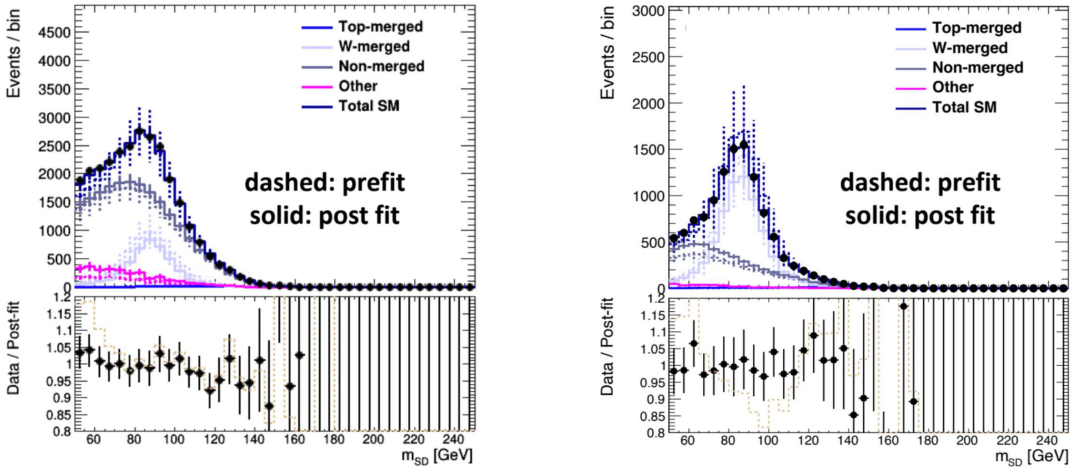


FIGURE 5.12: *Fail* (left) and *pass* (right) m_{SD} distributions used to derive the WvsQCD SF in the 2018 data taking year for a 10% misidentification rate for jets with 170 GeV $< p_T <$ 300 GeV. The lower panels show the ratio between the data (black markers) or the expectation from simulation before the fit (dashed line) and the post-fit distributions. Taken from Ref. [175].

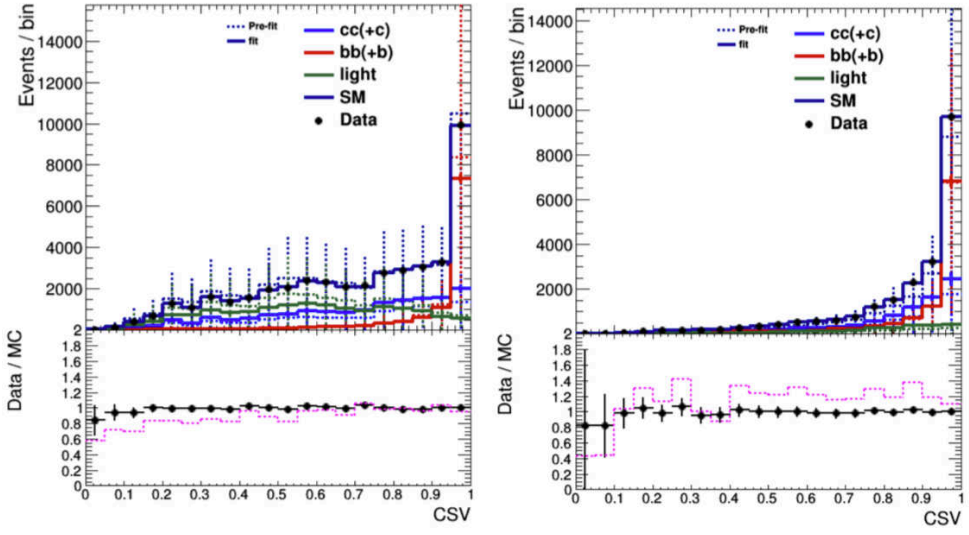


FIGURE 5.13: *Fail* (left) and *pass* (right) CSV distributions used to derive the ZHbbvsQCD SF in the 2017 data taking year for a 2% misidentification rate for jets with p_T above 600 GeV. The lower panels show the ratio between the data (black markers) or the expectation from simulation before the fit (dashed line) and the post-fit distributions. Taken from Ref. [175].

6 | Search strategy and event selection

All the BSM models introduced in Sections 2.3.1–2.3.2 to solve the hierarchy problem predict the existence of new (heavy) particles, with a high branching fraction for the decay to two bosons in several scenarios. Searches for diboson resonances allow to probe many different signal hypotheses at once, thus being a particularly sensitive channel for potential discoveries.

Traditionally, to identify boson-initiated jets, the m_{SD} of the jets is required to fall in an interval around the mass of the considered bosons. Thus, searches considering the $WW/WZ/ZZ$ (also indicated as VV) decay modes are usually performed separately from searches in the WH/ZH (referred as VH) or HH channel. The first step of searches performed in the VV all hadronic final state consists of selecting two jets with high transverse momentum p_{T}. Then, the above mentioned requirement on the m_{SD} is introduced to discriminate the V boson-initiated jets from the QCD multijet background, and constraints on τ_{21} (or similar substructure observables) are used to identify jets with two prongs [158]. The invariant mass of the two jets (*dijet mass* m_{jj}) satisfying these selections is reconstructed as candidate resonance mass. The m_{jj} spectrum of the QCD multijet background is smoothly falling, while a BSM signal would peak at the resonance mass. A *bump hunt* search in the m_{jj} spectrum is thus performed to probe the existence of the new particle.

As Fig. 5.3 shows, signal and QCD multijet background are, respectively, resonant and smoothly falling also in the m_{SD} distribution. The approach developed in Ref. [9, 177, 178] takes advantage of these features to perform a three-dimensional bump hunt in the m_{jj}-m_{jet1}-m_{jet2} plane, removing the range selection requirement on the jet masses. In this way, the sensitivity to resonances decaying to a V pair in the dijet final state is improved by up to 30% with respect to the 1-dimensional approach [9, 177, 178]. The search presented in Ref. [9, 177, 178], which demonstrated the efficacy of the multidimensional approach, targeted only VV decay channels. However, removing the mass categorization allows to simultaneously test for more diboson decay modes such as VV, VH, HH or even containing more exotic particles yet to be discovered. Additionally, it increases the signal detection efficiency, as up to 20% of the reconstructed W, Z, or H bosons may fall outside the traditionally used jet mass windows.

The analysis presented in this thesis extends the method developed in Ref. [9, 177, 178] to hadronic VV and VH final states. Furthermore, dedicated categories have been introduced to increase the sensitivity to resonances produced through Vector Boson Fusion (VBF) and final states with a Higgs boson. While the inclusion of the HH final state would be straightforward, this channel has not been considered here. The 138 fb^{-1} of data analyzed in this thesis have been collected by the CMS Collaboration in 2016 (36 fb^{-1}), 2017 (42 fb^{-1}) and 2018 (60 fb^{-1}).

In this chapter, the analysis methodology and the criteria used to select the phase space of interest are described. Chapter 7 illustrates the signal and background modelling, including the related systematic uncertainties and the multidimensional fit validation procedure. Finally, results are presented in Chapter 8. First, the analysis strategy is explained and the signal and background characteristic signatures are described (Sec. 6.1). Section 6.2 presents the event reconstruction. In Sec. 6.3 the selections applied to jet substructure observables and the optimization of the working points are described. Given the broad set of resonances decay channels and production modes studied in this thesis, dedicated categories have been introduced to maximize the sensitivity to the different signal hypotheses, also introduced in Sec. 6.3.

The work presented in this chapter and in Chapters 7–8 has been developed by the author of this thesis under the supervision of Dr. Andreas Hinzmann, in collaboration with Dr. Clemens Lange and Dr. Jennifer Ngadiuba. The author of this thesis is the principle analyzer for this study, the analysis contact person and regularly reported in the CMS working group meetings. In Chapter 6, the author performed all the presented studies, with the exception of the contributions listed below. The procedure for the tagger decorrelation with 2D maps has been developed by Dr. Anna Benecke, while the author of this work carried out the validation in the analysis framework, produced the final set of maps with the mass decorrelated version of the chosen taggers and performed the optimization of the working points and the event categorization. The trigger study presented in Sec. 6.2.1 has been performed by Dr. Jennifer Ngadiuba and the author performed additional studies leading to the final choice of the threshold requirement. The quality criteria for the identification of VBF jets in Table 6.5 where studied by Dr. Anna Benecke, while the requirements for the classification of ggF/DY and VBF events were optimized by the author of this work.

6.1 Analysis strategy

The analysis aims to firstly discover the new resonances predicted by many BSM theories and secondly set exclusion limits independent of the model, to test a large variety of hypotheses. While the bulk graviton model (Sec. 2.3.1) and the HVT framework (Sec. 2.3.2) are used to produce the signal samples and to set limits, the limits provided are valid for all models predicting spin-0, spin-1 or spin-2 resonances coupling to bosons. The search is performed under the assumption that the decay width of the resonance is smaller than the detector resolution (*narrow-width approximation*), allowing to factorize the resonance production cross section from the decay branching ratio.

The search targets heavy (1.3–6 TeV) resonances decaying to either two vector-bosons (VV channel) or one Higgs and one vector boson (VH channel). The hadronic decays of the bosons are considered, because they have the highest branching fraction. The analysis is performed in a boosted regime, where the two quarks originated from the decaying bosons are merged into a single jet, leading to dijet final states. Jet substructure and b-tagging techniques are needed to identify the V- and H-jets.

The three-dimensional bump hunt concept is illustrated in Fig. 6.1, where the m_{jet1}-m_{jet2}-m_{jj} plane is shown. The orange peaks represents the resonant behaviour of the signal being probed, and the red curve the smoothly falling QCD multijet background.

Two high p_{T} large radius jets satisfying some *tagging* requirements are combined in a dijet system, from which the invariant mass m_{jj} is calculated. In the 1D approach, the m_{jj} distribution was then fitted with an analytical function modelling

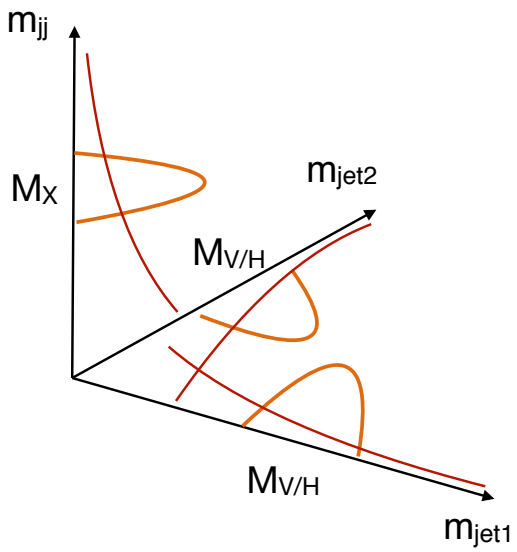


FIGURE 6.1: Schematic representation of the signal (orange) and QCD multijet background (red) in the m_{jet1}-m_{jet2}-m_{jj} plane.

the smoothly falling SM background. Instead, the method presented in this thesis features a three-dimensional maximum likelihood fit to the m_{jet1}-m_{jet2}-m_{jj} space. The signal plus background contributions can be fitted simultaneously, with the background constrained by the regions of the jet mass where no signal is expected.

The removal of the mass categorization allows to simultaneously test different bosonic final states, and the multi-dimensional approach has shown to significantly increase the sensitivity of the search. However, the modelling of the SM backgrounds presents several challenges. For instance, correlations present in the dijet system kinematic between the jet masses and the reconstructed m_{jj} must be taken into account. Since the theoretical predictions of QCD multijet process and the jet substructure modelling in simulation strongly depends on the showering and hadronization models, a data-driven background estimation is employed.

While QCD multijet is the dominant background, other partially resonant processes must be considered, especially V+jets and top pair production ($t\bar{t}$). The latter had a small contribution in the analysis presented in Ref. [9, 177, 178]. In contrast, it is the second most relevant background in the analysis presented in this thesis for the adoption of b-tagged enriched categories, necessary for identifying $H \rightarrow b\bar{b}$ decays, and different taggers used.

The m_{jj} spectrum is smoothly falling for all SM backgrounds. The sub-dominant backgrounds can present peaks in the m_{SD} distribution at the W, Z or top quark mass. However, fitting these real SM particles in the final state helps constrain the uncertainties affecting both the signal and these partially resonant backgrounds.

Resonances produced through Vector Boson Fusion (VBF) can be identified by the presence of two additional small cone jets. Signal events are divided into two categories depending on the presence (VBF category) or absence (ggF/DY category) of such extra jets.

In order to perform the data-driven background estimation, the background (and signal) models contain *nuisance parameters*, whose values are obtained from the fit to the data, i.e. histograms of the two jets masses and the dijet mass. The maximum

likelihood method is used to measure the set of (nuisance) parameters that maximizes the probability of obtaining the analyzed dataset. To either declare the discovery (or the exclusion) of the new particles, a likelihood ratio is used as test statistic, described in the following.

6.1.1 Event topology

The predicted signal cross sections are several orders of magnitude smaller than the cross sections of the SM processes. Dedicated requirements are chosen to reject the background while maximizing the signal detection efficiency. The selections are optimized based on the signal and background topology described below. The full list of samples can be found in Appendix A, where also the cross sections of the different processes are reported. A short description of the generators used for the sample simulation was provided in Sec. 4.2.1.

6.1.1.1 Signals

The theoretical models predicting the signals were introduced in Sec. 2.3.1 and Sec. 2.3.2 for the bulk scenario of the RS warped extra-dimension model and the heavy vector triplet (HVT) framework, respectively. The first predicts spin-0 and spin-2 resonances that can be produced through gluon-gluon fusion (ggF) and vector boson fusion (VBF). The latter predicts spin-1 resonances, whose production modes are quark-antiquark annihilation (Drell-Yan, DY) and VBF. Figure 6.2 shows the Feynman diagrams of the described signals. All simulated signal samples are produced with a relative resonance width of 0.1%, in order to be within the validity of the narrow-width approximation.

For processes produced through ggF and DY, no significant differences are observed in the kinematic distributions of resonances with different spins. While the m_{jj}^{AK8} distributions will be shown in Sec. 6.2.4 comparing signal and background samples, Fig. 6.3 shows the p_T of the leading AK8 jet (left) and the $\Delta\eta$ of the two AK8 boson-jets (right) for ggF/DY produced resonances with three different spins. The η variable was introduced in Eq. 3.1. It depends on the boson scattering angle, while it is independent of the scattering energy. $\Delta\eta \approx 0$ between the two boson-jets indicates that they are back-to-back. On the contrary, signals produced through VBF present polarization effects between the additional jets and the final state bosons, resulting in different distributions depending on the resonance spin, as shown in Fig. 6.4. In Fig. 6.5, p_T (left) and η (right) distributions of the leading VBF (AK4) jet are presented. The m_{jj}^{AK4} and $\Delta\eta^{AK4}$ of the two AK4 VBF-jets are shown later in Fig. 6.14, compared to the background distribution.

While this analysis is not designed to distinguish signals with different spins, it will be shown that the detection efficiency of VBF signals with different spins is differently affected by the background rejection selections.

Overall, the analysis investigates 16 signal hypotheses with masses between 1.3 and 6 TeV, taking into account the resonances' various production and decay modes, as summarized in Table 6.1. The resonances can decay into multiple combinations of boson pairs and, thus, the final results are interpreted also combining different decay modes in Table 6.1 .

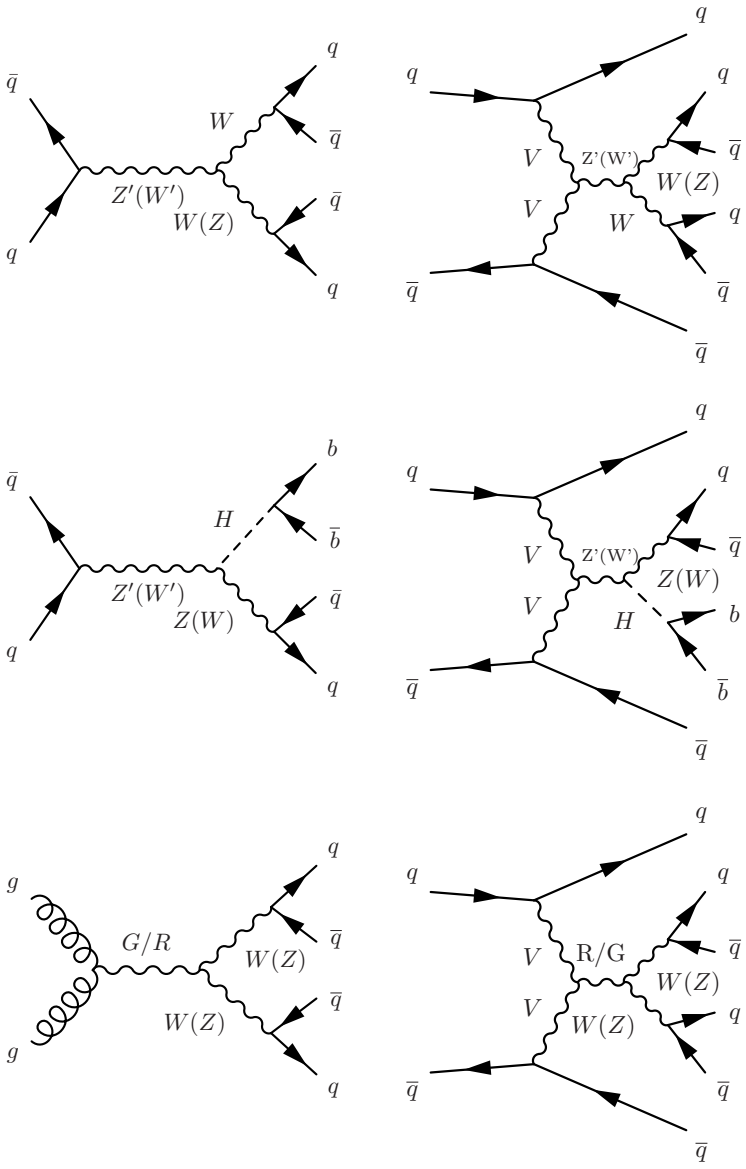


FIGURE 6.2: Feynman diagrams summarizing the production and decay modes of the considered signals. Top (middle): spin-1 resonances produced through quark-antiquark annihilation (left) or vector boson fusion (right) and decaying to two vector bosons (one Higgs and one vector boson) in the hadronic final state. Bottom: spin-2/0 resonances produced through gluon-gluon fusion (left) or vector boson funsion (right) in the VV hadronic final state.

TABLE 6.1: Summary of the possible, and thus considered, signal hypotheses.

	Resonance decay channel				
	WW	WZ	ZZ	WH	ZH
Spin-0	ggF/VBF Radion	-	ggF/VBF Radion	-	-
Spin-1	DY/VBF Z'	DY/VBF W'	-	DY/VBF W'	DY/VBF Z'
Spin-2	ggF/VBF G_{bulk}	-	ggF/VBF G_{bulk}	-	-

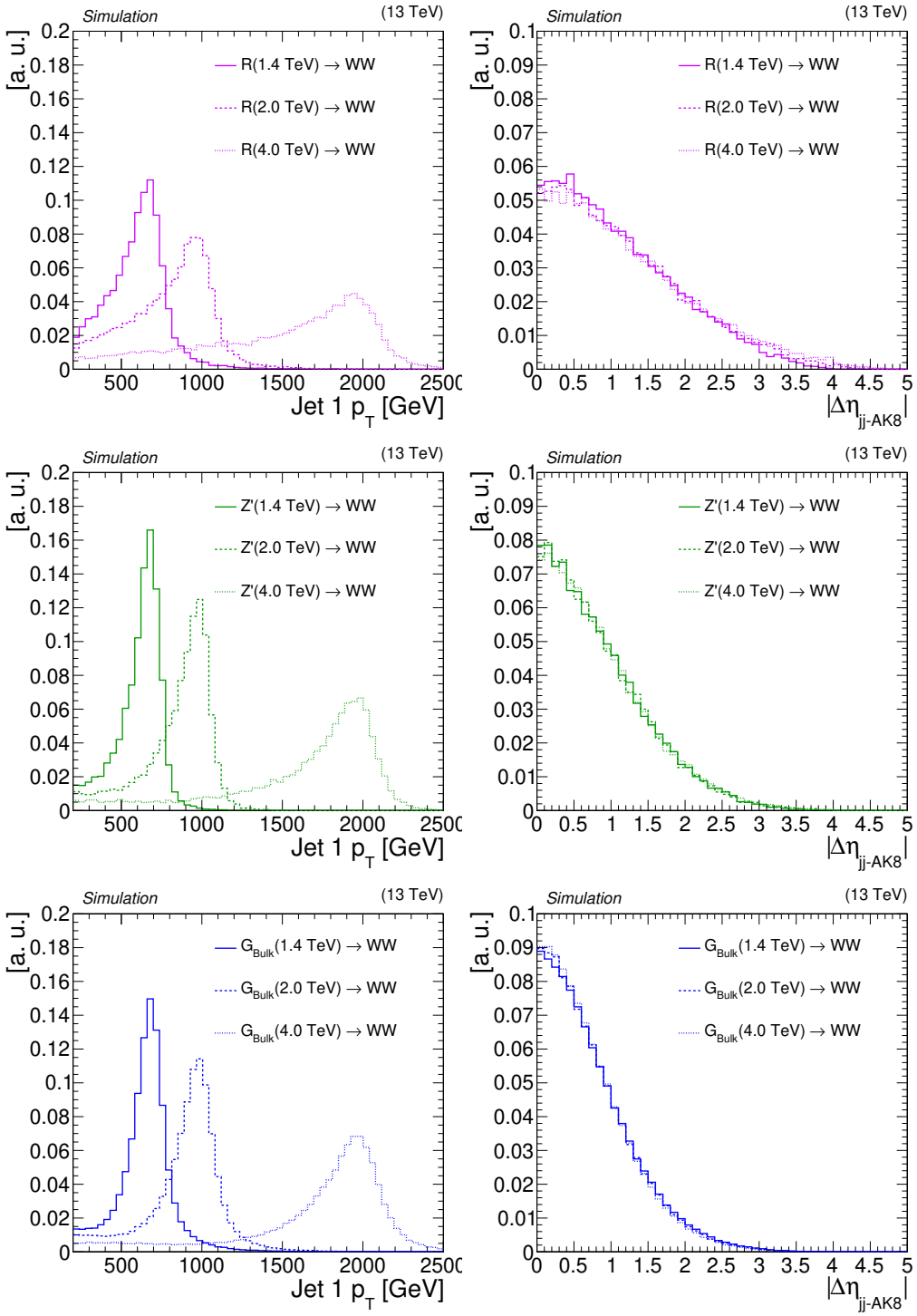


FIGURE 6.3: The p_T of the leading jet (left) and the $\Delta\eta$ of the two jets (right) is shown for ggF/DY produced resonances with three different spins, spin-0 (top), spin-1 (middle) and spin-2 (bottom).

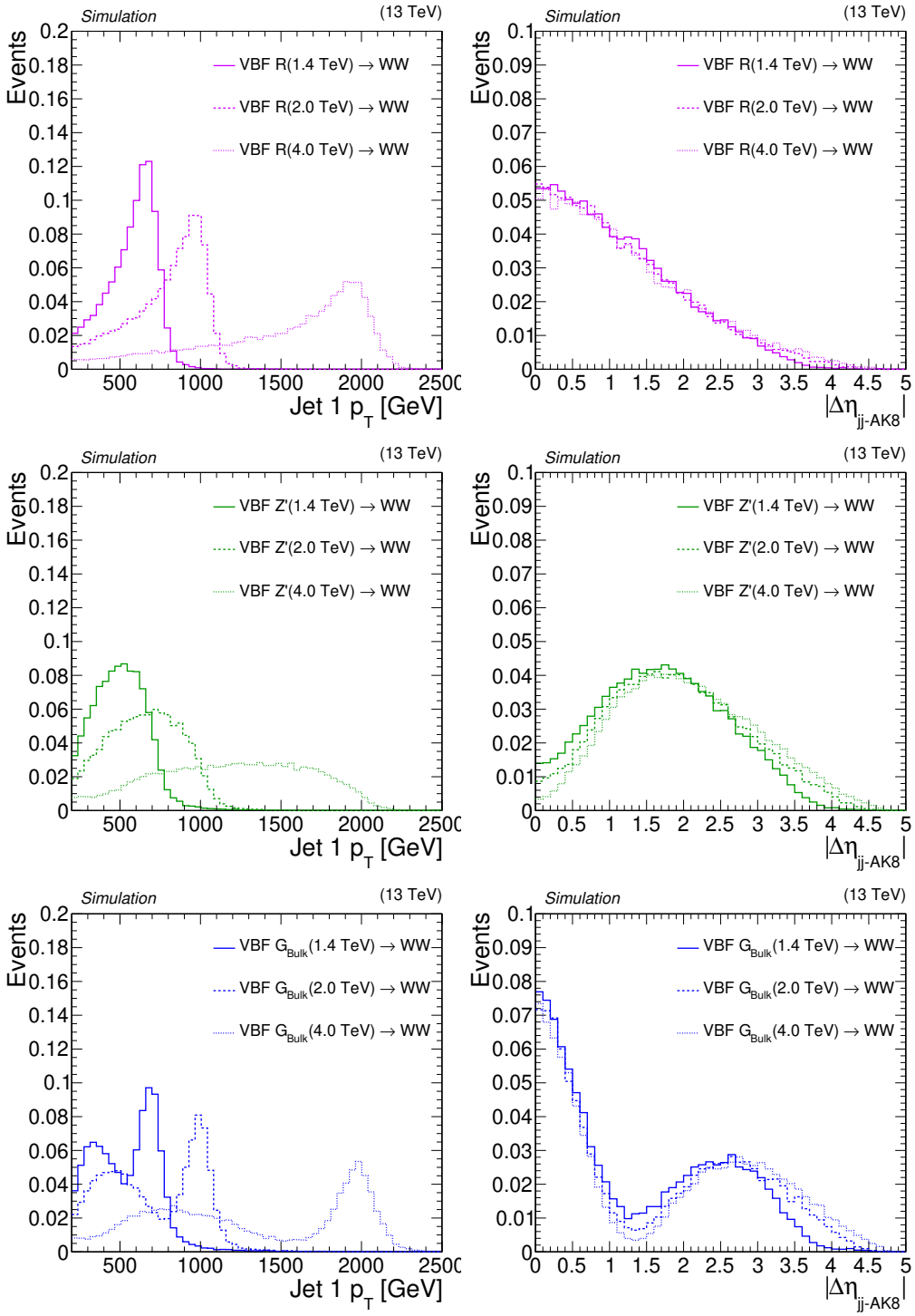


FIGURE 6.4: The p_T of the leading jet (left) and the $\Delta\eta$ of the two jets (right) is shown for VBF produced resonances with three different spins, spin-0 (top), spin-1 (middle) and spin-2 (bottom).

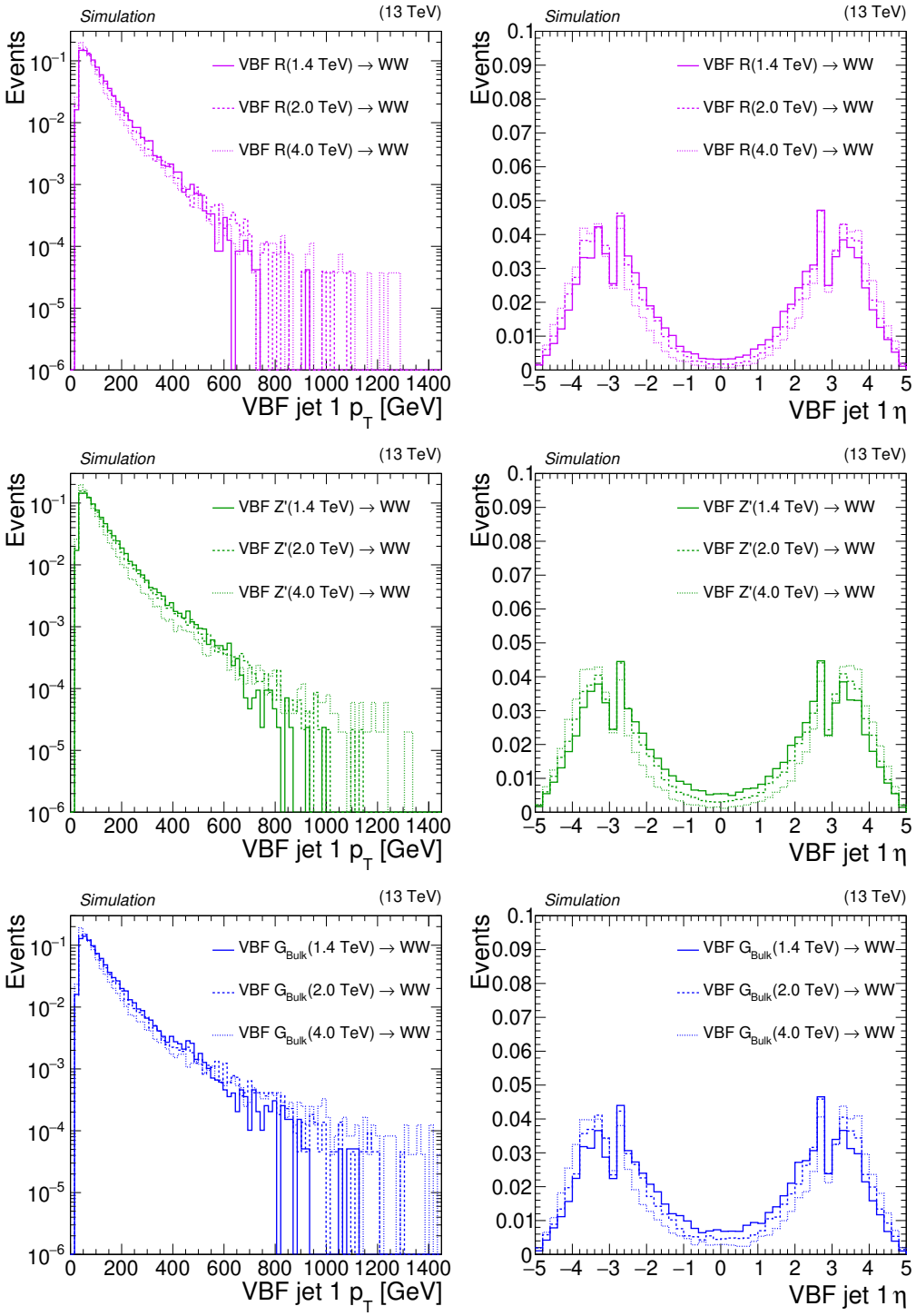


FIGURE 6.5: The p_T (left) and the η (right) distribution of the leading VBF jet are shown for VBF produced resonances with three different spins, spin-0 (top), spin-1 (middle) and spin-2 (bottom).

6.1.1.2 SM backgrounds

The relevant SM backgrounds are listed below, in order of importance for the analysis:

- **QCD multijet** describes all processes that produce two or more jets in the final state. It presents a smoothly falling spectrum in both the mass spectrum of the jets and in the invariant mass distribution of the reconstructed dijet system. Selected LO Feynman diagrams contributing to this process are shown in Fig. 6.6.
- **Top pair production ($t\bar{t}$)**. The diagrams illustrating the $t\bar{t}$ production are shown in Fig. 6.7 (top). The top quark decays almost exclusively into a b-quark and a W boson. With the addition of b-tagging requirements to select the $H \rightarrow b\bar{b}$ decay, this process is the second most relevant background. In the hadronic top decay, i.e. $t \rightarrow bW \rightarrow bq\bar{q}$, the final state has three quarks that can be merged in different combinations depending on the top p_T, as was shown in Eq. 5.1. Thus, the top-originating reconstructed jets present a resonant behaviour with a peak at the top mass or at the W mass or can be non-resonant. The fraction of the different contributions depends on the top p_T, and it is parametrized accordingly. When considering both top-quarks in the event (see Fig. 6.7 bottom), six different contributions are possible, shown in Fig. 6.8. The dijet mass spectrum of the $t\bar{t}$ system is smoothly falling.
- **V+jets** events consist of a real V boson and a QCD-like jet. While the cross section of this process is small compared to QCD or $t\bar{t}$ backgrounds, the V-tagging requirements enhance the contribution from this background. The real boson in the event appears as a resonance in one of the two jet masses. It is used to constrain major systematic uncertainties affecting both signal and background, such as the tagging efficiency and the jet mass resolution. Relevant production diagrams for this contributions are shown in Fig. 6.9.
- **Single top** and (SM) **WW** production contribute less than 1.5% and 0.5%, respectively, in the analysis phase space and are included as a part of the $t\bar{t}$ background model. The graphs illustrating the production mechanism of these backgrounds are presented in Fig. 6.10. Given the small production cross section, contributions from (SM) WZ, ZZ, WH and ZH processes in the analysis phase-space are well covered by the resonant background normalization uncertainties and thus these backgrounds are not included.

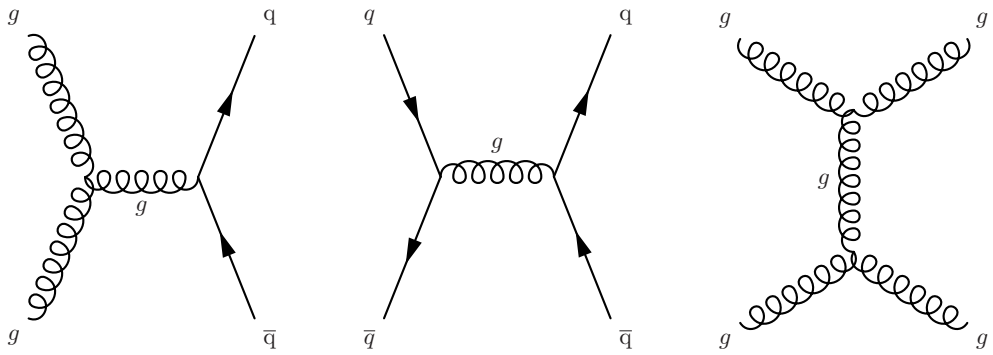


FIGURE 6.6: Selected LO Feynman diagrams contributing to the QCD multijet background.

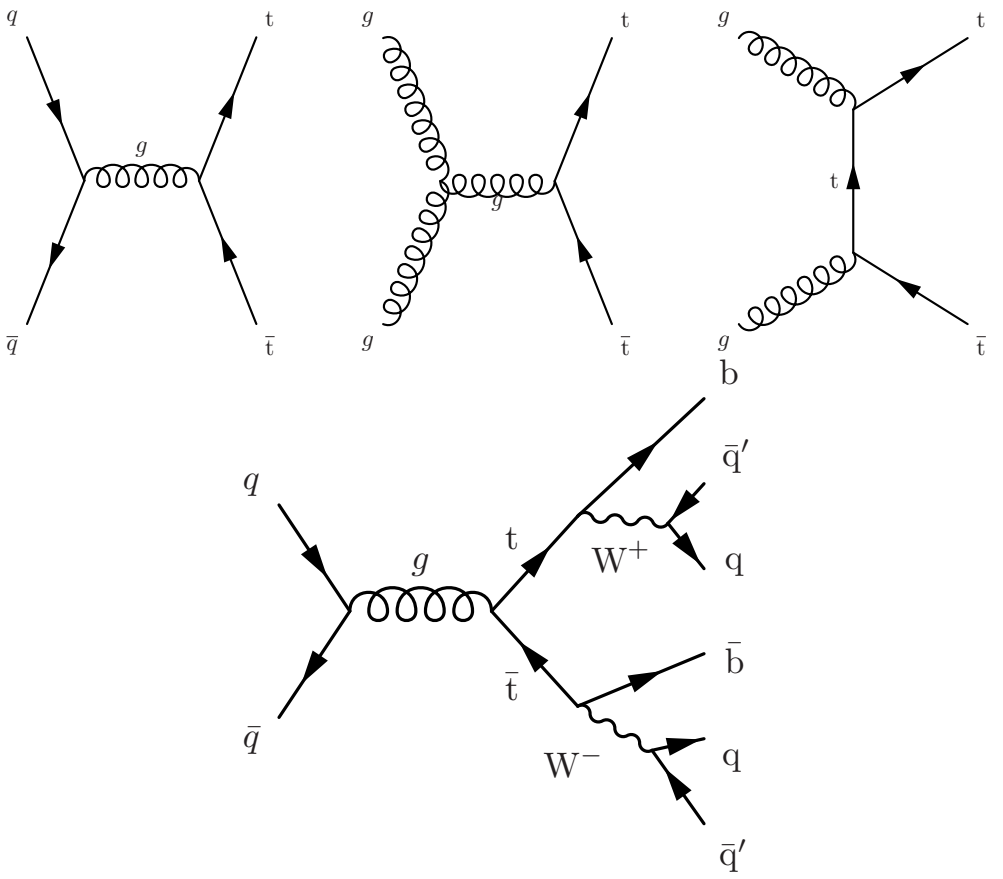


FIGURE 6.7: Top: LO Feynman diagrams contributing to the $t\bar{t}$ background. Bottom: Top quark pair decaying into the all-hadronic final state.

6.2 Event selection and reconstruction

In this section, the event selection criteria applied to ensure high-quality data sample and event reconstruction are summarized. Preselection requirements to reduce the SM background contributions based on the event topology described in Sec. 6.1.1 are also introduced. Section 6.3 describes the full list of requirements.

6.2.1 Trigger selections

Given the high rate of QCD multijet production, high p_T thresholds are applied on jets at the trigger stage to reduce the number of events to a recordable level. Therefore, typically m_{jj} above ≈ 1 TeV are accessible after trigger requirements, and the analysis is performed in a boosted regime.

In this analysis, the online event selection is based on the jet with the highest p_T in the event, the sum of the p_T of all jets in the event (H_T), and jet mass based triggers. Jet mass triggers pose additional requirements on the trimmed[1] [179] jet mass. For the three years of data taking, different online thresholds for unprescaled triggers have been used, with higher ones for the later periods.

H_T-triggers are applied on a collection of anti-k_T jets with radius parameter $R = 0.4$ (AK4), while mass and p_T triggers on AK8. The jet mass trigger requires the trimmed jet mass to be higher than 30–50 GeV and the event H_T to be above a certain

[1] Trimming reclusters jets from subjets passing a certain threshold in p_T.

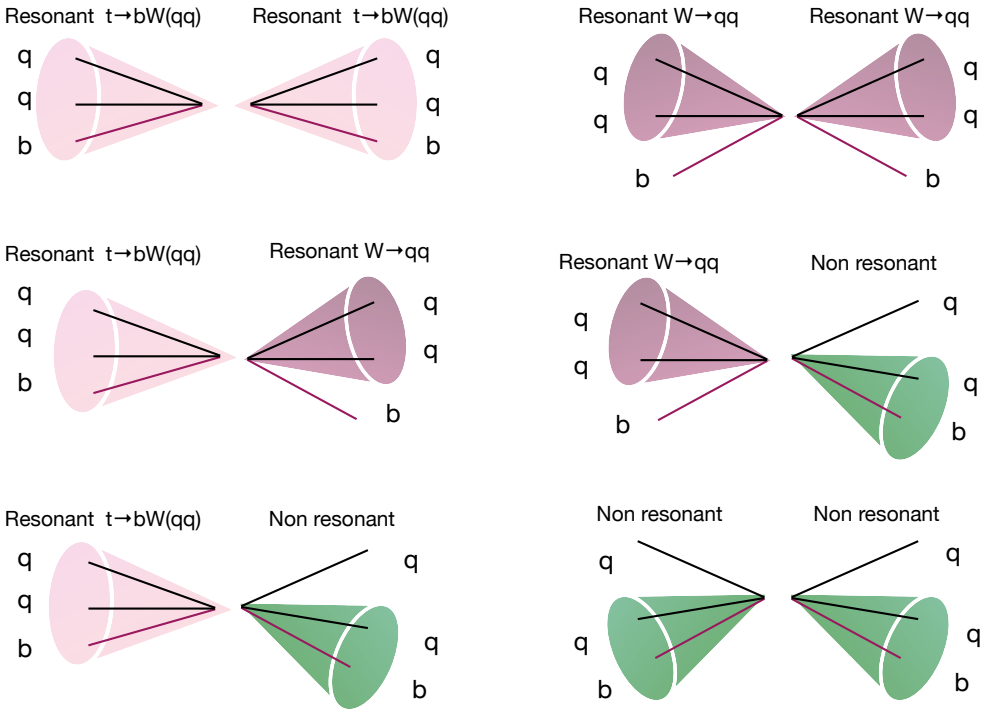


FIGURE 6.8: The six different possibility of reconstructed jets in the $t\bar{t}$ hadronic decay. In the hadronic top decay ($t \to bW \to bq\bar{q}$), the final state has three quarks that can be merged in three different combinations depending on the top p_T. The fraction of the different contributions depends on the top p_T. When considering both top-quarks in the event, six different contributions are possible. For the non-resonant contribution, the case in which only one quark is included in the chosen jet is possible and considered in the analysis, but not drawn.

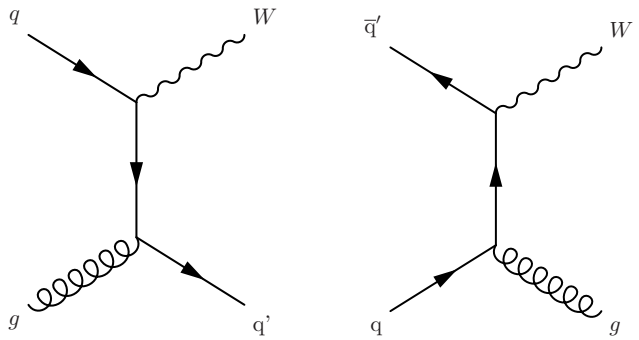


FIGURE 6.9: LO Feynman diagrams contributing to the W+jet background. Similar processes for the Z+jet production are also considered in this analysis.

threshold. The introduction of the trimmed jet mass trigger allows lowering the jet p_T threshold, thus decreasing the analysis m_{jj} threshold.

A summary of the triggers for the three years is presented in Tables 6.2–6.3.

The analysis threshold is set by the trigger turn-on point (where the combination of all triggers are $> 99\%$ efficient). The trigger turn-on is evaluated in a muon enriched dataset separately for the three years, using the `HLT_Mu50` and `HLT_IsoMu27` paths as reference triggers. For these measurements all the analysis preselections

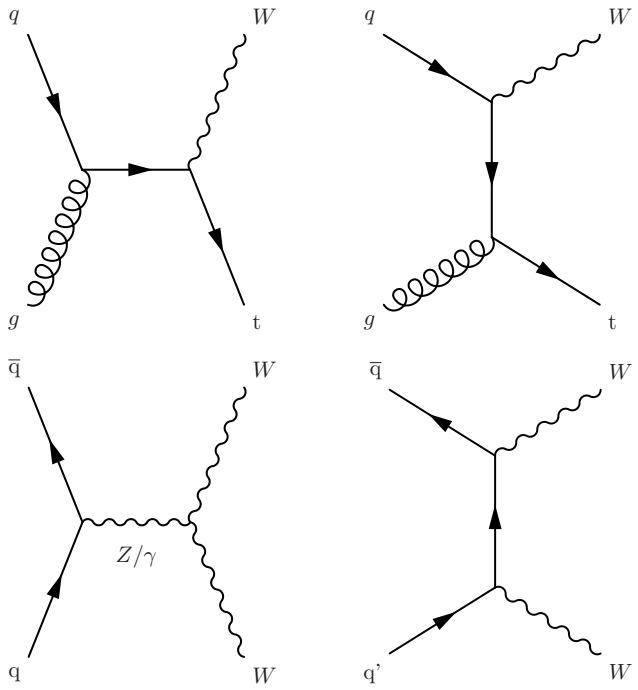


FIGURE 6.10: LO Feynman diagrams contributing to the single top (top) and WW (bottom) backgrounds.

TABLE 6.2: Jet p_T and H_T based triggers used to collect data online during the three data-taking years.

Period	H_T trigger	jet p_T trigger
2016 Runs B–G	HLT_PFHT800 HLT_PFHT650_WideJetMJJ900DEtaJJ1p5	HLT_PFJet450
2016 Runs H	HLT_PFHT900 HLT_PFHT650_WideJetMJJ900DEtaJJ1p5	HLT_PFJet450 HLT_AK8PFJet450
2017 Run B	HLT_PFHT1050	HLT_AK8PFJet500
2017 Runs C–D	HLT_PFHT1050	HLT_AK8PFJet500
2017 Runs E–F	HLT_PFHT1050	HLT_AK8PFJet500
2018 Runs A–D	HLT_PFHT1050	HLT_AK8PFJet500

described in Sec. 6.2.3 are applied. The trigger efficiency as a function of the dijet invariant mass using a combination of all triggers (left), and as a function of the jet soft drop mass for triggers based only on the jet mass (right) are shown in Figure 6.11 for the three years.

The trimming-based triggers were not available in the first 2017 data-taking period (corresponding to 4.8 fb^{-1}), causing the dataset to not fully reach the trigger plateau. The corresponding turn-on curves excluding 2017 Run B are also shown in Figure 6.11. With the jet substructure triggers not being part of the menu for part of 2017 data-taking period, and due to the higher online thresholds to handle the higher data rate, the trigger threshold in 2017 is significantly higher than in 2016. During the 2018 data-taking, the triggers thresholds were kept the same as in the later periods of 2017, however, since jet substructure triggers were online for the whole period the efficiency is slightly higher than in 2017. The combination of all triggers are $> 99\%$ efficient above a dijet invariant mass of 992 GeV, 1226 GeV, and 1131 GeV, for the 2016, 2017, and 2018 dataset, respectively. In order to minimize trigger turn-on effects, the threshold is set to the first m_{jj} bin used in the previous version of this analysis [9] where the 99% efficiency turn on point is achieved for all

TABLE 6.3: Jet mass triggers used to collect data online during the three data-taking years.

Period	H_T + jet mass	jet p_T + jet mass
2016 Runs B–G	HLT_AK8PFHT700_TrimR0p1PT0p03Mass50	HLT_AK8PFJet360_TrimMass30
2016 Runs H	HLT_AK8PFHT700_TrimR0p1PT0p03Mass50	HLT_AK8PFJet360_TrimMass30
2017 Run B	–	–
2017 Runs C–D	HLT_AK8PFHT750_TrimMass50	HLT_AK8PFJet360_TrimMass30
2017 Runs E–F	HLT_AK8PFHT800_TrimMass50	HLT_AK8PFJet400_TrimMass30
2018 Runs A–D	HLT_AK8PFHT800_TrimMass50	HLT_AK8PFJet400_TrimMass30

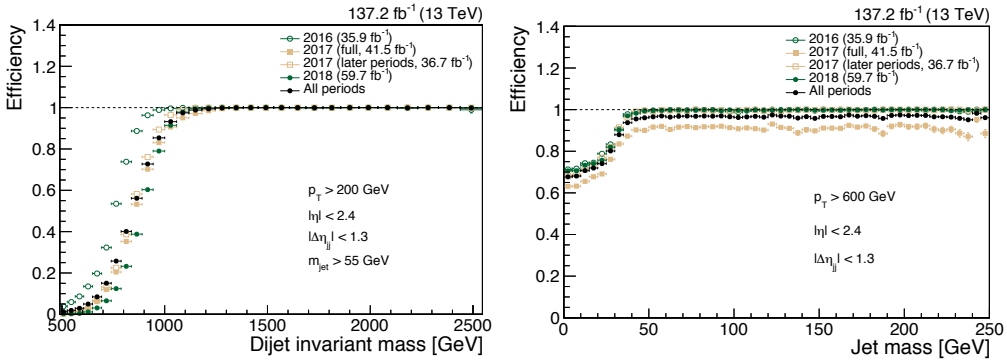


FIGURE 6.11: Left: trigger efficiency as a function of the dijet invariant mass using a combination of all analysis triggers. Right: trigger efficiency as a function of the jet softdrop mass for triggers requiring an online trimmed mass of at least 30 or 50 GeV.

three datasets, which is found to be at 1246 GeV.

During the 2016 and 2017 data taking, the CMS L1 trigger was affected by a slowly increasing shift of the reconstructed cluster time in the ECAL, predominantly at high η and more probable for higher p_T jets. This led to a loss of a fraction of events in the trigger, as these clusters could be assigned to the wrong LHC bunch crossing ("trigger prefiring") [180]. As this effect is not present in simulation, simulated events are reweighted to account for this inefficiency: the event weight is calculated from the probability of each AK4 jet (or photon) to cause the prefiring. A dedicated uncertainty estimates the effect of the weights on the signal yields, described in Sec. 7.3.

6.2.2 Noise filters and vertex selection

Instrumental effects such as anomalous signals in HCAL or energy deposits in ECAL can result in events in which the missing transverse energy (E_T^{miss}, the imbalance of energy measured in the transverse direction) is misreconstructed. To remove such events, additional selection criteria are applied to each jet [181].

Additionally, events must contain at least one primary vertex reconstructed within a 24 cm window along the beam axis, with a transverse distance from the nominal pp interaction region of less than 2 cm [182]. In the presence of more than one vertex passing these requirements, the reconstructed vertex with the largest value of summed physics-object p_T^2 is taken to be the primary pp interaction vertex.

6.2.3 Event reconstruction and preselection

Event reconstruction is based on the particle flow (PF) algorithm [183] described in Sec. 4.3, which reconstructs and identifies individual particles with information from the various elements of the CMS detector. Jets are reconstructed from PF particles,

$\lvert\eta\rvert < 2.4$	2016	2017	2018
Neutral Hadron Fraction	< 0.90	< 0.90	< 0.90
Neutral EM Fraction	< 0.90	< 0.90	< 0.90
Number of Constituents	> 1	> 1	> 1
Charged Hadron Fraction	> 0	> 0	> 0
Charged Multiplicity	> 0	> 0	> 0
Charged EM Fraction	< 0.99	–	–

TABLE 6.4: Jet ID criteria per year applied to AK8 CHS jets with $|\eta| < 2.4$.

using the anti-k_T jet clustering algorithm [144], illustrated in Sec. 5.1. Two different distance parameters are used. For the boson reconstruction, jets with $R = 0.8$ (AK8 jets) are used. For the VBF jets, where no substructure is expected, $R = 0.4$ (AK4 jets) is chosen.

Different pileup mitigation techniques are used in the analysis. As illustrated in Sec. 4.4.3, the PUPPI algorithm provides the best performance for jet mass and substructure observables. Therefore jet substructure observables, such as m_{SD}, are reconstructed from jet clustered after the PUPPI algorithm has been applied to the event. However, it was also shown that for high p_T jets, the CHS algorithm achieves a better momentum resolution. A similar effect has been observed also in the m_{jj} spectrum, where the CHS method obtains a $\approx 10\%$ better resolution [177, 178]. Therefore, CHS jets are used for the event reconstruction and jet substructure observables are then obtained from the PUPPI softdrop jet matched to the ungroomed AK8 CHS jet with the smallest angular separation ΔR. All AK8 jets are further required to satisfy the quality requirements provided in Ref. [81] and listed in Tab 6.4 for the three different data-taking periods.

VBF jets present a softer p_T spectrum and are usually located in the forward region of the detector. Therefore, it is non-trivial to distinguish them from pileup jets. Furthermore, higher noise level were present in the region $2.5 < |\eta| < 3$ of the ECAL (Sec. 3.2.1) in 2017. This effect, shown in Fig. 6.12, is mitigated when PUPPI is used. Thus, PUPPI AK4 jets are used to reconstruct the VBF jets and partially remove the effect. Jet energy calibration uncertainties are also larger in that pseudorapidity region to cover the remaining part of the issue. Moreover, the noise only affects the number of jets identified as VBF jets but does not change the m_{jj}-m_{jet1}-m_{jet2} distribution. Given the importance that the AK4/VBF jets have in this analysis and the challenges arising in the discrimination of such jets from pileup jets, the recommended quality criteria have been adapted to increase the detection efficiency of such jets as summarized in Table 6.5.

The energy of the reconstructed jets is corrected for detector non-linearities in p_T and η using the standard jet energy corrections and the four-momentum of jets is smeared in simulation so that the jet p_T resolution matches the one observed in data. The jet energy correction procedures are described in Sec. 5.2.

Events are then selected by requiring at least two AK8 jets with $p_T > 200\,\text{GeV}$ and $|\eta| < 2.4$. The two jets with the highest p_T in the event are selected as potential boson candidates.

The events analyzed in this search must not overlap with the ones used in complementary searches in the semi-or all-leptonic decay channels to allow the future combination of the results. Therefore, events are rejected if at least one of the two jets has an angular separation ΔR smaller than 0.8 from any electron or muon in the event. Leptons used for this veto need to have a p_T greater than $35(30)\,\text{GeV}$, an

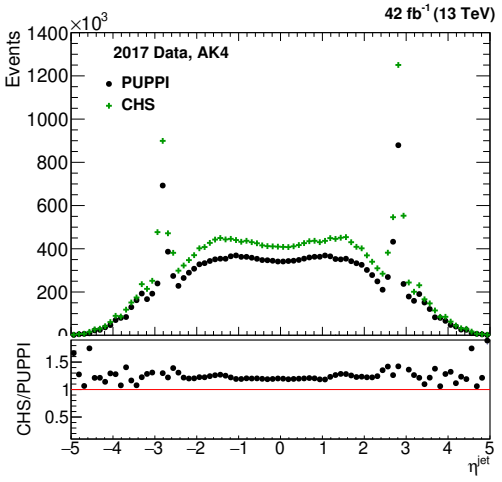


FIGURE 6.12: η distribution of the two p_T-leading AK4 jet of 2017 data with the ECAL noise problem between $2.5 < |\eta| < 3$.

absolute pseudorapidity smaller than 2.5 (2.4), and pass quality criteria that were optimized for high-momentum electrons (muons). In particular, these working points have been chosen to match the selections used in the search for high-mass WV/WH diboson resonances decaying in the semi-leptonic channel [63].

The two AK8 jets are further required to have a separation of $|\Delta\eta_{\mathrm{AK8}}| < 1.3$ in order to considerably reduce the QCD multijet background. This selection has been implemented also for consistency with the previous searches for VV resonances [9, 158, 177, 178], which did not consider the VBF production mode of the resonances. Figure 6.13 compares the $|\Delta\eta_{\mathrm{AK8}}|$ distribution of the QCD background with different signal hypotheses, showing that this requirement is not optimized for VBF produced resonances. The η distribution of the high p_T jets of VBF produced resonances is strongly influenced by the particle spin, as shown in Fig. 6.4. As a consequence of the $|\Delta\eta_{\mathrm{AK8}}| < 1.3$ requirement, a loss of signal efficiency is observed for VBF produced spin-2 and, especially, spin-1 resonances. However, the QCD multijet background is significantly reduced by this requirement.

Additionally, the two AK8 jets must have an invariant mass $m_{\mathrm{jj}} > 1246\,\mathrm{GeV}$ in order to be on the trigger efficiency plateau (see Section 6.2.1). Moreover, it is required that the two AK8 jets in the event have a soft drop mass in the range $55\,\mathrm{GeV} < m_{\mathrm{SD}} < 215\,\mathrm{GeV}$ as further discussed in Sec. 6.3. Finally, a loose requirement of $\rho = \ln((m_{\mathrm{jet}}^{\mathrm{AK8}})^2/p_T^2) < -1.8$ is applied in order to veto events in which the jet mass is high, but the jet p_T is low. In these cases the cone size of $\Delta R = 0.8$ is too small to contain the full jet, affecting both the jet mass resolution and the $\mathrm{q\bar{q}}$ tagging efficiency, which is not well modelled in simulation.

A summary of the preselections applied to AK8 jets is as follows:

- PF jet Tight ID applied
- Jet $\eta < 2.4$
- Jet $p_T > 200$ GeV
- $|\Delta\eta_{jj}| < 1.3$
- $m_{\mathrm{jj}} > 1246\,\mathrm{GeV}$
- $55\,\mathrm{GeV} < m_{\mathrm{SD}} < 215\,\mathrm{GeV}$

TABLE 6.5: Jet ID criteria applied to AK4 PUPPI jets.

2016				
	$\|\eta\| < 2.4$	$\|\eta\| < 2.7$	$2.7 < \|\eta\| < 3$	$3 < \|\eta\| < 5$
Charged Hadron Fraction	> 0	–	–	–
Charged Multiplicity	> 0	–	–	–
Neutral Hadron Fraction	< 0.90	< 0.90	< 0.9999	> 0.02
Neutral EM Fraction	< 0.90	< 0.90	–	< 0.90
Number of Constituents	> 1	> 1	–	–
2017				
	$\|\eta\| < 2.4$	$\|\eta\| < 2.7$	$2.7 < \|\eta\| < 3$	$3 < \|\eta\| < 5$
Charged Hadron Fraction	> 0	–	–	–
Charged Multiplicity	> 0	–	–	–
Neutral Hadron Fraction	< 0.90	< 0.90	< 0.9999	> 0.02
Neutral EM Fraction	< 0.90	< 0.90	–	< 0.90
Number of Constituents	> 1	> 1	–	–
2018				
	$\|\eta\| < 2.6$	$2.6 < \|\eta\| < 2.7$	$2.7 < \|\eta\| < 3$	$3 < \|\eta\| < 5$
Charged Hadron Fraction	> 0	–	–	–
Charged Multiplicity	> 0	–	–	–
Neutral Hadron Fraction	< 0.9	< 0.9	< 0.9999	> 0.02
Neutral EM Fraction	< 0.9	< 0.99	–	< 0.9
Number of Constituents	> 1	–	–	–

- $\rho = \ln((m_{\text{jet}}^{\text{AK8}})^2/p_{\text{T}}^2) < -1.8$.

Additional requirements are needed to select the VBF production mode. Since VBF (AK4) jets are expected to be in the forward region, η up to 5 are considered. VBF jets are required to have $p_{\text{T}} > 30\,\text{GeV}$ to take into account the softer p_{T} spectrum. If more than two such candidates are present in the event, the two jets with the highest p_{T} are selected. The selected leading AK4 jets will not be considered in the following if they have an overlap with AK8 jets in a radius $\Delta R < 1.2$.

Additional selections are then implemented to reduce the QCD multijet background based on the event topology: the two AK4 jets must have a separation of $|\Delta\eta^{\text{VBF}}| > 4.5$ and their invariant mass m_{jj}^{VBF} is required to be above 800 GeV. Figure 6.14 shows the $\Delta\eta^{\text{VBF}}$ and m_{jj}^{VBF} distributions before the above mentioned requirements are applied. The red lines indicate the cut values, chosen to maximize the signal-to-background ratio and achieve the best sensitivity. The selection optimization was performed by the author of this thesis in the context of a 1-D search based on Ref. [158] with the 2016 dataset and is described in the CMS internal note in Ref. [184] .

Events not passing the additional VBF selections are included in the ggF/DY dataset forming two orthogonal event samples.

6.2.4 Data to simulation comparison

This section presents comparisons of data and simulation for the relevant kinematic distributions and jet tagging observables after the ggF/DY and VBF channel selections described in the previous section are applied. For the former case, signals produced through quark-antiquark annihilation (W' and Z' signals) or gluon fusion

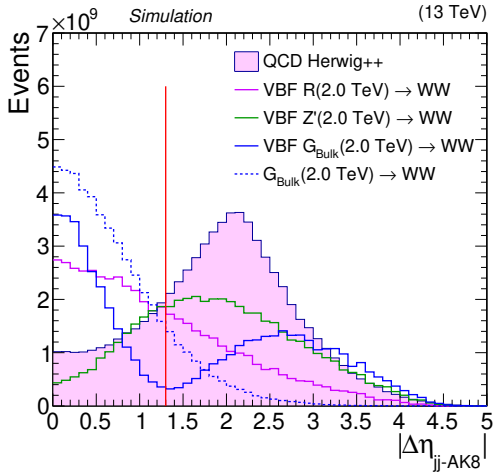


FIGURE 6.13: $\Delta\eta_{\text{jj-AK8}}$ distributions for 2018 QCD HERWIG++ simulation (filled pink area) and three VBF produced signals with different spins (solid lines). A ggF/DY signal is also overlaid for comparison (dashed line). The signals are rescaled by an arbitrary amount for visibility. The cut value applied on the $\Delta\eta$ of the AK8 jets is indicated by the vertical red line.

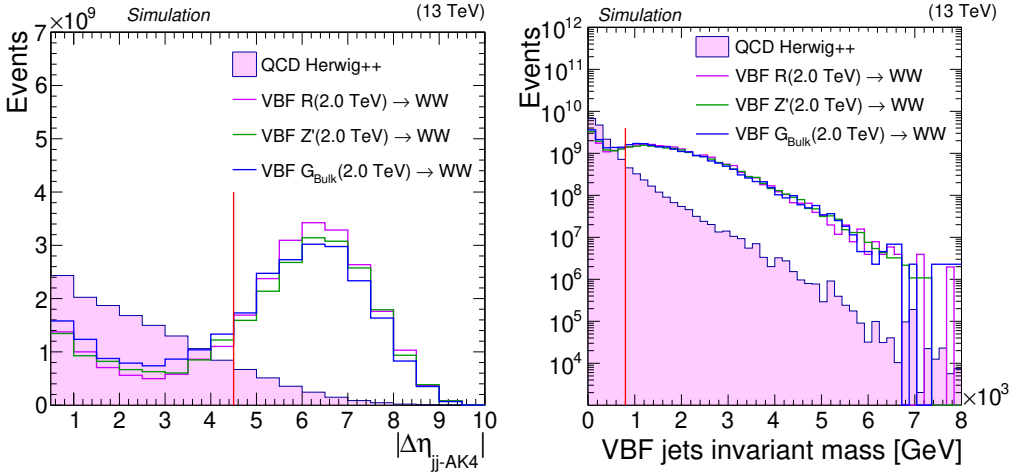


FIGURE 6.14: $\Delta\eta^{\text{VBF}}$ (left) and m_{jj}^{VBF} (right) distributions for 2018 QCD HERWIG++ simulation (filled pink area) and three VBF produced signals. The optimized cut values are indicated by the vertical red lines.

(G_{bulk}) are also shown, while for the latter case the signals produced through vector boson fusion are overlaid. In all the plots, the signals are scaled by an arbitrary number for visibility. The data corresponding to the sum of the three 2016, 2017, and 2018 datasets is compared to the sum of the three sets of MC samples, each weighted by the corresponding integrated luminosity. The contributions from minor backgrounds such as V+jets processes and $t\bar{t}$ processes are scaled to their SM expectation. Given the large uncertainties on the modelling of the QCD multijet background and strong dependence on the generators, the MC simulation of this process is scaled to the data, after the expected events from minor backgrounds have been subtracted. Overall a good agreement between data and simulation is observed, especially in the variables used for the final fit (the m_{SD} of the two jets and the dijet invariant mass). The VBF sample has less events compared to the ggD/DY one. For the QCD multijet simulation, the three generators (PYTHIA, MADGRAPH5_aMC@NLO interfaced

with PYTHIA and HERWIG++) introduced in Sec. 4.2.1 are shown. Differences in their distributions are observed. The first and second leading jets have similar distributions, thus only the ones for the jet with the highest p_{T} are shown, unless otherwise specified.

6.2.4.1 Kinematics observables

Figure 6.15 shows the transverse momentum p_{T} (top), the pseudorapidity η (middle) and the azimuthal angle ϕ (bottom) of one of the two p_{T}-leading selected AK8 CHS jets in the event in the ggF/DY (left) and VBF (right) channel.

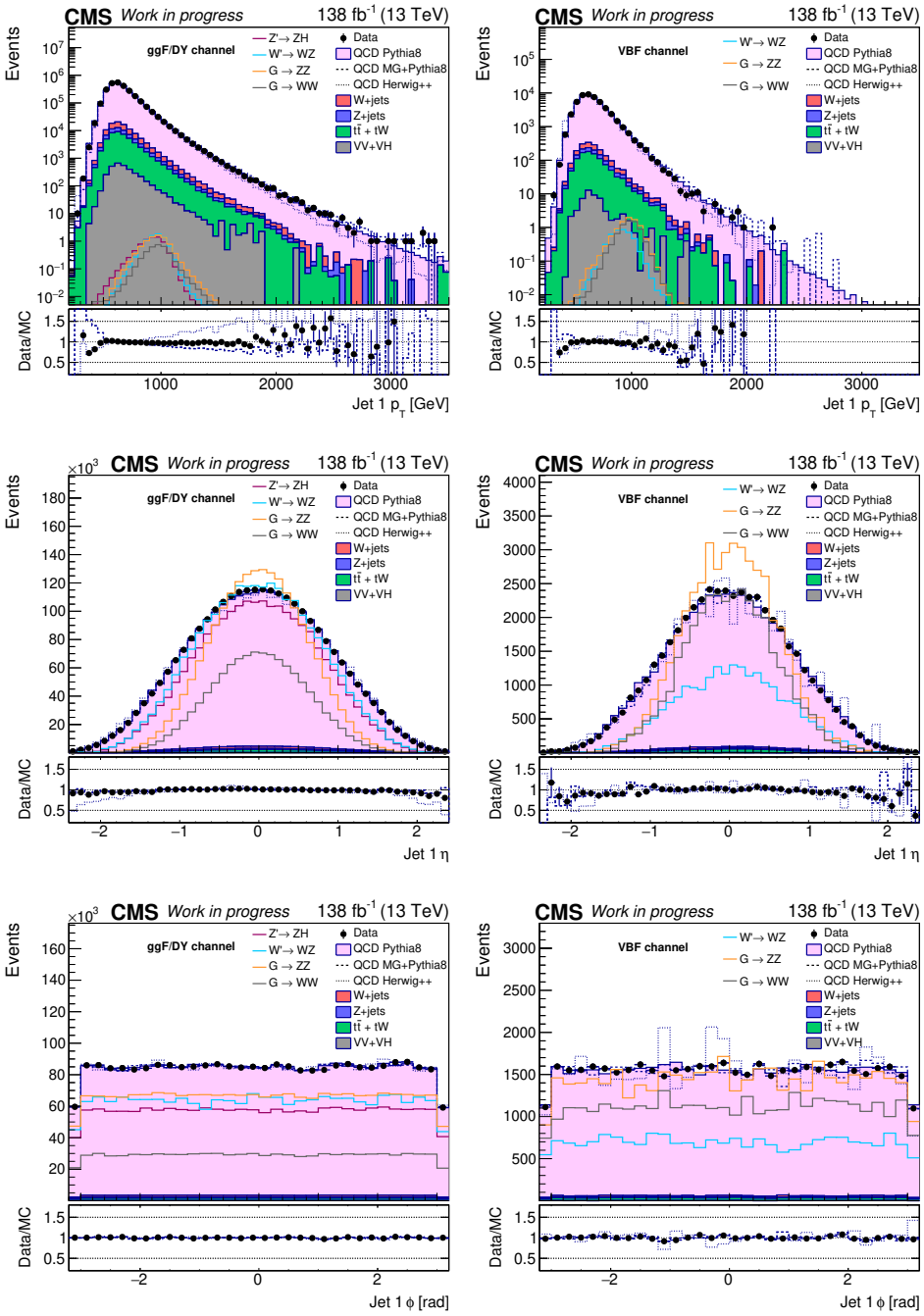


FIGURE 6.15: Transverse momentum p_T (top), pseudorapidity η (middle) and azimuthal angle ϕ of on of the p_T-leading selected AK8 CHS jets in the event in the ggF/DY (left) and VBF (right) channel. For the QCD multijet simulation, several alternative predictions are shown. The simulation is rescaled as described in the text. The ratio plots show the fraction of data over QCD multijet simulation for PYTHIA (black marker), MADGRAPH5_aMC@NLO interfaced with PYTHIA (dashed line) and HERWIG++ (dotted line).

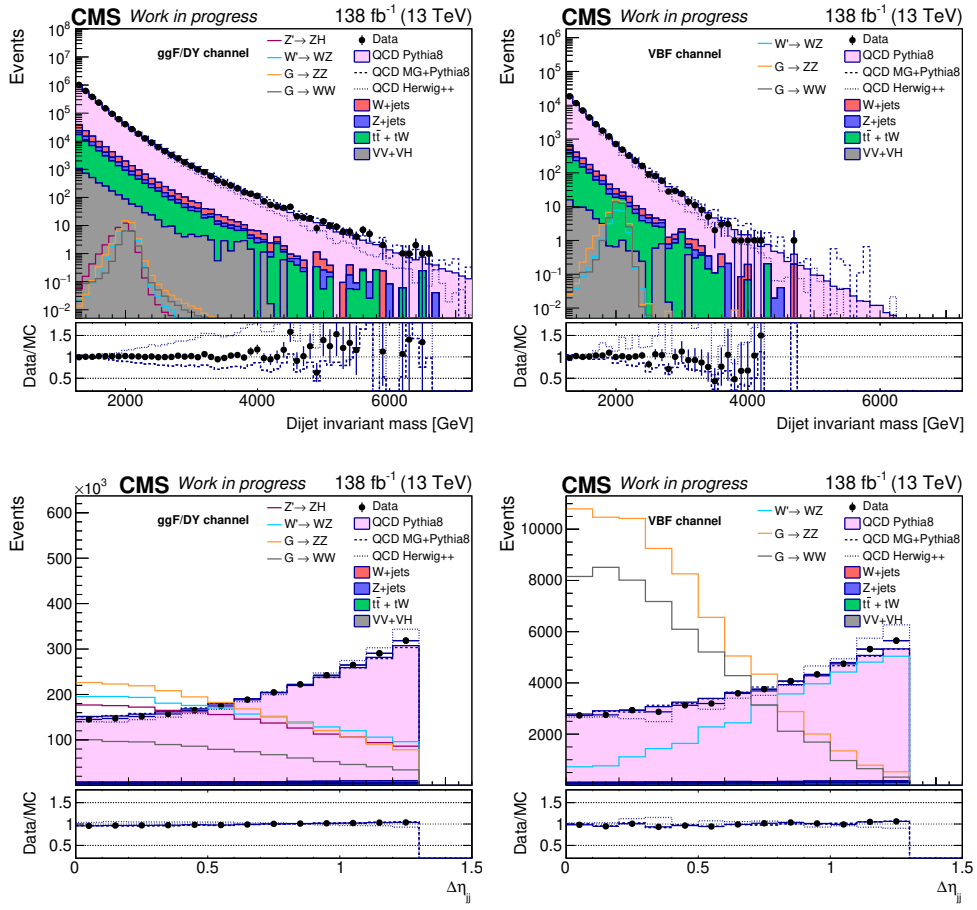


FIGURE 6.16: The invariant mass (top) and $|\Delta\eta_{jj}|$ (bottom) for the two selected AK8 CHS jets in the event in the ggF/DY (left) and VBF (right) channel. For the QCD multijet simulation, several alternative predictions are shown. The simulation is rescaled as described in the text. The ratio plots show the fraction of data over QCD multijet simulation for PYTHIA (black marker), MADGRAPH5_aMC@NLO interfaced with PYTHIA (dashed line) and HERWIG++ (dotted line).

Figure 6.16 (top) shows the invariant mass (m_{jj} for the multi-dimensional fit) for the selected AK8 CHS jets in the event in the ggF/DY (left) and VBF (right) channel. A good agreement between data and simulation is observed. Figure 6.16 (bottom) shows the $|\Delta\eta_{jj}|$ distribution. Figure 6.17 shows the relevant distributions characterizing the kinematics of the VBF jets, where the simulation does not describe the data to a perfect agreement. However, these variables are not used as observables in the fitting procedure and the observed disagreement does not constitute a limitation to the analysis.

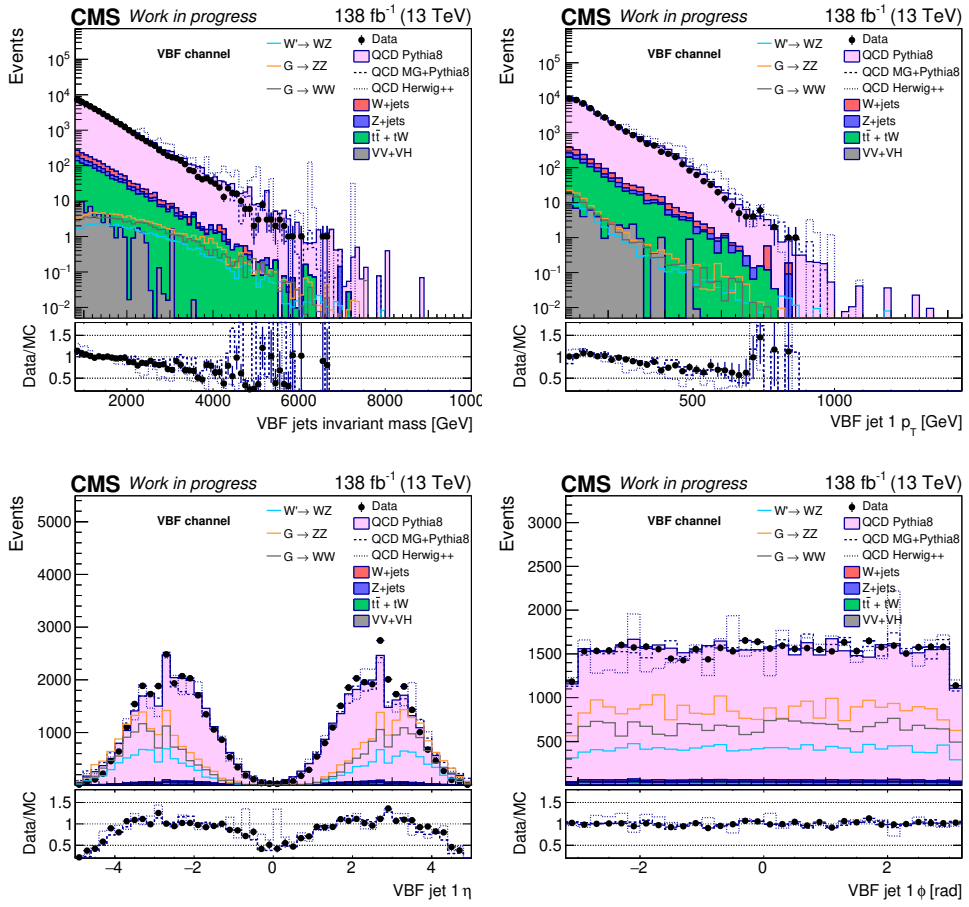


FIGURE 6.17: Distributions of the invariant mass of the two forward AK4 PUPPI jets (top left), first leading jet p_T (top right), η (bottom left), and ϕ (bottom right) of the two forward AK4 PUPPI jets after VBF preselections are applied. For the QCD multijet simulation, several alternative predictions are shown. The simulation is rescaled as described in the text. The ratio plots show the fraction of data over QCD multijet simulation for PYTHIA (black marker), MADGRAPH5_aMC@NLO interfaced with PYTHIA (dashed line) and HERWIG++ (dotted line).

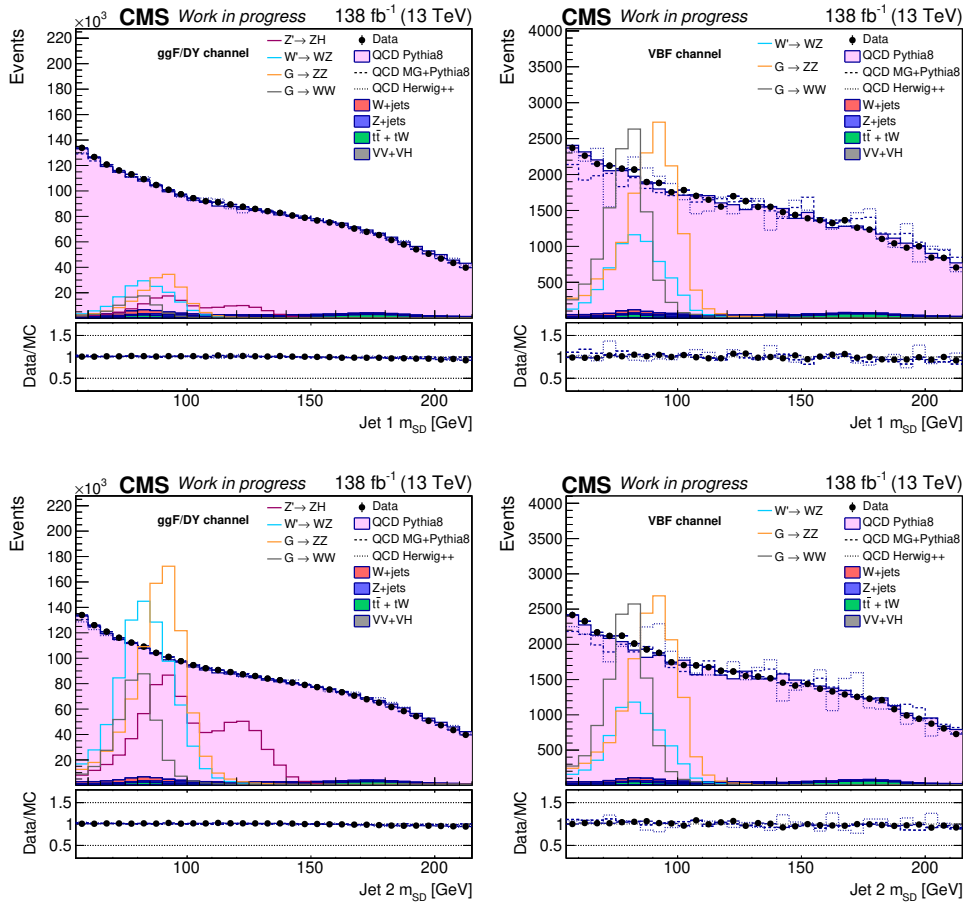


FIGURE 6.18: Softdrop mass of the two selected AK8 PUPPI jets in the event after the ggF/DY (left) and VBF (right) preselections are applied. For the QCD multijet simulation, several alternative predictions are shown. The simulation is rescaled as described in the text. The ratio plots show the fraction of data over QCD multijet simulation for PYTHIA (black marker), MADGRAPH5_aMC@NLO interfaced with PYTHIA (dashed line) and HERWIG++ (dotted line).

6.2.5 Jet substructure and tagging observables

In Fig. 6.18 the softdrop mass distribution of the two selected AK8 PUPPI jets in the event after the ggF/DY (left) and VBF (right) preselections are applied. The signal softdrop mass distribution peaks nicely around the boson mass, while the multijets background spectrum is peaked at lower softdrop masses. A good agreement between data and simulation is observed.

Figure 6.19 shows the scores of the DeepAK8-MD WvsQCD (top) and DeepAK8-MD ZHbbvsQCD (bottom) tagger, respectively. The PYTHIA and MADGRAPH5_aMC@NLO interfaced with PYTHIA simulations show a reasonable agreement to the data, while HERWIG++ better reproduces the data spectra. However, these variables are not used as observables in the fitting procedure.

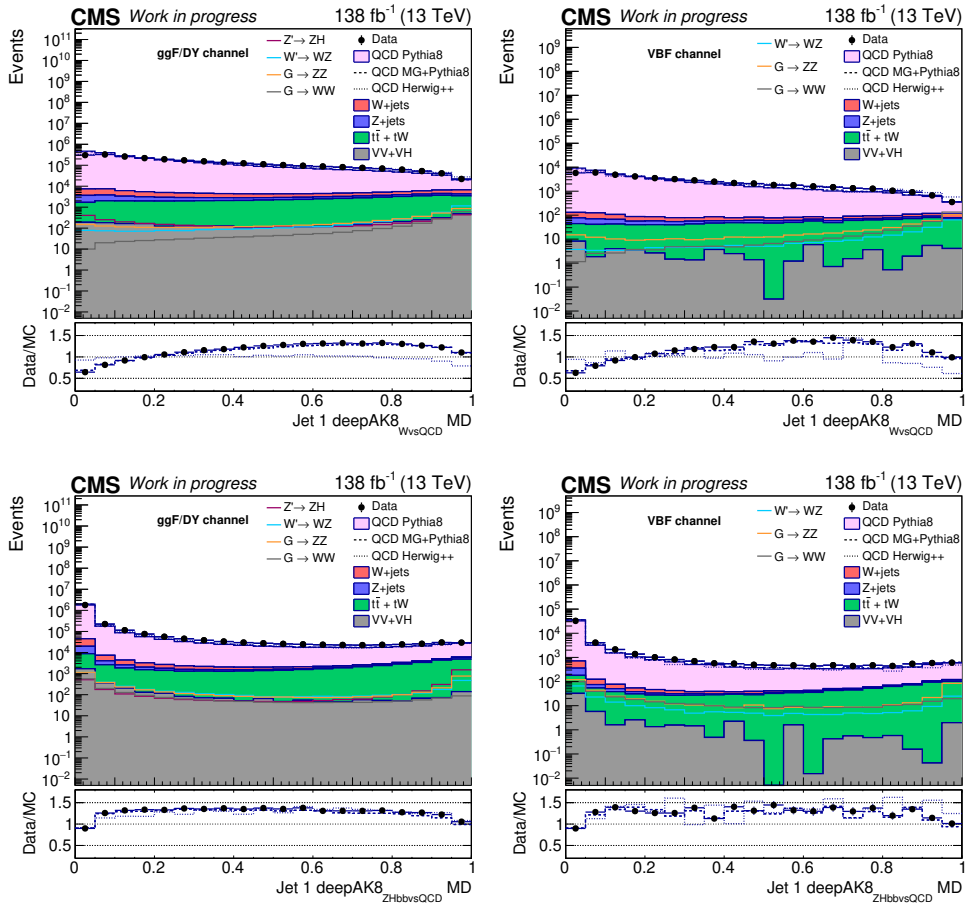


FIGURE 6.19: Distributions of the mass-decorrelated DeepAK8 WvsQCD (top) and ZHbbvsQCD (bottom) score of one of the two leading-p_T selected AK8 PUPPI jets in the event after the ggF/DY (left) and VBF (right) preselections are applied. For the QCD multijet simulation, several alternative predictions are shown. The simulation is rescaled as described in the text. The ratio plots show the fraction of data over QCD multijet simulation for PYTHIA (black marker), MADGRAPH5_aMC@NLO interfaced with PYTHIA (dashed line) and HERWIG++ (dotted line). For the QCD multijet simulation, several alternative predictions are shown. The simulation is rescaled as described in the text. The ratio plots show the fraction of data over QCD multijet simulation for PYTHIA (black marker), MADGRAPH5_aMC@NLO interfaced with PYTHIA (dashed line) and HERWIG++ (dotted line).

6.3 Jet substructure selections and event categorization

Jet substructure selections take advantage of the different inner characteristics of signal and background jets to increase the first and reduce the latter. In particular, for this analysis, the selections are used to identify boosted W, Z and Higgs jet candidates. The soft drop algorithm introduced in Sec. 5.3.1 is applied to the selected AK8 PUPPI jets. The m_{SD} distribution was shown in Fig. 6.18. The softdrop-jet mass is required to be between 55 and 215 GeV, to include the W/Z/H/top mass while excluding QCD multijet events with softer m_{SD} spectrum.

The previous CMS search in this channel [9, 177, 178] employed a decorrelated version of the N-subjettiness ratio τ_{21} (τ_{21}^{DDT}) to further enhance the analysis sensitivity to final states with boosted W/Z jets. While repeating the same jet tagging strategy followed in the past provides a solid starting point, it was shown in Sec. 5.3.4 that the CMS Collaboration developed and commissioned the novel and more performant DeepAK8 [165] jet tagging algorithm based on machine learning methods. The DeepAK8 is a multiclass classifier able to discriminate a variety of boosted signal jets (W, Z, H, top quarks) against QCD jets. Furthermore, it has been demonstrated (Sec. 5.3.4.1) that applying the 2D-decorrelation method to the mass-decorrelated version of the DeepAK8 tagger (DeepAK8-MD) prevents the sculpting of the m_{SD} and m_{jj} distributions. Hence, in the analysis presented in this thesis, the performance of the 2D decorrelated DeepAK8-MD (referred as 2D- or "map-decorrelated" DeepAK8-MD) has been validated in terms of decorrelation of the relevant quantities and discriminating power. In particular, the WvsQCD (defined in Eq. 5.6) and the ZHbbvsQCD (defined in Eq. 5.5) have been adopted to identify W/Z bosons decaying to two light quarks and Z/H bosons decaying to a pair of b quarks, respectively. For the WvsQCD (ZHbbvsQCD) 5%, 10% and 20% (2%, 5% and 10%) QCD misidentification rates are considered. The maps used in the working points optimization process (i.e. corresponding to the different misidentification rates) are shown in Fig. 6.20 for the WvsQCD (top) and the ZHbbvsQCD (bottom) taggers, respectively.

6.3.1 Optimization of the event categories

Two main categories are introduced:

VV: enriched in WW, WZ and ZZ final states

VH: enriched in WH and ZH final states.

It has to be noted that the VH category, to maximize the efficiency of the $Z(\mathrm{b\bar{b}})H(\mathrm{b\bar{b}})$ final state, may also pick up events from HH resonances although not optimized for that and hence not as sensitive as the dedicated search. The WvsQCD tagger is used to identify W/Z bosons decaying to two light quarks (Vqq-tag or V-tag in the following) and the ZHbbvsQCD tagger is employed for the identification of Z/H bosons decaying to a pair of b quarks (ZHbb-tag or H-tag in the following).

To find suitable working points, the approach of defining “high-purity” (HP) and “low purity” (LP) categories, similar to those applied in Refs. [9, 177, 178], has been adopted. The high-purity category maximizes the expected signal significance, while low-purity categories pick up a larger number of background events, allowing to constrain the background uncertainties. Moreover, LP categories still have a significant signal efficiency, making them the most sensitive at the highest resonances

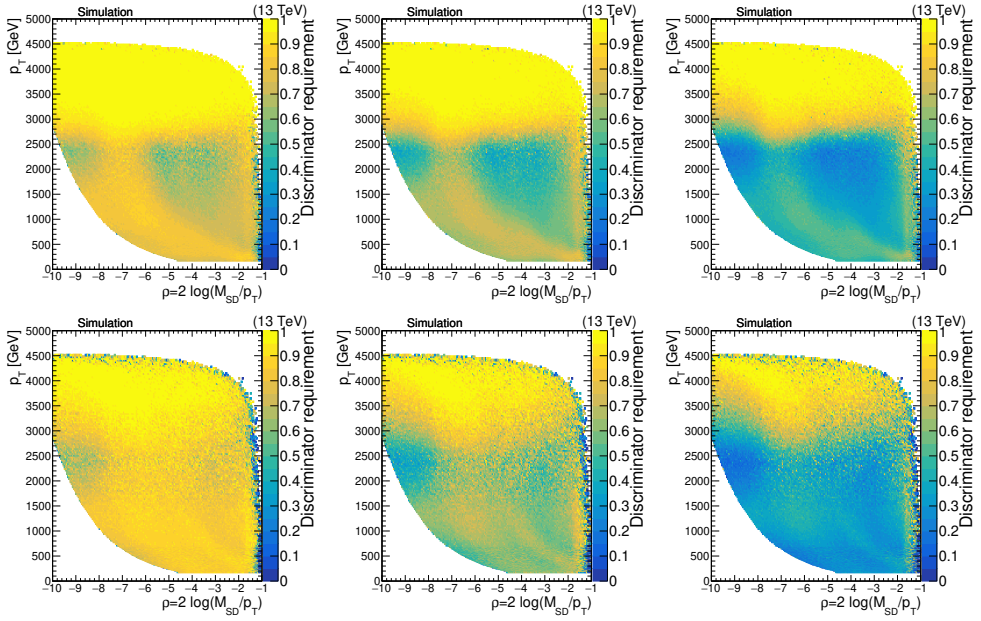


FIGURE 6.20: Decorrelation maps derived with QCD PYTHIA8 for DeepAK8-MD WvsQCD (top) tagger for the 5%, 10% and 20% QCD misidentification rates ZHbbvsQCD (bottom) tagger for the 2%, 5% and 10% QCD misidentification rates considered in the analysis optimization.

masses, where the background is low. Each of the VV and VH categories is then divided in sub-categories based on the number and type of HP/LP tagging. The five tagging categories are illustrated in Table 6.6 and can be summarized as follows:

- **VH HPHP**: 1 Vqq-tagged jet (HP) + 1 ZHbb-tagged jet (HP) or 2 ZHbb-tagged jets (HP)
- **VH LPHP**: 1 Vqq-tagged jet (LP) + 1 ZHbb-tagged jet (HP) or 1 ZHbb-tagged jet (LP) + 1 ZHbb-tagged jet (HP)
- **VH HPLP**: 1 Vqq-tagged jet (HP) + 1 ZHbb-tagged jet (LP)
- **VV HPHP**: 2 Vqq-tagged jets (HP)
- **VV HPLP**: 1 Vqq-tagged jet (HP) + 1 Vqq-tagged jet (LP).

Since the V and H jet taggers use different classifiers, they do not necessarily lead to orthogonality between the VH and VV categories, not allowing a straightforward combination. In previous analyses, either only one jet tagging algorithm was used to define the categories (as τ_{21}^{DDT} in Refs. [9, 177, 178], where H jet tagging is not considered), or disjoint jet mass windows were applied to distinguish a V tagged from a H tagged jet (as in the latest all hadronic VH results from the CMS Collabooration [185]). Hence, in this analysis a prioritization of the categories is needed to ensure orthogonality. Two different prioritization schemes have been studied. In the first scheme (scheme 1) all the VH categories are prioritized over the VV categories, i.e. an event falls in one of the two orthogonal VV categories only if it did not fall in one of the VH orthogonal categories. In the second scheme (scheme 2), the higher purity categories with two HP-tagged jets are first checked with priority on the VH ones over the VV ones, as follows:

$$\text{VH HPHP} \rightarrow \text{VV HPHP} \rightarrow \text{VH LPHP} \rightarrow \text{VH HPLP} \rightarrow \text{VV HPLP} \qquad (6.1)$$

TABLE 6.6: Two VV and three VH categories, for a total of five tagging categories, are introduced based on the number and type of high purity tagging.

$jet_1 \backslash jet_2$	V(qq) HP	V(qq) LP	ZH(bb) HP	ZH(bb) LP
V(qq) HP	• VV HPHP	▪ VV HPLP	✪ VH HPHP	◇ VH HPLP
V(qq) LP	▪ VV HPLP		❑ VH LPHP	
ZH(bb) HP	✪ VH HPHP	❑ VH LPHP	✪ VH HPHP	❑ VH LPHP
ZH(bb) LP	◇ VH HPLP		❑ VH LPHP	

In this order, an event falls in the next category if it did not fall in the previous one.

To optimize the choice of tagger working points and prioritization schemes, the $G_{bulk} \rightarrow WW$ and the $Z' \rightarrow ZH$ signals have been chosen as reference, having different final states and covering all the possible bosons in the decay. The optimization is carried out for the ggF/DY samples and applied also to the VBF ones.

The choice of map-decorrelated DeepAK8-MD discriminants working points has been performed by comparing their performance with the performance of τ_{21}^{DDT} and the DeepDoubleB [164] (which makes use of substructure, lifetime and vertex information to probe if a large radius jet contains two b-jets) in terms of both sensitivity and jet mass and p_T (or m_{jj}) decorrelation. The sensitivity has been estimated with the Punzi significance [186] defined as

$$s_P = \frac{\epsilon_s(t)}{\sqrt{a/2 + B(t)}}$$

where ϵ_s is the signal efficiency with respect to an applied tagging requirement t. The background rate B is evaluate in a 15% window around the signal mass and also depends on the tagging requirement t. In this test, the number of standard deviations of a Gaussian distribution corresponding to the chosen confidence level of the search (a) is set to 2. Figure 6.21 shows the Punzi's significance for three reference resonance mass points for the $G_{bulk} \rightarrow WW$ (top) and $Z' \rightarrow ZH$ (bottom) signals in the VV HPHP (left) and VH HPHP (right) categories.

Different markers and colors correspond to the different jet tagging algorithms, working points and category prioritization schemes described above. For the DeepAK8 algorithms two possible scenarios are investigated. In both cases 5% (10%) misidentification rate is chosen for the HP (LP) working point of the WvsQCD tagger and 10% for the LP working point of the ZHbbvsQCD tagger. For the HP working point of the ZHbbvsQCD tagger 2% (blue squares) and 5% (red circles) misidentification rates are considered. The combination of τ_{21}^{DDT} and DeepDoubleB is indicated by the green triangles. For the τ_{21}^{DDT} the same working points optimized in Refs. [9, 177, 178] are considered. The performance of the DeepDoubleB in comparison to the DeepAK8 algorithms was shown in Fig. 5.5. For this study the tight (T) and medium (M1) [164] working points are used for HP and LP tagging, respectively. Concerning the category prioritization, the first scheme (scheme1 and dashed lines in the distributions) always performs worse than the corresponding variant (scheme 2 and

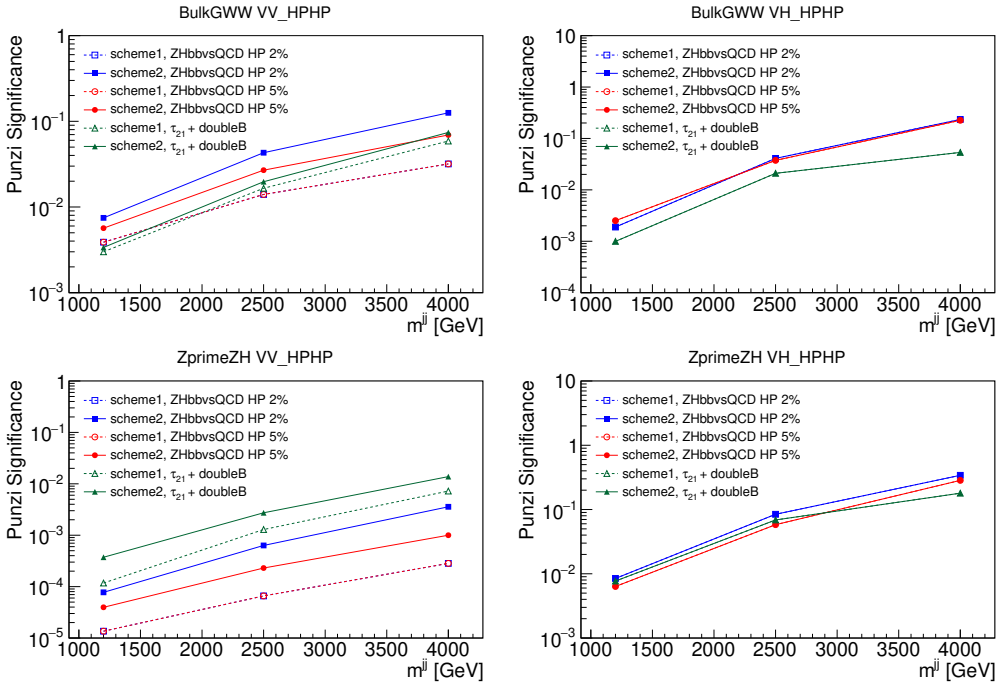


FIGURE 6.21: Punzi significance as a function of the dijet invariant mass, for $G_{bulk} \to WW$ (top) and $Z' \to ZH$ (bottom) in the VV HPHP (left) and VH HPHP (right) categories. Dashed lines and empty markers correspond to VH categories prioritized over the VV categories (scheme 1), while solid lines and filled markers refer to the second scheme described in Eq. 6.1. Markers of different colors and shape refer to different tagger choices, with the working points described in the text.

solid lines in the distributions) in the VV HPHP category (the most sensitive for VV signals), while it does not affect the sensitivity of the VH HPHP category. Thus, the second scheme is adopted in the following.

For the $G_{bulk} \to WW$, the DeepAK8 discriminants perform better than categorization using the combination of τ_{21}^{DDT} and DeepDoubleB in both the VV HPHP and VH HPHP categories. A tighter cut on the ZHbbvsQCD HP working point largely improves the sensitivity of the VV HPHP category without decreasing the sensitivity of the VH HPHP category. For the $Z' \to ZH$, the combination of τ_{21}^{DDT} and DeepDoubleB performs better than the DeepAK8 discriminants in the VV HPHP category while in the VH HPHP the DeepAK8 discriminants and a tighter cut on the ZHbbvsQCD HP is the one providing the best sensitivity. Since the Punzi's significance in the VV HPHP category for this signal is several order of magnitude smaller than the one in the VH HPHP, this latter category is the one driving the sensitivity to the $Z' \to ZH$ signal. Therefore the chosen taggers/working points are WvsQCD at 5% misidentification rate and ZHbbvsQCD at 2% misidentification rate for HP tagged jets. It has to be noted that this optimization step was performed with the map-decorrelated DeepAK8 (i.e. the non mass decorrelated version) as at this stage it was not yet known that such maps would results in sculpting of the data distributions, discussed in Sec. 5.3.4.1. However, the conclusions from the Punzi's significance optimization reported above are not affected by the choice of a different map-decorrelated tagger. In fact, the background rates that enter the denominator of the Punzi's significance are fixed by construction, as maps are derived for fixed background rates independently on which version of the DeepAK8 discriminant is used. At the same time it is true that the MD version of the DeepAK8 tagger is

less performant in terms of signal efficiencies. This was shown in Fig. 5.10 which compares the ROC curves for both MD and nominal DeepAK8 tagger together with the performance of the 2D decorrelation at the fixed misidentification rates of 2, 5, and 10% when using the nominal (red stars) and the MD (black dots) version of the DeepAK8 tagger to derive the map. This degradation in efficiency would only result in a global rescaling of the Punzi's significance towards lower values as the signal efficiency enters the numerator. Hence, the optimization studies have not been repeated and it was assumed that the same conclusions hold using maps derived from the DeepAK8-MD as well. However, it was found that the signal efficiency obtained using the deepAK8-MD tagger decreases more than for the nominal deepAK8 tagger for resonance masses above ≈5 TeV. To partially recover the drop in efficiency a looser misidentification rate of 20% has been introduced for the LP working point of the WvsQCD tagger. Once the full analysis procedure was in place and the tagging efficiency corrections became available, the performance in terms of sensitivity of the chosen working points has been validated comparing the expected limits obtained with the above mentioned combinations of working points.The final configuration of working points for the map-decorrelated DeepAK8-MD taggers used in the following is reported in Table 6.7.

TABLE 6.7: Optimized map-decorrelated DeepAK8-MD jet tagging working points defined at a fixed misidentification rate.

	HP	LP
V-tagging (WvsQCD)	5%	5–20%
H/Z-tagging (ZHbbvsQCD)	2%	2–10%

6.3.2 Event yield corrections for jet substructure selections

Jet tagging efficiency scale factors (SF) taking into account differences in tagging efficiencies between data and simulation for the map-decorrelated DeepAK8-MD taggers are estimated with the procedures described in Sec. 5.3.4.2. The applied SFs are summarized in Table 6.8 for the working points in Table 6.7. For WvsQCD the SFs are measured from jets with $400\,\text{GeV} < p_\text{T} < 800\,\text{GeV}$. The SFs for the HP working point in 2017/2018 are significantly lower due to the different underlying event tune used in simulation. Top mistagging SFs are reported in Table 6.9. The ZHbbvsQCD SFs are measured from jets with $p_\text{T} > 600\,\text{GeV}$.

TABLE 6.8: Scale factors for the 2D-decorrelated DeepAK8-MD W- and H-taggers.

WvsQCD	HP (5%)	LP (5–20%)
2016	$1.287^{+0.197}_{-0.147}$	$1.098^{+0.101}_{-0.098}$
2017	$0.823^{+0.059}_{-0.059}$	$1.185^{+0.094}_{-0.091}$
2018	$0.757^{+0.057}_{-0.052}$	$1.064^{+0.073}_{-0.072}$
ZHbbvsQCD	HP(2%)	LP (2–10%)
2016	$1.049^{+0.038}_{-0.040}$	$0.981^{+0.023}_{-0.026}$
2017	$0.998^{+0.035}_{-0.035}$	$1.066^{+0.027}_{-0.046}$
2018	$1.079^{+0.099}_{-0.092}$	$1.053^{+0.063}_{-0.069}$

Given the mixture of $V \to \text{q}\bar{\text{q}}$ and $Z/H \to \text{b}\bar{\text{b}}$ tagging discriminants used in the categorization, the tagging SFs are applied on an event by event basis keeping into account both the MC-truth of the jet and how it has been tagged. A reconstructed AK8 jet matched to a generator level $\text{V} \to \text{q}\bar{\text{q}}$ boson may still be tagged as a $\text{Z} \to \text{b}\bar{\text{b}}$

TABLE 6.9: Scale factors for misidentification of merged top quark jets when applying the 2D-decorrelated DeepAK8-MD W-tagger.

WvsQCD	HP (5 %)	LP (5-20 %)
2016	$0.949^{+0.065}_{-0.065}$	$0.990^{+0.042}_{-0.041}$
2017	$0.990^{+0.033}_{-0.032}$	$1.050^{+0.041}_{-0.041}$
2018	$0.919^{+0.029}_{-0.029}$	$1.050^{+0.034}_{-0.033}$

or a H $\rightarrow b\bar{b}$ boson, or viceversa a reconstructed AK8 jet matched to a generator level Z $\rightarrow b\bar{b}$ or a H $\rightarrow b\bar{b}$ can be tagged as a V $\rightarrow q\bar{q}$ boson, because of the different algorithms used in the analysis. A SF is applied to all AK8 jets which match to a V $\rightarrow q\bar{q}$, Z $\rightarrow b\bar{b}$ or H $\rightarrow b\bar{b}$. If the jet is tagged with the ZHbbvsQCD, the SFs of the ZHbbvsQCD tagger are applied, while if the jet is tagged with the WvsQCD the SFs of the WvsQCD tagger are applied. For AK8 jets originated from the decay of a top, if the reconstructed jet is matched to a V $\rightarrow q\bar{q}$ but not to a generator level top, SFs for either the WvsQCD or the ZHbbvsQCD are also applied based on which tagger discriminant they pass. SFs to take into account top quarks misidentified as W bosons are applied to reconstructed jets matched to a generator level t $\rightarrow$ bq$\bar{q}$.

7 | Signal and background modelling

In the multi-dimensional fit approach described in this thesis, signal and background shapes are parametrized in three dimensions, namely the m_{SD} of the leading-p_{T} large radius jets (m_{jet1} and m_{jet2}) and their dijet invariant mass m_{jj}.

The three-dimensional approach takes advantage of the fact that the signals are resonant in all three dimensions of the m_{jet1}-m_{jet2}-m_{jj} plane.

The probability density functions (pdfs) for signal and background processes used in the fit to data are derived from full simulation, considering also the detector conditions. Thus, a separated sample is employed for each of the three years to take into account the different data taking conditions. While the detector components and conditions varied across the 3 years of data taking, the detector performance relevant to this analysis was very similar. The samples are then added together with weights corresponding to the relative luminosity of the year. The modelling procedures benefit from the obtained large sample.

Since the two leading-p_{T} jets in the event are randomly labelled, all distributions for the first and second leading-p_{T} jet are the same except for statistical fluctuations.

The chapter describes pdfs derivation for signal and background simulation and the final fit procedure to the data. It is structured in the following way: first, the signal model is described in Sec. 7.1; then, the modelling of the different background contributions is illustrated in Sec. 7.2; in Sec. 7.3 the systematic uncertainties are discussed; finally, the maximum likelihood fit method and the fit validation are explained in Sec. 7.4 and Sec. 7.5, respectively. The results are presented and discussed in Chapter 8.

The work presented in this chapter has been developed by the author of this thesis under the supervision of Dr. Andreas Hinzmann, in collaboration with Dr. Clemens Lange and Dr. Jennifer Ngadiuba. Dr. Daniela Schäfer contributed to the modelling of the VH signals and Dr. Thea K. Årrestad to the modelling of the $t\bar{t}$ jet mass distribution. The modelling of the $t\bar{t}$ background was developed in collaboration with Dr. Daniela Schäfer and Dr. Jennifer Ngadiuba. Dr. Jennifer Ngadiuba assisted in the modelling of the dijet mass distribution of the partially resonant backgrounds and carried out the estimation of part of the systematic uncertainties affecting the signal. The signal PDF uncertainties were estimated by Dr. Gerrit van Onsem, who performed also the goodness-of-fit and bias tests. The author of this thesis contributed to all of the above and integrated the various contribution in the main analysis framework, carried out the combination of the datasets, performed all the stages of the signal and background modelling, including the implementation of the systematic uncertainties, and validated the three-dimensional fitting procedure.

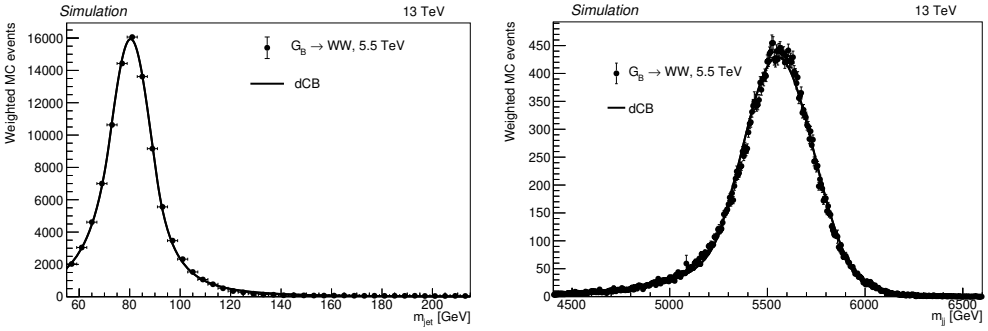


FIGURE 7.1: Jet mass (left) and dijet mass (right) distributions for a $G_{bulk} \rightarrow WW$ with M_X= 5.5 TeV. The fit with a double sided Crystal Ball (dCB) function is also shown.

7.1 Signal modelling

In order to use the new multi-dimensional fit method, the signal has to be parametrized in three dimensions. Two slightly different approaches are used for resonances decaying to WW/ZZ and decay channels with two different bosons. The modelling of the WW/ZZ signals uses the same approach described in Ref. [9, 177, 178]. Because of the simple confined peak structure of the signal in all three dimensions, in the WW/ZZ case, the signal pdf is defined as a product of the shape of the resonance mass and the jet masses:

$$P_{\text{sig,VV}}(m_{\text{jj}}, m_{\text{jet1}}, m_{\text{jet2}}|\overline{\theta}^{\text{S}}(\text{M}_\text{X})) = P_{\text{VV}}(m_{\text{jj}}|\overline{\theta}^{\text{S}}_1(\text{M}_\text{X}))\, P_{\text{j1}}(m_{\text{jet1}}|\overline{\theta}^{\text{S}}_2(\text{M}_\text{X}))\, P_{\text{j2}}(m_{\text{jet2}}|\overline{\theta}^{\text{S}}_3(\text{M}_\text{X})). \tag{7.1}$$

Here, the shapes for m_{jj}, m_{jet1} and m_{jet2} denoted as P_{VV}, P_{j1} and P_{j2}, respectively, are parametrized with double-sided Crystal Ball (dCB) functions [187] for each signal mass M_X. The dCB function has a Gaussian core (described by the mean and width/sigma parameters) with power-tails on each side (each described by two parameters commonly referred as N and α). The parameters $\overline{\theta}^{\text{S}} = (\overline{\theta}^{\text{S}}_1, \overline{\theta}^{\text{S}}_2, \overline{\theta}^{\text{S}}_3)$ denote the free parameters of each of the dCB functions. For the VV channel, given the resonant and consequently localized distributions of the signal processes in all considered mass dimensions, the m_{jj} and m_{jet} distributions can be considered as uncorrelated and these distributions can therefore be fitted separately from each other. The jet mass distributions of jet_1 and jet_2 are combined, which is possible due to the use of the random sorting, to attain a high statistics sample and fit this combined distribution once – ensuring that $P_{\text{j1}} = P_{\text{j2}}$. An example fit for one signal mass point is shown in Fig. 7.1.

For the VH and WZ channels, correlations between the m_{jet} and m_{jj} distributions can no longer be completely neglected because of the larger spread of the m_{jet} distribution, where now contributions from different bosons must be taken into account. Instead of the ansatz given in Eq. 7.1, the following model is used for the VH channel:

$$\begin{aligned} &P_{\text{sig,VH}}(m_{\text{jj}}, m_{\text{jet1}}, m_{\text{jet2}}|\overline{\theta}^{\text{S}}(\text{M}_\text{X})) = \\ 0.5 \cdot &\left(P_{\text{VH}}(m_{\text{jj}}|\overline{\theta}^{\text{S}}_1(\text{M}_\text{X}), m_{\text{jet1}}, m_{\text{jet2}})\, P_{\text{H}}(m_{\text{jet1}}|\overline{\theta}^{\text{S}}_2(\text{M}_\text{X}))\, P_{\text{V}}(m_{\text{jet2}}|\overline{\theta}^{\text{S}}_3(\text{M}_\text{X}))\right) \\ +0.5 \cdot &\left(P_{\text{VH}}(m_{\text{jj}}|\overline{\theta}^{\text{S}}_1(\text{M}_\text{X}), m_{\text{jet1}}, m_{\text{jet2}})\, P_{\text{V}}(m_{\text{jet1}}|\overline{\theta}^{\text{S}}_3(\text{M}_\text{X}))\, P_{\text{H}}(m_{\text{jet2}}|\overline{\theta}^{\text{S}}_2(\text{M}_\text{X}))\right). \end{aligned} \tag{7.2}$$

The pdf for the WZ channel is obtained with the substitution V $\rightarrow$ W and H $\rightarrow$ Z in Eq. 7.2. In this case, the probability distribution of the dijet invariant mass $P_{\mathrm{VH}}(m_{\mathrm{jj}}|\vec{\theta}_1^{\mathrm{S}}(\mathrm{M_X}), m_{\mathrm{jet1}}, m_{\mathrm{jet2}})$ gets a dependence on the two jet masses, due to the larger jet mass spread of VH and WZ signals. This is modelled by scaling the mean and width of the dCB fitted to m_{jj} depending on the two jet masses m_{jet1} and m_{jet2}. The scaling factors are derived by fitting the m_{jj} spectrum with a dCB function in four bins of the jet masses, and then fitting with a two dimensional polynomial the relative difference in mean and width with respect to the m_{jj} distribution obtained when one jet has a mass confined in a small window around the V boson and the other jet has a mass confined in a small window around the H boson. The same mass window is used for both jets in the WZ case. This shift in mean and width of the signal m_{jj} with the two jet masses has only a small dependence on the resonance mass $\mathrm{M_X}$, and therefore it is calculated by averaging all simulated signal mass points to obtain a larger sample. The 3D signal modelling is shown in Fig. 7.2 for ggF/DY produced $\mathrm{W'} \rightarrow \mathrm{WH}$, $\mathrm{Z'} \rightarrow \mathrm{ZH}$ and $\mathrm{W'} \rightarrow \mathrm{WZ}$ signals. The respective VBF signals are shown in Fig. B.1 in Appendix B.1. The jet masses are modelled separately for the two bosons by fitting the distributions of AK8 jets matched to generator level V or H bosons with a dCB centered at the corresponding boson mass to obtain the pdfs P_{H} and P_{V} in Eq. 7.2. An example fit for one signal mass point is shown in Fig. 7.3. The final signal pdf is then the sum of two contributions: one where the first jet is an H boson and the second a V boson, and one where this is reversed. Both contributions are added with a factor of 0.5 due to the random labelling of jets in the event.

It has been verified with signal simulation that the signal shapes in all mass dimensions are affected only marginally by applying tagging requirements. Therefore, the shapes for all categories are derived once, after the preselections for AK8 jets described in Sec. 6.2 are applied, and are subsequently used in all categories. This approach is beneficial since some tagging categories contain a very small amount of simulated signal events, as explained in the following, which makes a stable fit challenging.

In order to parametrize signals with arbitrary resonance masses $\mathrm{M_X}$, the parameters of the dCB functions are interpolated as a function of $\mathrm{M_X}$. The interpolation is obtained with a spline, linear or parabolic function to ensure smooth and stable shapes for each interpolated variable at each mass point. Figure 7.4 shows the mean (left) and sigma (right) of the signal m_{jj} (top) and m_{jet} (bottom) distributions as a function of $\mathrm{M_X}$, as they are obtained from the mean and width of the dCB function fitted to the spectrum of ggF/DY produced resonances. The other dCB fit parameters and parameters for VBF produced signals can be found in Appendix B.1.

The jet mass mean is stable as a function of the resonance mass, while the width increases with $\mathrm{M_X}$. In general the jet mass width of W bosons is around 7–8 GeV, the jet mass distribution of a Z boson shows widths around 9–10 GeV and the Higgs boson has a larger width of $\approx$ 12–14 GeV. The WZ final state has a larger width compared to WW or ZZ as the W and Z peaks cannot be resolved. The measured width difference between W and Z bosons, relative to their masses, is due to the higher production rate of B-mesons in Z decays that can decay semileptonically, eventually producing neutrinos, resulting in a larger tail of the distribution to low jet masses and a consequent decrease in resolution.

The signal shapes resulting from the interpolation between the sampled mass points are shown in Fig. 7.5 for the m_{jj} (left), and the m_{jet1} (right) distributions. The distribution for m_{jet2} is omitted since it is identical to m_{jet1}, because of the random jet labelling.

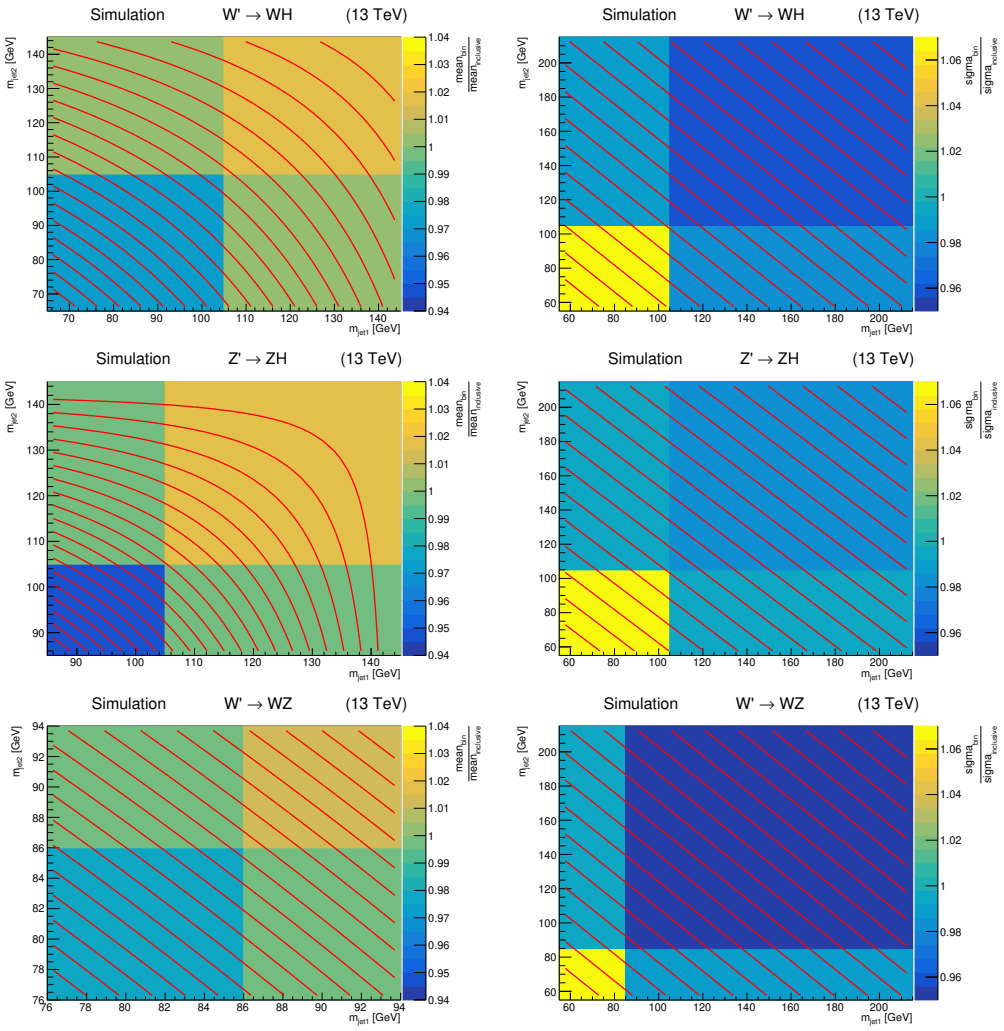


FIGURE 7.2: Two-dimensional dependence of the mean (left) and width (right) of the dijet invariant mass distribution on the two jet masses for the $W' \rightarrow WH$ (top), $Z' \rightarrow ZH$ (middle), and $W' \rightarrow WZ$ signal (bottom). The Z-axis shows the relative difference in mean or width with respect to the m_{jj} distribution obtained with jet masses confined in windows around the relevant boson mass and in the case where no jet mass selections are applied.

Finally, for each mass point M_X and each jet tagging category, the signal efficiency, defined as the total signal yield after passing all analysis selections divided by the number of generated events, is also interpolated as a function of M_X in order to extract signal yields for arbitrary resonance masses. The interpolation is obtained with a polynomial function. Figure 7.6 (Fig. 7.7) shows on top the two (VBF) high purity categories, while on the bottom the total signal efficiency is presented. Overall, ggF/DY produced signals have a higher efficiency (12-26%) than VBF produced signals (2-18%). Signals with $b\bar{b}$ decays in the final states have higher signal efficiencies due to the prioritization of VH categories over VV ones. For such signals, the VH HPHP category contains 40-60% of the signal, while for the other signals 10-40% of the signal is contained in the VH HPHP category and 15-40% in the VV HPHP. The VH LPHP, VH HPLP, VV HPLP categories contain 5-40% of the signal depending on resonance type and mass. While a similar behaviour is observed also for VBF produced signals, here spin effects changing kinematic distributions used for the event selection are more relevant and spin-0 (spin-1) signals have the highest

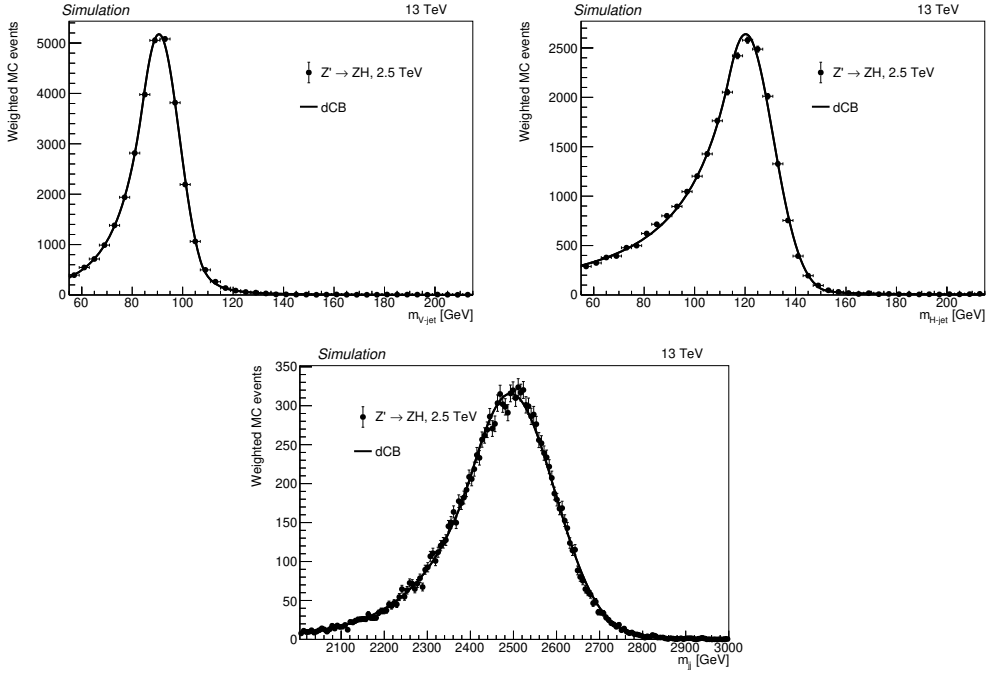


FIGURE 7.3: Jet mass (top) and dijet mass (bottom) distributions for a $Z' \rightarrow ZH$ with M_X = 2.5 TeV. The fit with a double sided Crystal Ball (dCB) function is also shown. On the top left (right), the jet mass distrtibution for the jet matched to a generator level V (H) boson is presented.

(lowest) efficiency. The VBF categories contain 25-40% of the VBF signals and less than 5% of the ggF/DY signals.

Both signal shapes and efficiencies are obtained by adding the three year simulation campaigns with weights corresponding to the relative luminosity of the year.

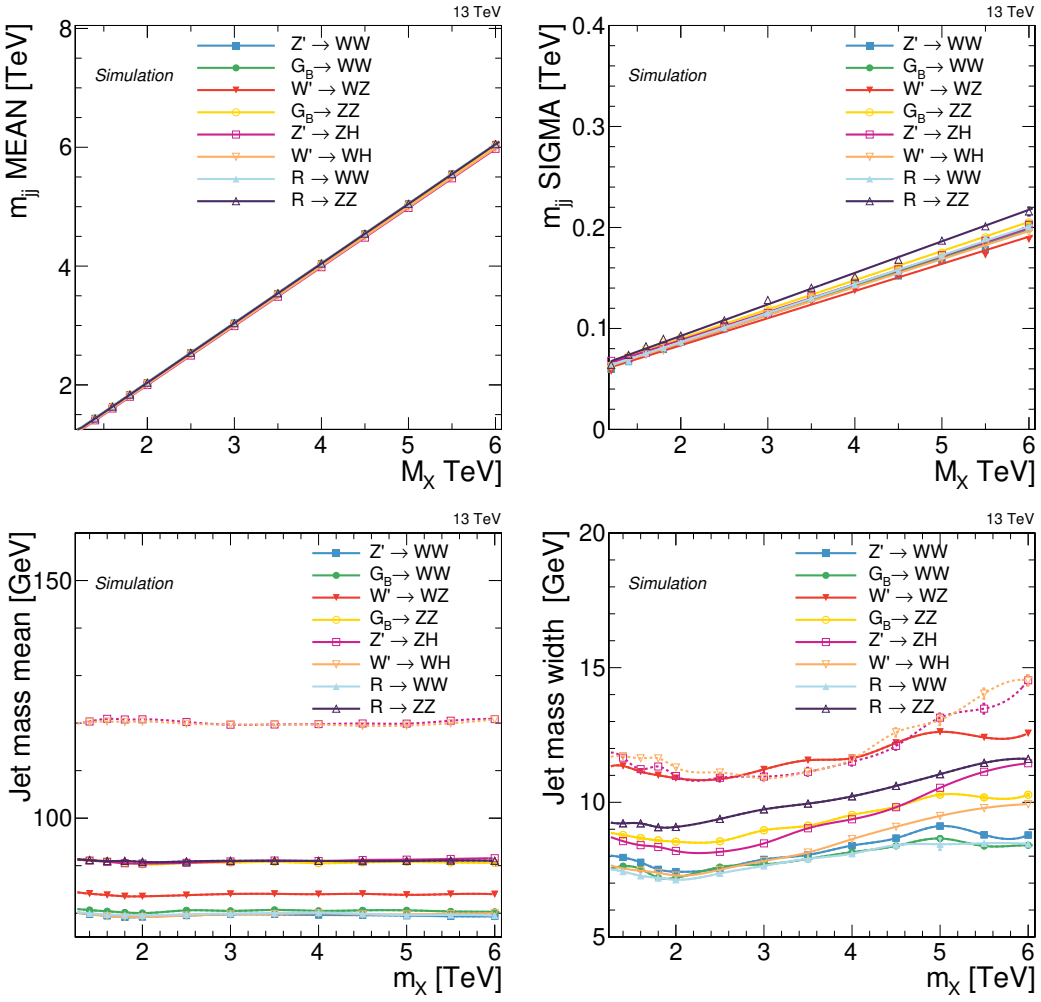


FIGURE 7.4: The mass scale (left) and resolution (right) of the signal m_{jj} (top) and m_{jet} (bottom) as a function of M_X, as they are obtained from the mean and width of the dCB function fitted to the spectrum of ggF/DY produced signals. For the m_{jj} a linear fit is performed while for m_{jet} a spline is used.

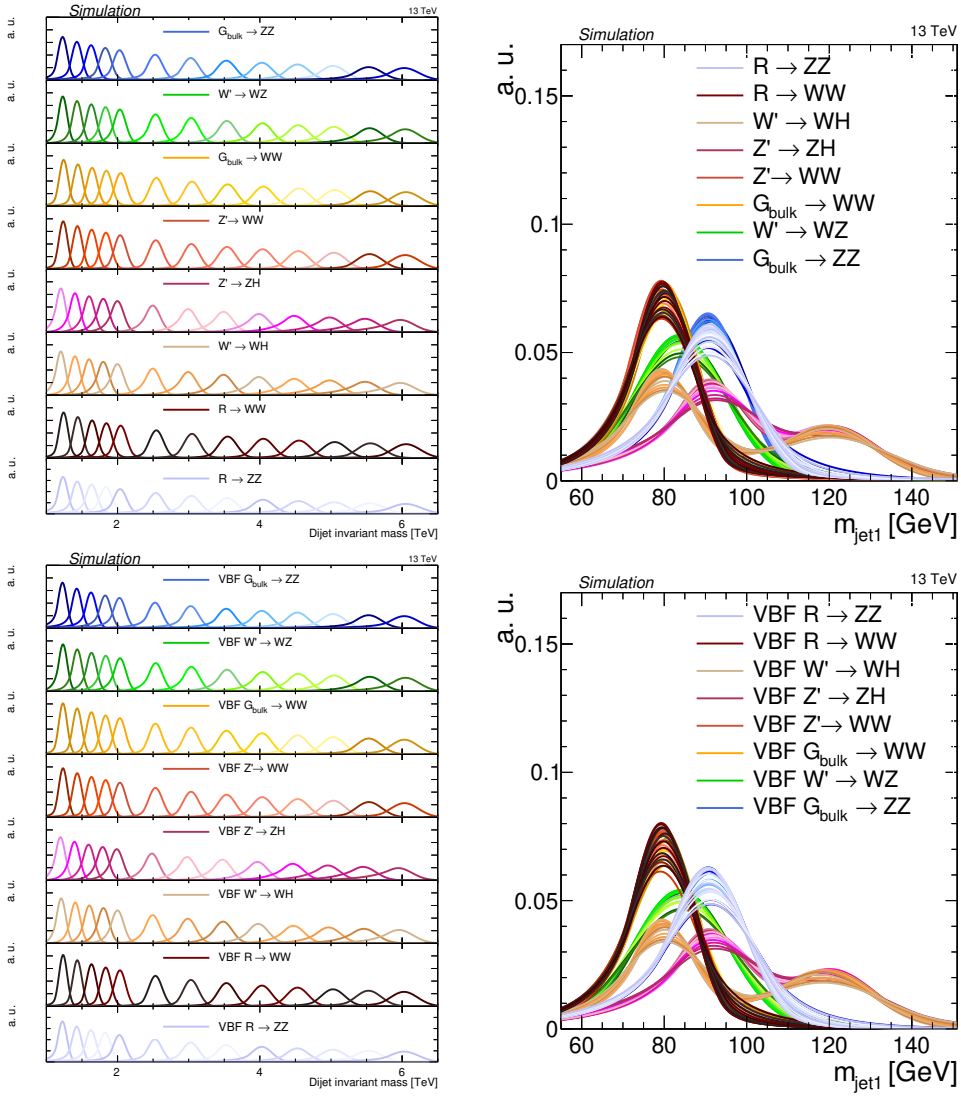


FIGURE 7.5: The final m_{jj} (left) and m_{jet1} (right) signal shapes extracted from the parametrization of the fitted dCB parameters. The same shapes are used for all tagging categories. The jet mass distributions are shown for a range of resonance masses between 1.3 and 6 TeV for one of the two jets in the event fitted without applying category selections. Since the jets are labeled randomly, the jet mass distributions for the second jet are the same and not shown here. The distributions for a G_{bulk} and Z' decaying to VV have the same shapes as those for the Radion signal and are therefore not visible. Different transparencies of a color represent different resonance masses as indicated on the left part.

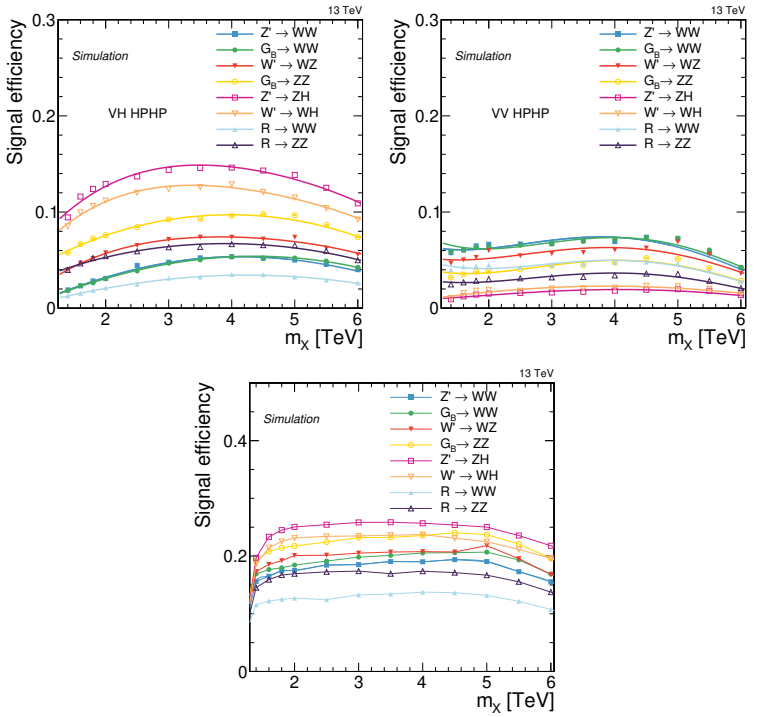


FIGURE 7.6: Signal efficiency as a function of M_X after all selection are applied, on the top left in the ggF/DY VH HPHP category and on the right in the ggF/DY VV HPHP category. The polynomial fit is also shown as a solid line. The bottom plot shows the total signal efficiency considering all 10 categories. Here the line has the only purpose to guide the eyes. The various ggF or DY produced signal models are shown.

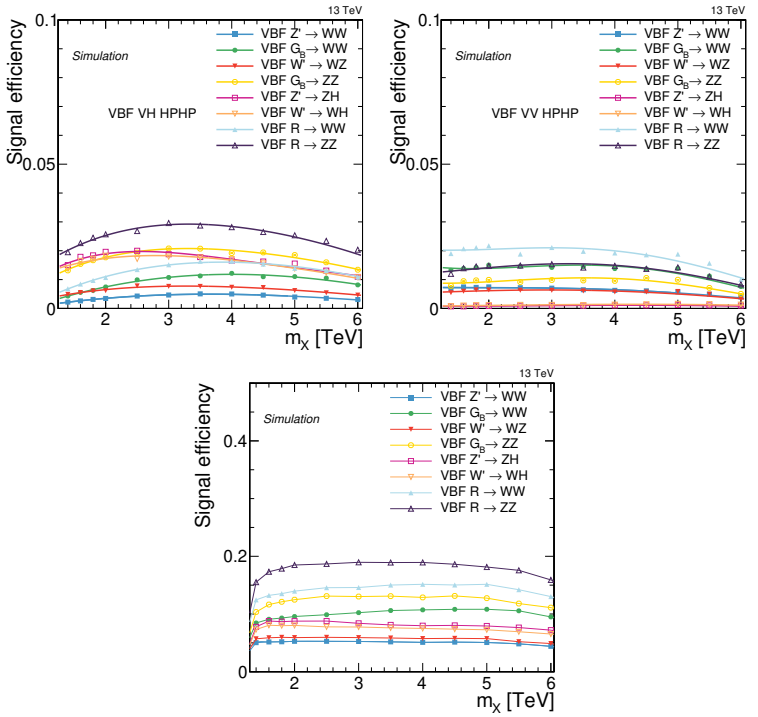


FIGURE 7.7: Signal efficiency as a function of M_X after all selection are applied, on the top left in the VBF VH HPHP category and on the right in the VBF VV HPHP category. The polynomial fit is also shown as a solid line. The bottom plot shows the total signal efficiency considering all 10 categories. Here the line has the only purpose to guide the eyes. The various VBF produced signal models are shown.

7.2 Background modelling

Standard Model backgrounds are also modelled in three dimensions. As introduced in Sec. 6.1.1.2, the contributions, from the most to the least abundant, are QCD multijet, top pair ($t\bar{t}$) and W/Z+jets production. Minor contributions come from single top and diboson production, included in the $t\bar{t}$ description in the following. Given the large number of categories and the different tagging requirements in each of them, the background composition in each tagging category is different, as shown in Table 7.1. Furthermore, the VBF categories have a limited amount of statistics. Therefore, background shapes are constrained in the ggF/DY region and for each VBF category the templates derived from the respective ggF/DY tagging category are used. The ability of shapes derived in the ggF/DY region to reproduce the distributions in the VBF category has been checked in simulation.

TABLE 7.1: Expected background yields in each category from simulation before the fit to data is performed. The QCD yield has been rescaled by a factor 1.8 to match the data.

Background	Total	QCD	$t\bar{t}$	W+jets	Z+jets
VH HPHP VBF	404	240	155	5	4
VV HPHP VBF	250	188	48	10	4
VH LPHP VBF	1073	872	182	9	10
VH HPLP VBF	729	583	131	10	5
VV HPLP VBF	1071	976	47	35	13
VH HPHP DY/gg	17732	9941	7181	343	267
VV HPHP DY/gg	11909	8661	2309	701	238
VH LPHP DY/gg	43408	33829	8406	534	639
VH HPLP DY/gg	32326	25208	5853	899	366
VV HPLP DY/gg	52525	47064	2367	2297	797

7.2.1 Modelling of the QCD multijet background

The QCD multijet background is estimated in a data driven way as the modelling strongly depends on which generator is chosen for the process simulation and has large uncertainties. However, simulation is used to derive QCD background shapes, including free parameters to allow the model to adapt to the data distribution in the final multidimensional fit.

The QCD multijet background presents a complex three-dimensional spectrum where the mass observables are correlated with each other. The use of three-dimensional binned template shapes takes into account these features. The same procedure described in Refs. [9, 177, 178] is used to ensure smooth shapes without empty bins. The following fit range and binning are used for the three axes: $m_{\text{jet1}}/m_{\text{jet2}}$ is fitted from 55 to 215 GeV using 2 GeV bins. m_{jj} is fitted from 1246 to 7600 GeV. The lower bound is chosen to avoid complications in the fitting procedure due to trigger turn-on effects, where sensitivity would also be affected by the reduced trigger efficiency (Sec. 6.2.1); the upper bound is chosen considering the highest dijet invariant mass found in the data, 5.6 TeV in the VV HPLP category. For m_{jj}, a binning corresponding to the actual dijet mass resolution of the order of ≈5% is implemented:
1246, 1313, 1383, 1455, 1530, 1607, 1687, 1770, 1856, 1945, 2037, 2132, 2231, 2332, 2438, 2546, 2659, 2775, 2895, 3019, 3147, 3279, 3416, 3558, 3704, 3854, 4010, 4171, 4337, 4509, 4686, 4869, 5058, 5253, 5455, 5663, 5877, 6099.

The conditional product

$$P^{\mathrm{QCD}}(m_{\mathrm{jj}}, m_{\mathrm{jet1}}, m_{\mathrm{jet2}}|\bar{\theta}^{\mathrm{QCD}}) = P_{\mathrm{jj}}(m_{\mathrm{jj}}|\theta_1^{\mathrm{QCD}}) \times P_{\mathrm{cond,1}}(m_{\mathrm{jet1}}|m_{\mathrm{jj}}, \theta_2^{\mathrm{QCD}}) \times P_{\mathrm{cond,2}}(m_{\mathrm{jet2}}|m_{\mathrm{jj}}, \theta_3^{\mathrm{QCD}}) \tag{7.3}$$

describes the three-dimensional m_{jj}-m_{jet1}-m_{jet2} plane, where $P_{\mathrm{cond,1}}$ and $P_{\mathrm{cond,2}}$ are two dimensional templates containing different jet mass shapes in intervals of m_{jj}. Due to the random jet sorting, the templates $P_{\mathrm{cond,1}}$ and $P_{\mathrm{cond,2}}$ are essentially the same, except for statistical fluctuations, and the choice to generate two templates is purely technical. P_{jj} is a one-dimensional template, describing the dijet invariant mass spectrum. The parameters $\bar{\theta}^{\mathrm{QCD}} = (\theta_1^{\mathrm{QCD}}, \theta_2^{\mathrm{QCD}}, \theta_3^{\mathrm{QCD}})$ are the free parameters of the background fit, included as nuisance parameters in the multi-dimensional fitting and limit setting procedure, as explained in Sec. 7.4.

The probability density in Eq. 7.3 requires the computation of the conditional two-dimensional shapes of m_{jet1} and m_{jet2} given m_{jj} and one-dimensional shape of the m_{jj} distribution. Similarly to VH signals, the correlations between m_{jet} and m_{jj} have to be modeled for the QCD multijet background.

The background model is built starting from simulation and encodes sufficient nuisance parameters into the fit to allow the shape to adapt itself to the data. A *forward-folding* approach ensures smooth shapes without empty bins, despite the large phase space and the limited number of available simulated events. In this method, instead of filling a histogram with reconstructed events, the relevant quantities of each event are calculated on generator level. The full jet clustering also happens on generator level particles. The generator quantities of each simulated event are then smeared with a Gaussian distribution (*kernel*), parametrizing the mass scale and resolution due to detector effects, resulting in smoothed histograms of reconstructed quantities. The mass scale and resolution are parametrized using the ratio of $m_{\mathrm{jet}}^{\mathrm{reco}}/m_{\mathrm{jet}}^{\mathrm{gen}}$ (and similarly $m_{\mathrm{jj}}^{\mathrm{reco}}/m_{\mathrm{jj}}^{\mathrm{gen}}$) as a function of the generated transverse momentum of the jet. Since a random labeling of the two jets in the event is used, only one such parametrization is derived. The final scale and resolution used in the smearing of each event is derived by fitting a Gaussian function to the histogram containing $m_{\mathrm{jet}}^{\mathrm{reco}}/m_{\mathrm{jet}}^{\mathrm{gen}}$ for different intervals of the generated jet p_{T}. Figure 7.8 shows the fit to $m_{\mathrm{jet}}^{\mathrm{reco}}/m_{\mathrm{jet}}^{\mathrm{gen}}$ (left) and $m_{\mathrm{jj}}^{\mathrm{reco}}/m_{\mathrm{jj}}^{\mathrm{gen}}$ (right) for an arbitrary bin. The Gaussian mean yields the mass scale and the Gaussian width the mass resolution.

The dijet mass and soft drop jet mass scale and resolution as a function of the generator-level jet p_{T} are shown in Fig. 7.9. The three different data-taking years and their combination (Run2) weighted by luminosity are shown. No particular differences are observed among years. Therefore, the m_{jj} and m_{jet} resolution parametrization and the QCD background *templates* are derived by adding the three year simulation campaigns with weights corresponding to the relative luminosity of the year.

For each MC event in the two (one)-dimensional m_{jj}-m_{jet} (m_{jj}) space, a 2D (1D) Gaussian kernel is built starting from generator level quantities. In the case of the m_{jj}-m_{jet} space, the derived mass scale and resolution are used to populate the conditional 2D histogram as follows. For each generated event i a 2D Gaussian kernel $K((m_{\mathrm{jet}}, m_{\mathrm{jj}}), m_{\mathrm{jet},i}^{\mathrm{gen}}, m_{\mathrm{jj},i}^{\mathrm{gen}}, p_{\mathrm{T},i}^{\mathrm{gen}})$ is calculated and used to fill a 2D histogram (*template*)

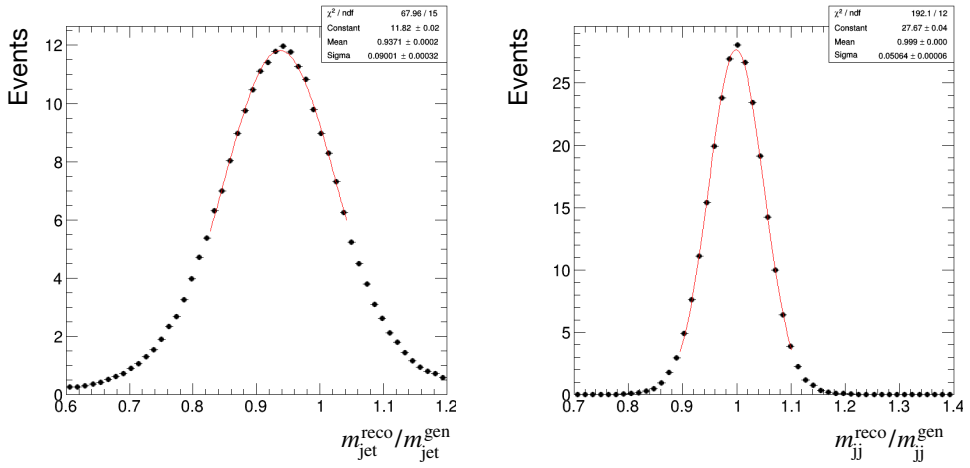


FIGURE 7.8: Fit to $m_{\text{jet}}^{\text{reco}}/m_{\text{jet}}^{\text{gen}}$ (left) and $m_{\text{jj}}^{\text{reco}}/m_{\text{jj}}^{\text{gen}}$ (right) with a Gaussian function in the 742–770 GeV and 636–660 GeV, respectively, generator-level jet p_{T} bin in the nominal PYTHIA8 sample. The mass resolution is taken as the width of the fitted Gaussian, while the Gaussian mean yields the mass scale.

at the $(m_{\text{jet}}, m_{\text{jj}})$ bin. The 2D Gaussian kernel is defined as

$$K((m_{\text{jet}}, m_{\text{jj}}), m_{\text{jet},i}^{\text{gen}}, m_{\text{jj},i}^{\text{gen}}, p_{\text{T},i}^{\text{gen}}) = \frac{\omega_i}{\sqrt{2\pi} R_{m_{\text{jj}}} \cdot R_{m_{\text{jet}}}} \cdot \exp\left(-\frac{1}{2}\left(\frac{m_{\text{jj}} - S_{m_{\text{jj}}} \cdot m_{\text{jj},i}^{\text{gen}}}{R_{m_{\text{jj}}} \cdot m_{\text{jj},i}^{\text{gen}}} \right)^2 - \frac{1}{2}\left(\frac{m_{\text{jet}} - S_{m_{\text{jet}}} \cdot m_{\text{jet},i}^{\text{gen}}}{R_{m_{\text{jet}}} \cdot m_{\text{jet},i}^{\text{gen}}} \right)^2 \right), \tag{7.4}$$

where $S_{m_{\text{jet}}} = S_{m_{\text{jet}}}(p_{\text{T},i}^{\text{gen}})$ and $R_{m_{\text{jet}}} = R_{m_{\text{jet}}}(p_{\text{T},i}^{\text{gen}})$ are the scale and the resolution derived in the previous step for the jet mass, $S_{m_{\text{jj}}} = S_{m_{\text{jj}}}(p_{\text{T},i}^{\text{gen}})$ and $R_{m_{\text{jj}}} = R_{m_{\text{jj}}}(p_{\text{T},i}^{\text{gen}})$ are the scale and the resolution for the dijet mass, and ω_i is the event weight. This procedure is performed separately for m_{jet1} and m_{jet2}.

To build the one-dimensional template for the dijet invariant mass the same procedure is performed with a one-dimensional Gaussian kernel depending only on m_{jj}.

As shown in Fig. 5.11, applying tagging requirements does not change the shape of the inclusive mass distributions. Therefore, the templates are derived from the inclusive distributions to benefit from all the available statistics. To remove residual bias due to generator events residing at the corners of the phase space, the inclusive templates constructed as above are fitted to the MC samples of a certain category. In this way, dedicated templates are derived for each category to be used as input to the final fit performed through vertical morphing, where alternative shapes are also included as shape systematic uncertainties, as described in Section 7.3. Example templates are shown in Fig. 7.10.

As a first validation of the templates generation, Fig. 7.11 shows a comparison of the three-dimensional pdf to simulated events projected on m_{jj}, m_{jet1} and m_{jet2}; the QCD PYTHIA8 samples, used for the nominal templates, are shown for the two high purity categories, VH HPHP (top) and VV HPHP (bottom). The other categories and distributions obtained from different MC generators to be used as alternative shapes in the final fit (see Section 7.3) can be found in Appendix B.2. Good agreement

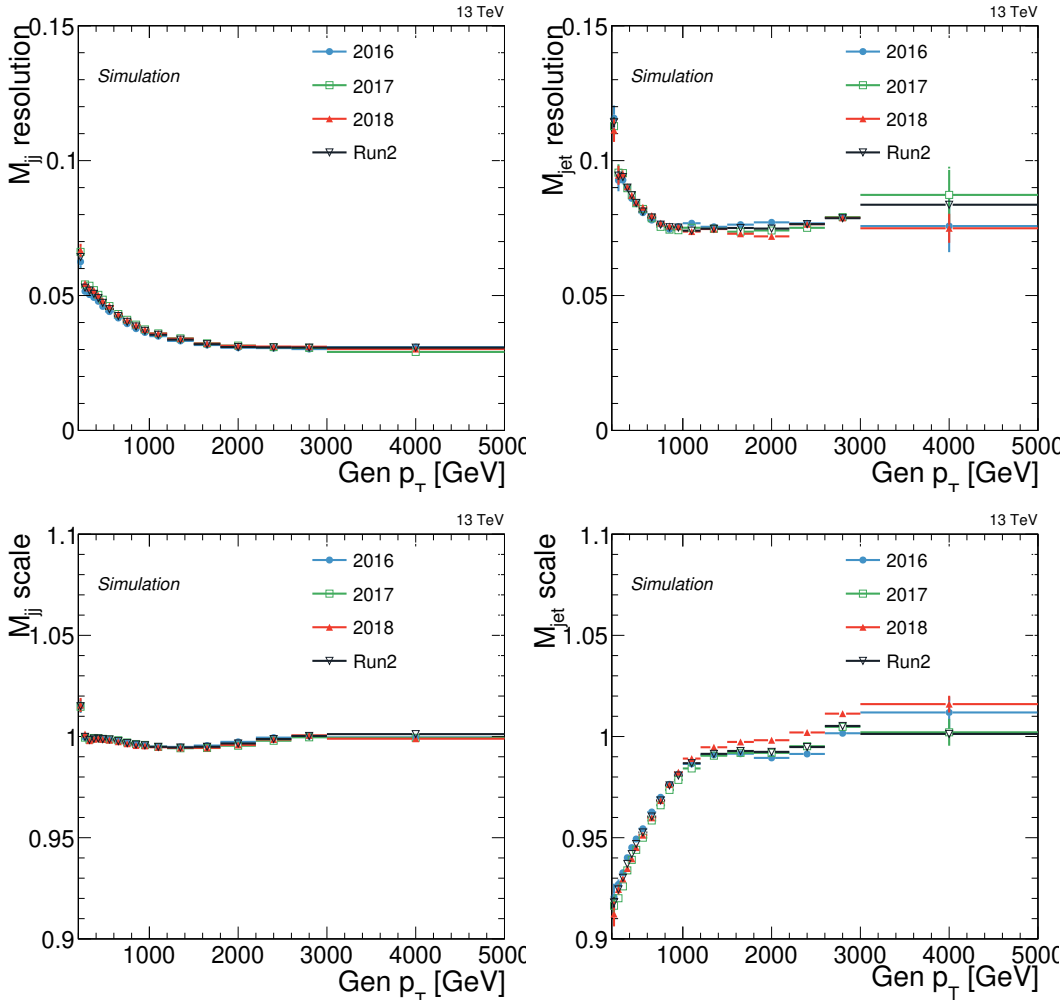


FIGURE 7.9: Resolution (top) and scale (bottom) for m_{jj} (left) and the m_{jet} (right) as a function of generator-level jet p_T derived from nominal PYTHIA8 sample. The three different data taking years as well as their combination weighted by luminosity are shown.

between templates derived from generator level quantities and the reconstructed event from simulation is observed. For each VBF category, the templates derived from the respective ggF/DY tagging category are used. The ggF/DY templates of a certain category fitted to the MC sample of the respective VBF category are shown in Sec. B.2.2 of Appendix B.2.

7.2.2 Modelling of the partially resonant backgrounds

In addition to the main background from QCD multijet processes, there are a few subdominant backgrounds to consider, namely top pair production ($t\bar{t}$) and W/Z+jets processes containing one real merged V jet and at least one additional jet, together called V+jets. The $t\bar{t}$ background can also contain jets from real merged vector bosons or real merged top quarks.

In this analysis the adoption of the DeepAK8 tagger and the addition of the Z/H→$b\bar{b}$ tagging categories changed the background composition significantly, compared to the previous version of this analysis focused on VV decay modes only [9, 177, 178]. In particular, the $t\bar{t}$ component, almost negligible before, now constitutes an overall 20% of the background (40% in the VH HPHP category). Therefore, while

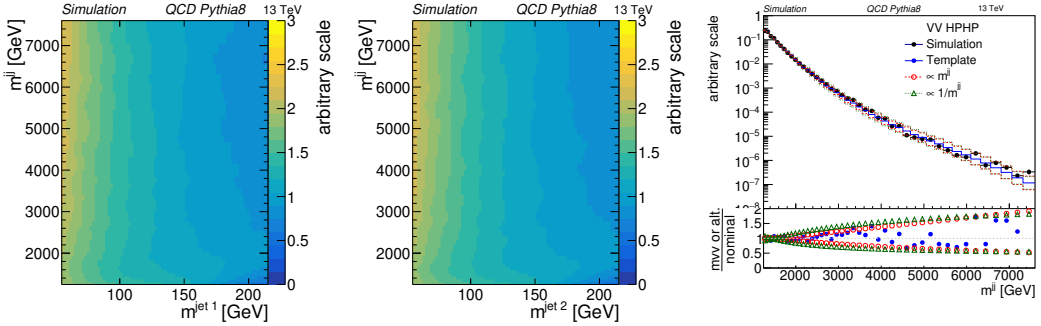


FIGURE 7.10: The left (middle) histogram shows the conditional m_{jet1} (m_{jet2}) pdf for a given interval in m_{jj} for the main QCD multijet background processes. Since the two jets in the event receive random labels these distributions are essentially the same barring effects from statistical fluctuations. On the right, the one-dimensional m_{jj} templates derived from the forward-folding procedure compared to reconstructed events taken from a simulated sample for the QCD multijet background. The green and red lines show alternative shapes derived by varying the m_{jj} spectrum proportional to m_{jj} or to $1/m_{\text{jj}}$.

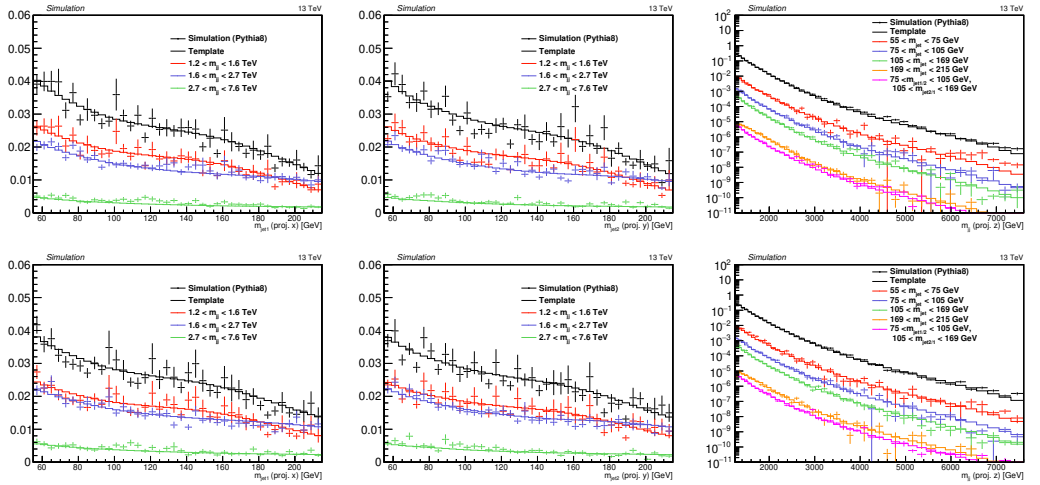


FIGURE 7.11: Comparison between simulation (markers) and templates derived from generator level quantities (lines) for the nominal QCD PYTHIA8 sample for the VH HPHP (top) and VV HPHP (bottom) category. The templates are shown for m_{jet1} (left), m_{jet2} (middle) and m_{jj} (right). In the m_{jj} distribution, the jet mass selections reported in the legend are applied to both jets. In the m_{jet1} and m_{jet2} distributions, the shown m_{jj} slices cover the whole phase space. This is not the case in the m_{jj} distribution where many more different combinations should be shown. However, slices covering the most relevant jet mass ranges are presented. Template and simulation distributions corresponding to the various m_{jj}, m_{jet1} and m_{jet2} intervals are rescaled for visualization purpose.

similar modelling as in Ref [9, 177, 178] has been used for V+jets, a new full separate modelling of $t\bar{t}$ is introduced.

The treatment of these processes is fundamentally different from that of the main QCD multijet background since these events contain a real merged vector boson or top quark jet, leading to a peak around the W/Z boson or top quark mass for the two jet mass dimensions. The dijet invariant mass has a monotonically falling distribution similar to QCD multijet processes.

For $t\bar{t}$, several contributions have to be considered as different jet mass and dijet invariant mass shapes are observed depending on which fractions of the jets have

merged top quarks, merged W bosons or not merged decay products. All the different contributions are separately modelled as described in the following. Single top and WW contributions to the SM background are modelled together with the different $t\bar{t}$ contributions described in the following.

7.2.2.1 $t\bar{t}$ modelling

The Standard Model $t\bar{t}$ background at high energies consists of several contributions, summarized in Fig. 6.8, resulting in different jet mass and dijet invariant mass shapes. Different $t\bar{t}$ contributions have different shapes and contribute in different amounts in the various categories, thus all the different contributions and tagging categories are modelled separately. The contributions are separated in simulation by matching the final state AK8 jets with generator level W bosons or top quarks. If the jet cannot be matched to a generator level W boson or top quark, it is labelled as "non-resonant".

The pdfs for each of the six contributions are built in the following way

$$
\begin{aligned}
P_{\mathrm{TT}}\left(m_{\mathrm{jet1}}, m_{\mathrm{jet2}}, m_{\mathrm{jj}}|\bar{\theta}\right) &= P_{\mathrm{jj}}^{\mathrm{TT}}(m_{\mathrm{jj}}|\bar{\theta}_1)\, P_{\mathrm{T},1}(m_{\mathrm{jet1}}|\bar{\theta}_2(m_{\mathrm{jj}}))\, P_{\mathrm{T},2}(m_{\mathrm{jet2}}|\bar{\theta}_3(m_{\mathrm{jj}})) \\
P_{\mathrm{WW}}\left(m_{\mathrm{jet1}}, m_{\mathrm{jet2}}, m_{\mathrm{jj}}|\bar{\theta}\right) &= P_{\mathrm{jj}}^{\mathrm{WW}}(m_{\mathrm{jj}}|\bar{\theta}_1)\, P_{\mathrm{W},1}(m_{\mathrm{jet1}}|\bar{\theta}_2(m_{\mathrm{jj}}))\, P_{\mathrm{W},2}(m_{\mathrm{jet2}}|\bar{\theta}_3(m_{\mathrm{jj}})) \\
P_{\mathrm{nonRes}}\left(m_{\mathrm{jet1}}, m_{\mathrm{jet2}}, m_{\mathrm{jj}}|\bar{\theta}\right) &= P_{\mathrm{jj}}^{\mathrm{nonRes}}(m_{\mathrm{jj}}|\bar{\theta}_1)\, P_{\mathrm{nonRes},1}(m_{\mathrm{jet1}}|\bar{\theta}_2(m_{\mathrm{jj}}))\, P_{\mathrm{nonRes},2}(m_{\mathrm{jet2}}|\bar{\theta}_3(m_{\mathrm{jj}})) \\
P_{\mathrm{TnonRes}}\left(m_{\mathrm{jet1}}, m_{\mathrm{jet2}}, m_{\mathrm{jj}}|\bar{\theta}\right) &= 0.5\,\left(P_{\mathrm{jj}}^{\mathrm{TnonRes}}(m_{\mathrm{jj}}|\bar{\theta}_1)\, P_{\mathrm{nonRes},1}(m_{\mathrm{jet1}}|\bar{\theta}_2(m_{\mathrm{jj}}))\, P_{\mathrm{T},2}(m_{\mathrm{jet2}}|\bar{\theta}_3(m_{\mathrm{jj}}))\right) \\
&+ 0.5\,\left(P_{\mathrm{jj}}^{\mathrm{TnonRes}}(m_{\mathrm{jj}}|\bar{\theta}_1(m_{\mathrm{jj}}))\, P_{\mathrm{T},1}(m_{\mathrm{jet2}}|\bar{\theta}_2(m_{\mathrm{jj}}))\, P_{\mathrm{nonres},2}(m_{\mathrm{jet1}}|\bar{\theta}_3(m_{\mathrm{jj}}))\right) \\
P_{\mathrm{WnonRes}}\left(m_{\mathrm{jet1}}, m_{\mathrm{jet2}}, m_{\mathrm{jj}}|\bar{\theta}\right) &= 0.5\,\left(P_{\mathrm{jj}}^{\mathrm{WnonRes}}(m_{\mathrm{jj}}|\bar{\theta}_1)\, P_{\mathrm{nonRes},1}(m_{\mathrm{jet1}}|\bar{\theta}_2(m_{\mathrm{jj}}))\, P_{\mathrm{W},2}(m_{\mathrm{jet2}}|\bar{\theta}_3(m_{\mathrm{jj}}))\right) \\
&+ 0.5\,\left(P_{\mathrm{jj}}^{\mathrm{WnonRes}}(m_{\mathrm{jj}}|\bar{\theta}_1(m_{\mathrm{jj}}))\, P_{\mathrm{W},1}(m_{\mathrm{jet2}}|\bar{\theta}_2(m_{\mathrm{jj}}))\, P_{\mathrm{nonres},2}(m_{\mathrm{jet1}}|\bar{\theta}_3(m_{\mathrm{jj}}))\right) \\
P_{\mathrm{WT}}\left(m_{\mathrm{jet1}}, m_{\mathrm{jet2}}, m_{\mathrm{jj}}|\bar{\theta}\right) &= 0.5\,\left(P_{\mathrm{jj}}^{\mathrm{WT}}(m_{\mathrm{jj}}|\bar{\theta}_1)\, P_{\mathrm{T},1}(m_{\mathrm{jet1}}|\bar{\theta}_2(m_{\mathrm{jj}}))\, P_{\mathrm{W},2}(m_{\mathrm{jet2}}|\bar{\theta}_3(m_{\mathrm{jj}}))\right) \\
&+ 0.5\,\left(P_{\mathrm{jj}}^{\mathrm{WT}}(m_{\mathrm{jj}}|\bar{\theta}_1(m_{\mathrm{jj}}))\, P_{\mathrm{W},1}(m_{\mathrm{jet2}}|\bar{\theta}_2(m_{\mathrm{jj}}))\, P_{\mathrm{T},2}(m_{\mathrm{jet1}}|\bar{\theta}_3(m_{\mathrm{jj}}))\right),
\end{aligned}
\tag{7.5}
$$

where T (W) stands for merged top quark (W boson) and "nonRes" for the resolved hadronic top decay. Merged top quark (W boson) means that the particle decay product are reconstructed in a single jet with the mass corresponding to the particle mass. In the resolved hadronic top decay the top or W decay products are not merged in a single jet and the jet considered in the final state does not have a characteristic mass, and thus it is referred in the following as "non-resonant" component, despite the top quark still being a resonance. The parameters $\bar{\theta}$ are the free parameters of the background fit.

In the following, first, the jet mass modelling is described, which is initially done inclusively with respect to the six components. Then the procedure used to obtain the m_{jj} templates is illustrated. Finally, the 3D pdf is built. An additional fit to the simulation is required to obtain the correct normalization of the six different $t\bar{t}$ components described above.

Jet mass modelling Two resonant contributions are present in the jet mass, one peak at ≈ 80 GeV due to merged fully-hadronic W boson decays and a second peak at ≈ 170 GeV due to merged hadronic decays of the top quark. In addition to these, there is also a non-resonant component coming from non-merged top or W bosons. The separation between quarks is related to the mass and transverse momentum

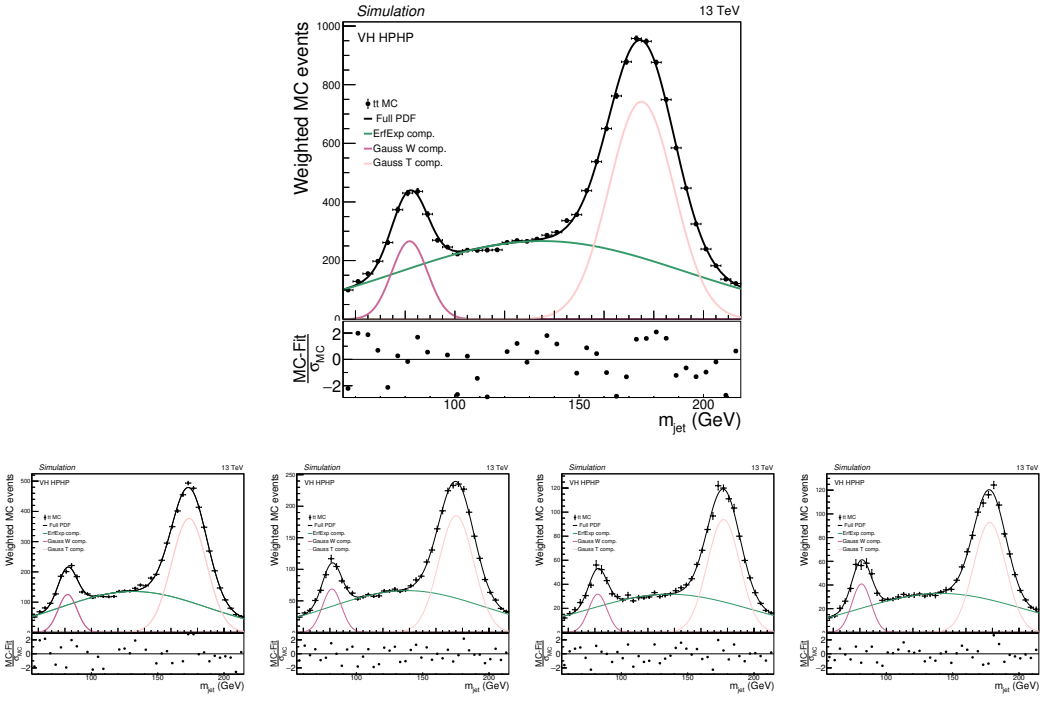


FIGURE 7.12: Fit to the inclusive (all bins of m_{jj}, top) and in increasing m_{jj} bins (bottom, from left to right: 1246–1404 GeV, 1404–1563 GeV, 1563–1722 GeV and 1722–7600 GeV jet mass distributions for the $t\bar{t}$ background in the VH HPHP.

of the decaying particle, as described in Eq 5.1. Therefore, with a W boson transverse momentum greater than 200 GeV [165], the two quarks from the W decay are merged into a single AK8 jet. With a top quark momentum greater than 450 GeV, the three quarks from the top quark decay become merged into an AK8 jet. With the increase of the top transverse momentum, the relative fraction of merged top quarks to merged W bosons rises. Due to the relationship between jet p_T and dijet invariant mass, the normalization of the resonant jet mass peaks depends on m_{jj} and needs to be modelled conditionally. A correlation between non-resonant and resonant components as a function of m_{jj} is also observed. The three (top, W, non-resonant) components are modelled separately. Thanks to the random sorting, m_{jet1} and m_{jet2} are added together to increase the sample size and, therefore, improve the stability of the fit. The VH HPHP category is used in the following to explain the modelling procedure. Distributions for the other categories can be found in Appendix B.3. In the VV categories, the top and W contributions are of similar size; in the VH categories, the top contribution dominates.

A Gaussian function is used to fit each of the resonant components, while an error function convolved with an exponential function (referred as ErfExp) is used for the non-resonant part of the jet mass spectrum. The shape of each of the three components is modelled as a function of m_{jj}, with the initial fit parameters obtained from the fit to the jet mass inclusive in m_{jj}, shown in Fig. 7.12 (top), to benefit from high statistics. Then the jet mass is sampled in bins of m_{jj}. The distribution in each bin is then fitted with the chosen pdf, i.e. the sum of the Erfexp with the two Gaussian functions modelling the resonant contributions. These fits are shown in Fig 7.12 (bottom).

The fit parameters are then extracted in each m_{jj}-bin, and a parametrization of the fit parameter distributions as a function of dijet invariant mass is obtained. The

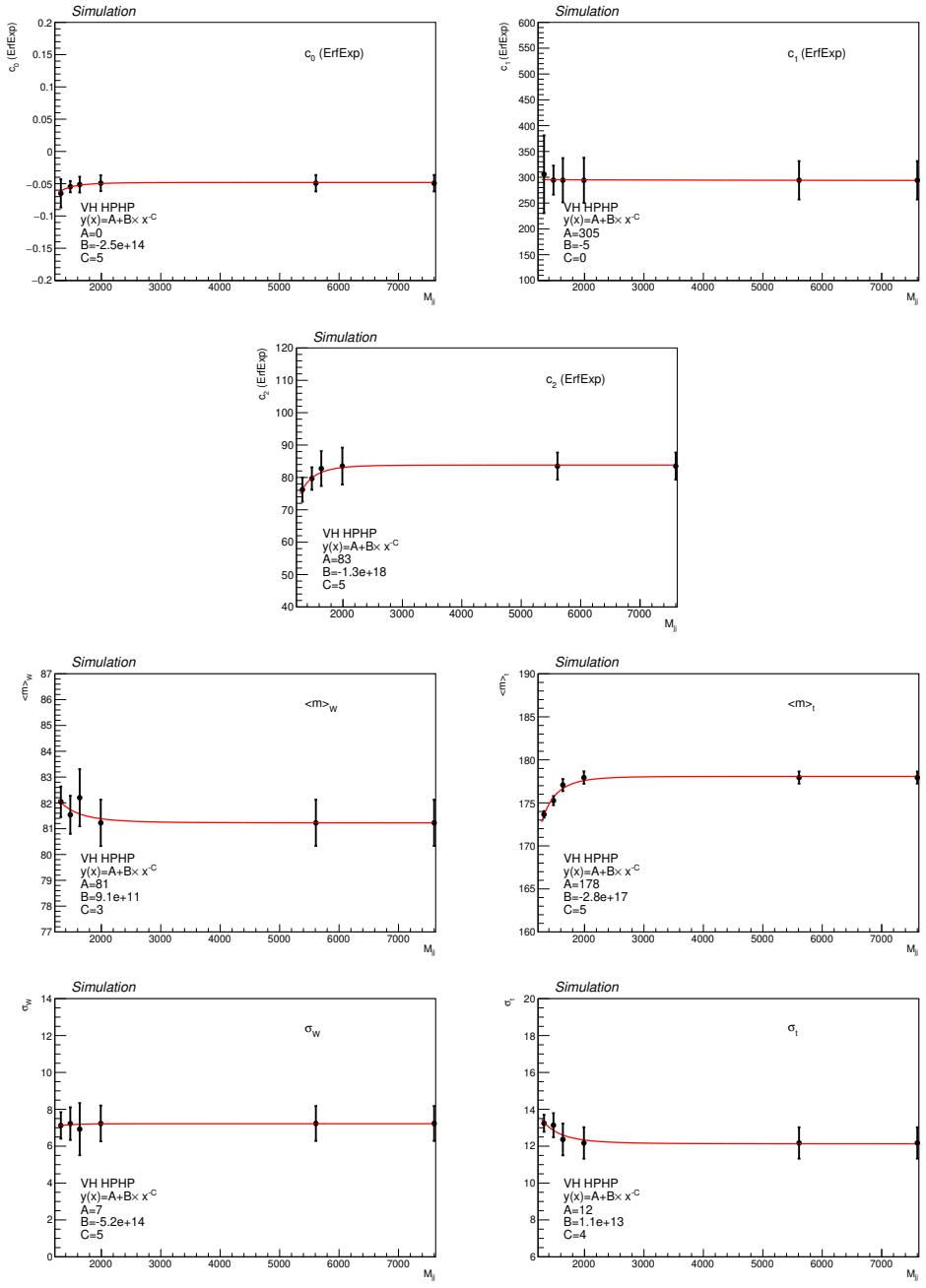


FIGURE 7.13: The parametrization of the fitted parameters for the $t\bar{t}$ background in the VH HPHP category. The first two rows show the parameters of the non resonant component while the second and third row show the mean and resolution for the W boson (left) and top quark (right) peaks.

m_{jj}-value for each bin is chosen to be the average m_{jj}-value in that bin. The parameter values in the highest fittable bin, at $\approx$ 1700 GeV, is used also at higher m_{jj} values to constrain the parametrization at high m_{jj}. An m_{jj} of $\approx$ 1700 GeV roughly corresponds to a jet p_T of 800–900 GeV, and thus to the regime where the top decay products are merged in a single jet. At even higher m_{jj} no significant changes in the shape are expected and thus the same parameter values as estimated at $m_{jj} \approx 1700$ GeV can be used. The parametrization of each fit parameter is shown in Fig. 7.13.

Finally, the validation is performed by refitting the $t\bar{t}$ MC using the parametrised

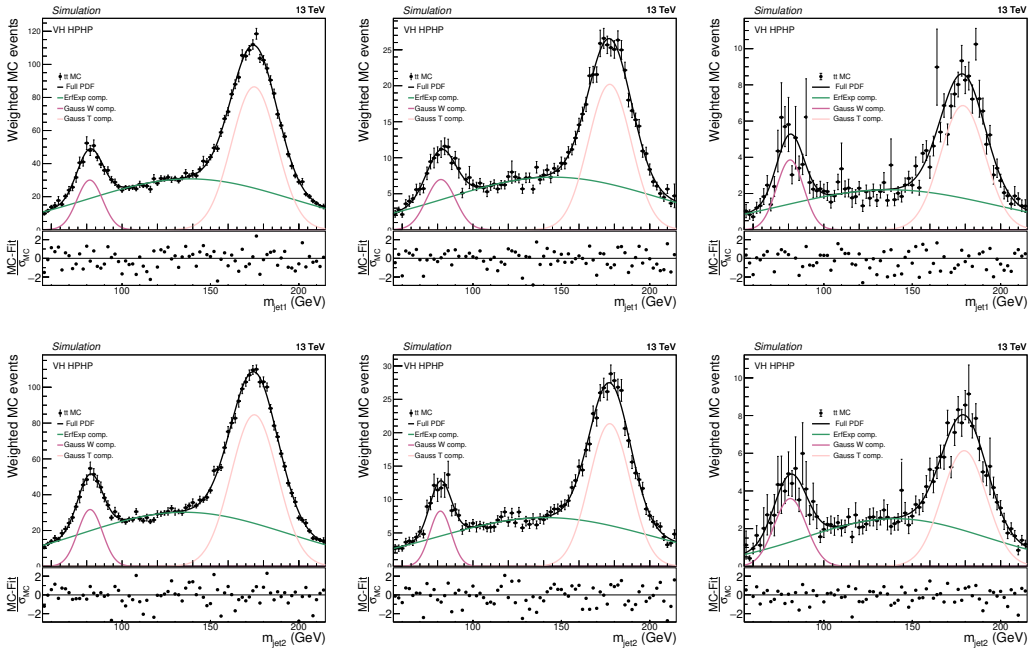


FIGURE 7.14: Closure test for the $t\bar{t}$ background pdfs in the VH HPHP category. Here, m_{jet1} (top) and m_{jet2} (bottom) are fit separately starting from the parametrizations of the $t\bar{t}$ background as a function of m_{jj}. This is shown in increasing m_{jj} bins: 1325–1563 GeV (left), 1643–1881 GeV (middle), 2040–7600 GeV (right).

shapes. Different distributions are generated to test the stability of the parametrization: 1) Each jet mass axis is fit separately (when doing the parametrization, the two contributions were added); 2) The default jet mass binning is used (80 bins, whereas for the parametrization a binning of 40 was used to increase the stability of the fit); 3) The dijet invariant mass bins are different than those used for the parametrization. The fitted jet mass distributions for m_{jet1} and m_{jet2} are shown in Figure 7.14. The pdfs can generalize to distributions not yet seen and accurately describe the jet mass dependence on m_{jj}.

Dijet mass modelling The MC simulations of the minor backgrounds contain fewer events than for the QCD multijet background. Therefore, a different approach is developed for the m_{jj} template building of the $t\bar{t}$ and V+jets components to obtain smooth templates with reduced sample size. The m_{jj} distribution is fitted with a three-parameters dijet function

$$\frac{\mathrm{d}N}{\mathrm{d}m} = \frac{P_0(1-m/s)^{P_1}}{(m/s)^{P_2}}\,, \tag{7.6}$$

where s is the center-of-mass energy and $P_{0/1/2}$ are free parameters. The value assumed by the fit function in each m_{jj} bin is used to build the template histogram, rather than the kernel-based template. Five alternative functions with 2, 4 and 5 parameters have also been tested and the chosen three-parameter function provides the best χ^2/N_{dof}, where N_{dof} is the number of degrees of freedom. The final 1D m_{jj} templates for the VH HPHP category are shown in Fig. 7.15, for the 6 $t\bar{t}$ contributions, obtained by matching the jets with generator level W bosons or top quarks. The other categories can be found in Appendix B.3. The fitting procedures allows

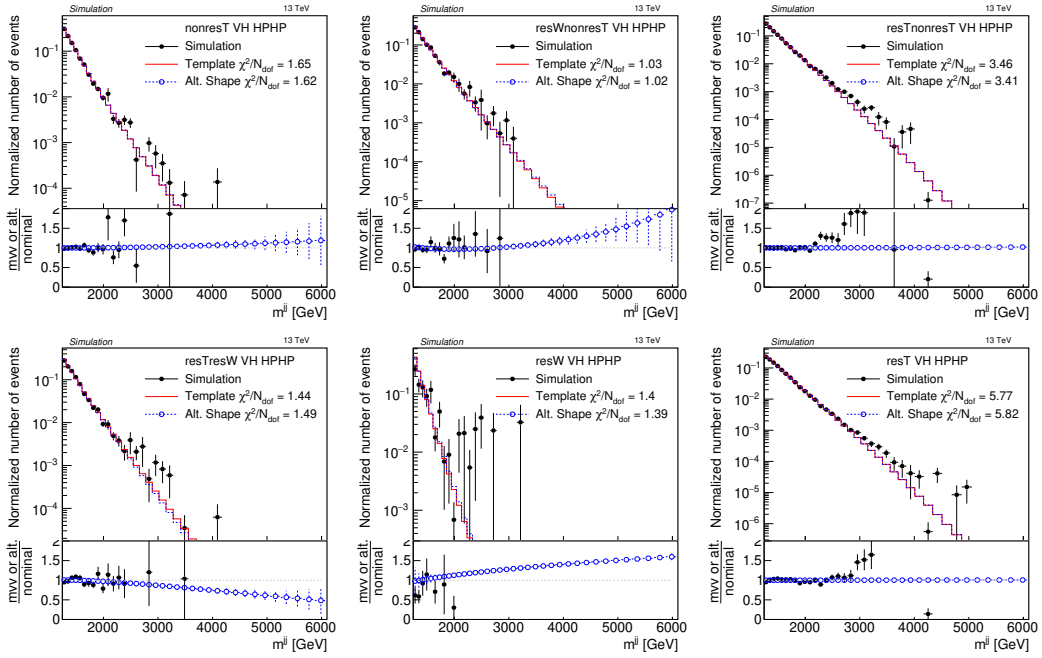


FIGURE 7.15: Final one-dimensional m_{jj} templates in the VH HPHP category. The simulation is shown as black markers, the nominal shape derived from the fitting procedure can be seen as red solid line, the alternative shape derived from varying the slope of the m_{jj} spectrum is shown as the blue dashed line. Top, from left to right: nonRes, WnonRes and TnonRes; bottom from left to right: WT, WW, TT.

to carefully model the lower part of the m_{jj} spectrum where these contributions are more relevant. While some disagreement is observed in the high m_{jj} tails, overall the procedure leads to smooth templates also with a reduced number of simulated events. Alternative dijet invariant mass shapes are also shown as derived using the same method as for the nominal shape but fitting the m_{jj} distribution without top p_T reweight corrections applied (see Sec. 7.3).

$t\bar{t}$ 3D pdf The modelling of the m_{jet} shape as a function of m_{jj} is obtained above. However, the relative fraction of the top/W/non-resonant components as a function of m_{jj} still has to be included. The final pdf is built summing the pdfs described in 7.5. Then, this model is fit to the full $t\bar{t}$ simulation, independently for each category, to obtain the initial normalization values for each $t\bar{t}$ contribution that will enter the final fit to data. This intermediate step is necessary because the generator-level matching procedure is not 100% efficient. The results of the fits are shown in Fig. 7.16 for the VH HPHP category, while the other categories can be found in Appendix B.3.

7.2.2.2 V+jets

For V+jets processes, since the jets in this analysis are randomly sorted, each jet mass distribution contains two contributions: the resonant part, with a real V-jet, shows a clear peak around the V boson mass, and the non-resonant part, composed of jets originating from a quark or a gluon, shows a distribution very similar to the one seen for the QCD multijet background.

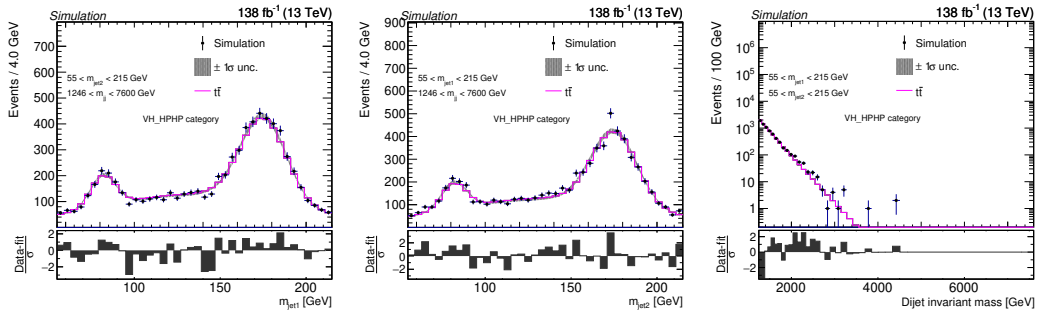


FIGURE 7.16: Post fit to the $t\bar{t}$ MC in the VH HPHP category.

The three-dimensional probability density function for the V+jets background is built as a product of three one-dimensional pdfs as follows:

$$P_{\text{V+jets}}\left(m_{\text{jet1}}, m_{\text{jet2}}, m_{\text{jj}}|\bar{\theta}\right) = 0.5\,\left(P_{jj}(m_{\text{jj}}|\bar{\theta}_1)\,P_{\text{res}}(m_{\text{jet1}}|\bar{\theta}_2)\,P_{\text{nonres}}(m_{\text{jet2}}|\bar{\theta}_3)\right) + 0.5\,\left(P_{jj}(m_{\text{jj}}|\bar{\theta}_1)\,P_{\text{res}}(m_{\text{jet2}}|\bar{\theta}_2)\,P_{\text{nonres}}(m_{\text{jet1}}|\bar{\theta}_3)\right)\,. \tag{7.7}$$

P_{res} (P_{nonres}) is the pdf describing the (non-)resonant contribution, while $\bar{\theta}$ are the fit free parameters. Exhaustive testing [9, 177, 178] shows that correlations between m_{jet} and dijet invariant mass are small enough to be neglected in the modelling of these backgrounds, i.e. the shape of the jet mass spectrum does not significantly depend on the momentum of the jets. There is, however, a correlation between the two jet masses. This correlation stems from the fact that if one of the two jets in the final state originates from a real boson, the second is a quark or gluon jet. Both of these jets have different distributions, which means that depending on the fraction of real boson jets in m_{jet1} the fraction of such jets in m_{jet2} will change. For this reason, these backgrounds are modelled as if there were two different contributions: one where the first jet is resonant ($P_{\text{res}}(m_{\text{jet1}}|\bar{\theta}_2)$) and the second non-resonant ($P_{\text{nonres}}(m_{\text{jet2}}|\bar{\theta}_3)$), and one where the second jet is resonant ($P_{\text{res}}(m_{\text{jet2}}|\bar{\theta}_2)$) and the first non-resonant ($P_{\text{nonres}}(m_{\text{jet1}}|\bar{\theta}_3)$). These two contributions are added together with a fraction $f = 0.5$ for the random labeling of jets as in Eq. 7.7. The resonant events (described by P_{res}) are separated from the non-resonant (P_{nonres}) ones by requiring that there is a generated V boson in a cone of $\Delta R = 0.8$ around the reconstructed merged jet. A double-sided crystal-ball function, the same function used to model the signal m_{jet} shapes, is fitted to the resonant spectrum for W+jets and Z+jets separately. This treatment allows to fully correlate the uncertainties on the mean and width of the m_{jet} distribution with the signal, since these uncertainties affect all jets originating from real vector bosons in the same way. In Fig. 7.17 the final fit of the dCB function to the resonant part of the W+jets (left) and Z+jets (right) spectrum, respectively, is shown for the VV HPHP category, while the other categories can be found in Appendix B.4.

The non-resonant part of the V+jets backgrounds is modelled using a simple fit with a Gaussian function to the non-resonant part of the spectrum as shown in Fig. 7.18 for the W+jets (left) and Z+jets (right) spectrum, respectively, for the VV HPHP category, while the other categories can be found in Appendix B.4.

For the dijet invariant mass pdf P_{jj}, the same approach as for the $t\bar{t}$ background is used to obtain smooth templates with reduced sample size. The m_{jj} distribution is fitted with the three-parameters dijet function in Eq. 7.6. The final 1D m_{jj} templates are

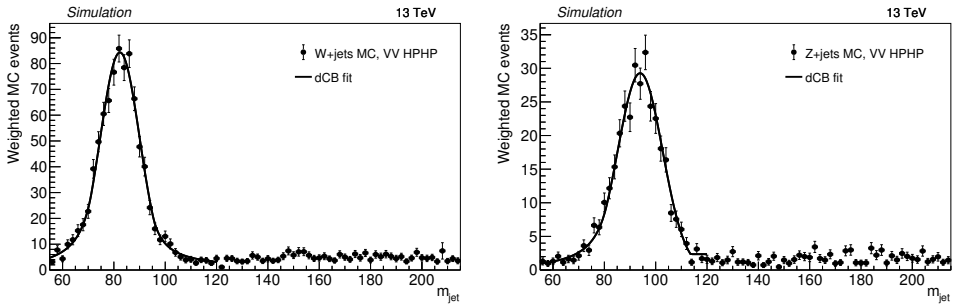


FIGURE 7.17: The fit to the resonant part of the W+jets (left) and Z+jets (right) spectrum (black markers) with a dCB function (black line) in the VV HPHP category.

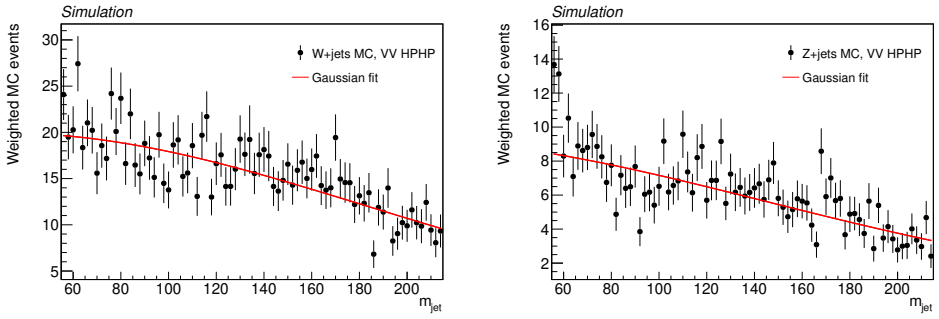


FIGURE 7.18: The fit to the non-resonant part of the W+jets (left) and Z+jets (right) spectrum (black markers) with a dCB function (red line) in the VV HPHP category.

shown in Fig. 7.19 compared to MC simulation, for W+jets (left) and Z+jets (right) in the VV HPHP category. The other categories can be found in Appendix B.4. Alternative dijet invariant mass shapes are also shown, derived using the same method (see Section 7.3). They are added to the final fit to accommodate mismodelling of the m_{jj} distribution, for example due to higher order QCD and electroweak corrections, not considered in the MC simulation samples.

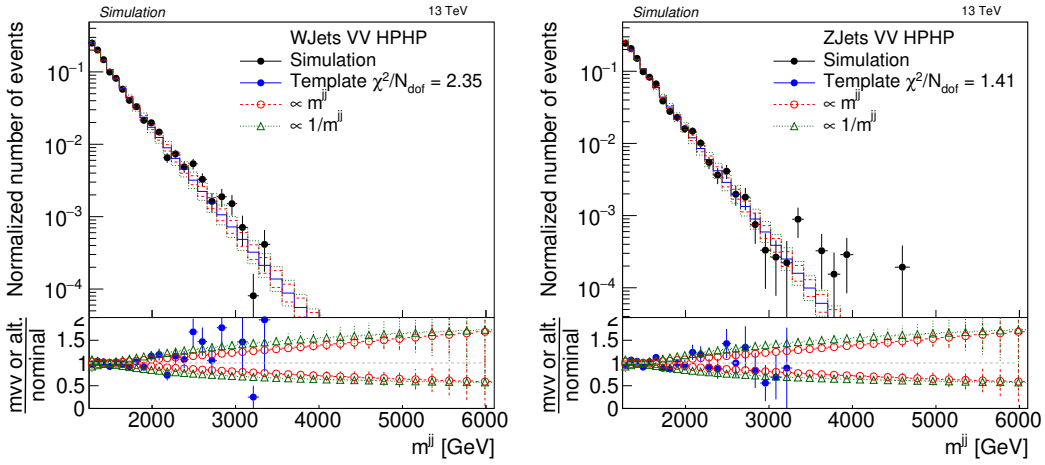


FIGURE 7.19: Final one-dimensional m_{jj} templates for the resonant W+jets (left) and Z+jets (right) background compared to MC simulation for the VV HPHP category. The nominal shape derived from the fitting procedure can be seen as blue line, alternative shapes derived from varying the slope of the m_{jj} spectrum are shown in green and red.

7.3 Systematic uncertainties

Systematic uncertainties affect both the yield and the distribution shapes of the background and signal models. This section discusses the systematic uncertainties considered in the analysis and included as *nuisance parameters* in the multidimensional fit. Specific uncertainties are added to the description of the signal model only. The nuisance parameters for each uncertainty source are profiled in the statistical interpretation using log-normal constraints for normalization uncertainties and Gaussian constraints for shape uncertainties.

While differences exist among the various signal hypothesis considered in the analysis, the most limiting uncertainties are represented by the uncertainties related to the estimation of the tagging efficiency, followed by the uncertainties in the resonance mass scale and resolution. Besides these, uncertainties in the prediction of the resonant and non-resonant background normalizations also have a significant impact.

First the normalization uncertainties are listed in Sec. 7.3.1, then the shape uncertainties are presented in Sec. 7.3.2.

7.3.1 Normalization uncertainties

7.3.1.1 Signal normalization uncertainties

Several systematic uncertainties affect the determination of the signal yield, as well as the sub-dominant processes. Here the uncertainties affecting the signal are listed. Contributions such as luminosity uncertainty or the tagging efficiency are assumed fully correlated with the partially resonant backgrounds.

- **Jet tagging efficiency** This uncertainty is evaluated for each category by varying up and down the correction factors (applied as described in Sec. 6.3.2) by their measured uncertainty (Table 6.8) and taking the relative difference in signal yields. Two nuisance parameters are obtained per each year, one for the $H/Z\to b\bar{b}$ and one for the $V\to q\bar{q}$ tagging, which are anti-correlated between VH and VV categories and between HP and LP categories. Therefore, six nuisance parameters in total are added to the fit. The impact on the signal normalization is estimated for each year. H-tagging uncertainties are of the order of 5%–10%. V-tagging uncertainties are $\approx$ 15%, with a larger 30% uncertainty observed in 2016 in the VV HPHP categories. However, such uncertainty is then rescaled by a weight corresponding to the fraction of luminosity collected in each year with respect to the total luminosity collected in 2016–2018.

- **Extrapolation of jet tagging efficiency to higher p_T**: An additional uncertainty arises from the extrapolation of the jet tagging efficiency scale factors (see Sec. 5.3.4.2), which are measured in $t\bar{t}$ or $g\to b\bar{b}$ events at relatively low p_T, towards higher transverse momenta. The extrapolation of a p_T-dependent uncertainty from the available scale factor measurement has too large statistical fluctuations to infer on the high p_T behaviour. In the previous version of this analysis a logarithmic dependence on the p_T was found for the tagging efficiency comparing the performance of the tagger on simulations obtained with different generators. This approach cannot be used with the ML-based taggers used in this thesis, as it would require to train again the DeepAK8. However, the p_T dependence $7\%\ln(p_T/300\,(\text{GeV}))$ estimated in Refs. [9, 177, 178] was found to be as large or larger than a linear extrapolation to high p_T

of the tagging efficiency estimated with the DeepAK8-MD. Furthermore, this analysis features a non-trivial combination of different V- and H-jet taggers that must be considered. For this reason, this uncertainty is implemented in a similar way as the nominal jet-tagging efficiency uncertainty. In particular, two separate nuisance parameters are again considered for the V- and H-jet tagging algorithms, anti-correlated among categories, as described above. As uncertainty, a value that is 1.5 times larger than the nominal V/H-jet tagging efficiency is introduced for the whole resonance mass range to simplify the implementation. The chosen value yields an uncertainty as large as the one provided by logarithmic extrapolation at high p_T and larger at low p_T, thus well covering the extrapolation uncertainty in a conservative approach.

- **Jet energy calibration uncertainties**: As it was described is Sec. 5.2, energy scale and resolution of reconstructed jets must be calibrated and uncertainties are associated to such procedure. The effect on the signal yield is evaluated separately for ggF/DY and VBF categories by varying up/down the central JES and JER values. This variation is evaluated for a luminosity-weighted sum of the three years simulation where the three years are varied together, obtaining one correlated uncertainty. It takes also into account the migration of events from the ggF/DY category to the VBF one (and vice versa), due to the requirement on the m_{jj} of the AK4 jets, described in Sec. 6.2. In general, JER uncertainties are lower than 1% in ggF/DY categories. In VBF categories, they are less than 2% for VBF signals and of the order of 4–20% for ggF/DY produced signals. Similarly, JES uncertainties are less than 1% ($\approx 5\%$) for ggF/DY (VBF) signals in ggF/DY categories. Larger uncertainties are observed in VBF categories, $\approx 10\%$ for VBF signals and up to 30% for spin-1 DY produced resonances. It has to be noted that extra jets are not expected for ggF/DY produced resonances, resulting in large uncertainties values in VBF categories, where a small amount of such signal events falls.

- **Luminosity**: The integrated luminosity of the recorded data is measured by the CMS Collaboration with an uncertainty of 1.6% [78–80] for the total 2016–2018 combined dataset, correlated among all categories;

- **PDF and factorization and renormalization scale uncertainties**: Uncertainties in the PDFs and of the choice of factorization and renormalization scales affect the expected signal cross section and acceptance. They can be evaluated by considering differences in the predicted kinematics of the resonance. The (large) uncertainties on the cross section are highly model-dependent and therefore not considered in the statistical analysis. Instead, an uncertainty on the signal acceptance dependent on the resonance mass is evaluated for each signal benchmark by reweighting the events according to the ≈ 100 available PDF variations and calculating the standard deviation of the distribution of acceptances corresponding to the variations. The relative variations for some of the signal benchmarks are shown in Appendix C. For each signal benchmark, the function obtained by fitting the variations corresponding to the first and last mass points with an exponential function is taken as M_X-dependent uncertainty, considering only the most sensitive category for each of the benchmarks. With this approach, the exact variations are considered at the extreme values of the probed masses, while a conservative value is taken in the intermediate mass region. At the lowest M_X the uncertainty are of the order of 1%

or below while at the highest M_X they are of the order of 5–10%. By considering only the most sensitive category, an overestimation of the uncertainties is avoided in categories of very low signal statistics where large fluctuations are expected due to the very small number of expected events.

- **Pileup reweighting**: As can be seen in Appendix A, data have a slightly different pileup profile than what was used to generate MC events. Therefore a weight is applied to the simulation to match the number of pileup interactions observed in data and an uncertainty is evaluated by shifting up and down the minimum bias cross section by 4.6%. For each shift a new weight is calculated and applied, obtaining up and down variations of the yields. These variations are at most 1% for all signals independently of the generated mass and in both VBF and ggF/DY categories. An exception is observed for ggF/DY signals in the VBF category where variations are found to be at most 4%. This is due to the fact that ggF/DY signals do not contain genuine forward jets and their tagging might be mainly due to the presence of additional pileup jets. This uncertainty is correlated among the years.

- **L1 prefiring**: During the 2016 and 2017 data taking, detector effects (see Sec. 6.2.1) caused a loss of a fraction of events in the trigger. A weight is applied to simulated events to mimic this effect and an uncertainty is evaluated by varying up/down the central event weight. Four nuisance parameters are included in the fit such that the effect in 2016 is uncorrelated with the one in 2017 and ggF/DY categories are uncorrelated from VBF ones. The largest measured uncertainty of 1% is applied on both VBF and ggF/DY signals of all masses.

7.3.1.2 Background normalization uncertainties

In the final fit, the QCD multijet background is allowed to float within 50% of the yield expected by the simulation of the SM process, independently in each category. To account for potential deviations due to missing higher-order corrections as well as significant mismodelling of the misidentification rate of the q/g jet [165], a flat uncertainty of 50% is applied to the W+jets and Z+jets background yields separately in each individual category. These normalization uncertainties are then constrained through the final fit to data.

For $t\bar{t}$, a cross section uncertainty of 6% is assigned from the sum in quadrature of the scale uncertainty and the PDF+α_S uncertainty on the cross section [188, 189]:

$$\sigma_{t\bar{t}} = 831.76\ ^{+19.77}_{-29.20}\ (\text{scale})\ ^{+35.06}_{-35.06}\ (\text{PDF} + \alpha_S).$$

This uncertainty is correlated among categories.

As for the signal, the V+jets and $t\bar{t}$ resonant background normalizations are affected by the uncertainties in jet tagging efficiency, derived with the same method described above. H-tagging uncertainties are of the order of 1–5% for both V+jets and $t\bar{t}$ contributions (larger values are observed in $t\bar{t}$ contributions with two resonant W boson because of limited statistics), while V-tagging uncertainties can be up to 15%. The $t\bar{t}$ sub-contributions with at least one merged top quark jet, are also affected in addition by the uncertainties in the top misidentification scale factors (between less than 1% to up to 10%). Uncertainties on the extrapolation of jet tagging efficiency at high p_T are also considered, with the same implementation described for the signal.

As for the signal, for V+jets and $t\bar{t}$ four additional nuisance parameters are included to take into account the uncertainty in the L1 prefiring in 2016 and 2017 and in ggF/DY and VBF categories, separately. Finally, the uncertainty due to PU reweighting is evaluated with the same method as for the signal and it is found to be 0.1% for the ggF/DY categories and 2% for the VBF categories. Also for V+jets and $t\bar{t}$, two additional nuisance parameters are included to take into account the yield variation due to JES and JER uncertainties, as for the signal. JES (JER) uncertainties are of the order of 3–4% (less than 1%) in ggF categories and 20–27% (6–10%) in VBF categories, with a similar behaviour as the one observed for ggF/DY produced signals.

7.3.2 Shape uncertainties

The shape uncertainties discussed in the following either affect the shape of the signal or background processes. For the background processes, artificially large uncertainties are presumed, due to the strong dependence on the simulation generators, and constrained through the multi-dimensional fitting procedure.

7.3.3 Signal shape uncertainties

The jet energy scale and resolution uncertainties affect the mean or standard deviation of the dCB function parametrizing the dijet invariant mass of the resonances. Jet mass scale and resolution uncertainties have a similar effect on the jet mass spectra.

The impact of the JES and JER uncertainties is evaluated by reweighting the transverse momentum of the event by up/down JES and JER variations [153]. The uncertainty is evaluated for a luminosity-weighted sum of the three years MC where the three years are varied together, obtaining one correlated uncertainty. For each signal and mass the dijet invariant mass distribution is fitted with a double crystal ball and the largest relative variations in the mean and sigma among signal and mass points is taken as the uncertainty, obtaining 1.2% and 6% for the mean and σ, respectively. A JER uncertainty of 1.2% is a representative value for high resonance masses and conservative at low resonance masses. A JES uncertainty of 6% is representative of the behaviour at the lower and higher mass points and conservative in intermediate mass values.

Effects of groomed mass scale and resolution affect the mean and σ of the double crystal ball used to fit the m_{jet} shape. Furthermore, they are correlated among categories and with the V boson and top quark jet peak in V+jets and $t\bar{t}$. These uncertainties were estimated simultaneously with the 2D-decorrelated DeepAK8-MD WvsQCD scale factors and found to be 2% and 8% for the mean and σ, respectively.

Additionally, PDF, factorization and renormalization scale uncertainties affect both the m_{jj} and m_{jet} shape. However, they were evaluated in Refs. [9, 177, 178] and found mostly smaller than 1%. They are therefore negligible compared to the effects from other uncertainty sources and not included in the fit.

7.3.4 Background shape uncertainties

7.3.4.1 QCD multijet background

Uncertainties in the QCD multijet background shape are included as alternative pdfs, derived with the template-building method described in Sec. 7.2.1. Shape uncertainties can simultaneously affect all three dimensions. In total 5 shape nuisance parameters are introduced in addition to the nominal shape obtained from PYTHIA

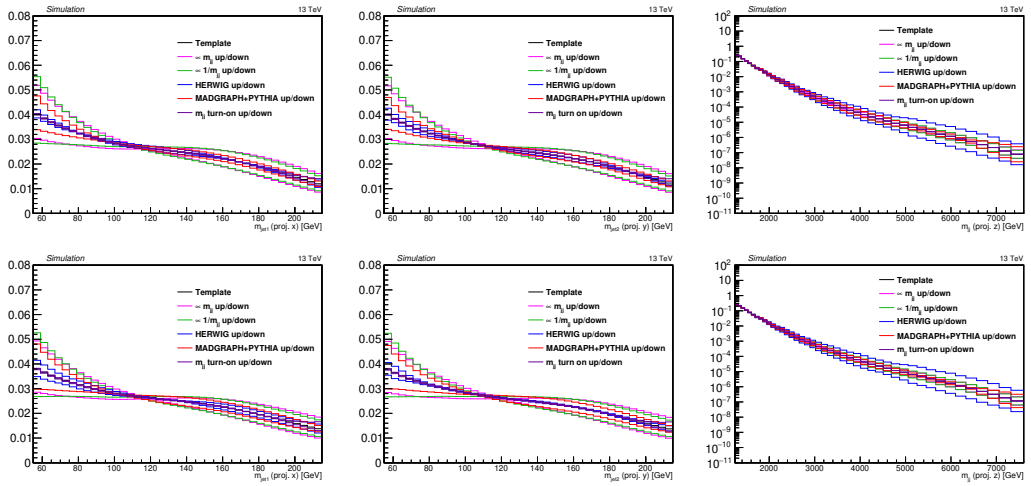


FIGURE 7.20: The QCD PYTHIA8 MC data and nominal template is shown together with the five alternate shapes added to the fit as shape nuisance parameters for the VH HPHP (top) and VV HPHP (bottom) categories.

MC simulation, for each tagging category. They cover the possible effects described below:

- Two (up/down) alternate shapes obtained by simultaneously varying jet masses and m_{jj} by a quantity proportional to m_{jj} and m_{jet} accounts for variations of the underlying transverse momentum spectrum.
- A variation of scale is taken into account with two alternate shapes obtained by simultaneously varying up and down jet masses and m_{jj} by a quantity proportional to $1/m_{jj}$ and $1/m_{jet}$.
- Differences in MC generation and modeling of parton shower are considered by alternate shapes that simultaneously affect resonance mass and jet groomed mass obtained using HERWIG++ and MADGRAPH+PYTHIA8 MC samples through the procedure described in Sec. 7.2.1, for a total of four shapes (considering the up/down variations for each generator).
- An additional shape takes into account mis-modeling of the extreme phase space given by large jet masses and low m_{jj}, where an inefficiency is expected inducing a turn on in m_{jj}. This effect was found in [9, 177, 178] hard to model both in MC and in real data after studies in a control region. This additional shape allow the nominal template to adjust the mis-modeling also in this part of the phase space.

Those five shape uncertainties are assigned very large pre-fit values (allowed to float within a sigma variation of the starting value), effectively allowing the simulation to take any value to fit the data and thus obtain the data-driven estimation of the QCD multijet background. The alternate shapes described above are shown in Figure 7.20 for high purity categories and in Appendix C for the lower purity categories.

7.3.4.2 Partially resonant backgrounds

The resonant part of the m_{jet} spectra are affected by the same uncertainties as the signal, namely jet mass scale and resolution uncertainties, as described above. Assuming these uncertainties as fully correlated between signal and the V+jets/$t\bar{t}$ backgrounds allows to constrain them, when fitting the W/Z boson and top peaks in the jet mass spectrum in data.

Analogous to the treatment of the non-resonant QCD backgrounds, alternative shapes obtained by varying the m_{jj} spectrum are derived for the m_{jj} templates of the resonant V+jets backgrounds and added to the fit with separate nuisance parameters for W+jets and Z+jets contributions. These nuisances account for variations in the jet p_{T} spectrum due to higher order corrections that are not included in the simulation.

In order to account for differences between the shapes of the measured and simulated top p_{T} spectra, due to the absence of NNLO order in the simulation, a reweight is applied to the $t\bar{t}$ MC samples using p_{T}-dependent scale factors. For this background, one alternative shape is introduced, where the NNLO top p_{T} weights are not included.

7.4 The 3D-fit

The signal extraction and background estimation procedures rely on a simultaneous fit to the three-dimensional m_{jj}-m_{jet1}-m_{jet2} data distributions in the 10 event categories of the signal region introduced in Sec. 6.3, five tagging categories for each of the two resonance production categories (ggF/DY and VBF). During the fit, the background normalizations and shapes vary within the uncertainties controlled by the nuisance parameters described in the previous section.

The following description of the maximum likelihood fitting procedure is based on Refs. [190, 191], where more details can be found.

The maximum likelihood method [191] is used to measure the set of nuisance parameters that maximizes the probability of obtaining the analyzed dataset. The dataset consists of the histograms of the kinematic variable of interest x, the m_{jj}-m_{jet1}-m_{jet2} plane in this case. The expectation value of each bin n_k is $E[n_k] = \mu s_k + b_k$ where s_k and b_k are the mean number of entries in the k-th bin from signal and background, respectively, and μ determines the strength of the signal process ($\mu = 0$ corresponds to the background-only hypothesis and $\mu = 1$ is the nominal signal hypothesis). The mean number of entries for the process p (with $p = s, b$) is given by

$$p_k = p_{\text{tot}} \int_{\text{bin } k} f_p(x; \bar{\theta}_p) dx \tag{7.8}$$

with $f_p(x; \bar{\theta}_p)$ the probability density function (pdf) of the variable x characterized by the parameters $\bar{\theta}_p$. The quantity p_{tot} is the total mean number of events of the process p. While b_{tot}, together with $\bar{\theta}_b$ and $\bar{\theta}_s$, is a nuisance parameter of the fit, the signal normalization s_{tot} is fixed to the value predicted by the nominal signal model. The likelihood function is then build as the product of the (Poissonian) probabilities $P(\mu s + b)$ for all bins k:

$$L(\mu, \bar{\theta}) = \prod_k P_k(\mu s_k + b_k) = \prod_k \frac{(\mu s_k + b_k)^{n_k}}{n_k!} \mathrm{e}^{-(\mu s_k + b_k)}. \tag{7.9}$$

By maximizing the likelihood function described in Eq. 7.9, it is possible to find the set of parameters $\hat{\hat{\theta}}$ that gives the highest probability of obtaining the observed dataset.

Considering a pdf $f(x|m)$ described by a morphing parameter m, such pdf can be sampled for different values m_i of the morphing parameter, each representing a known input template shape. This approach is followed in the pdfs characterizing the background prediction, where alternative binned template shapes are derived, corresponding to the m_i morphing parameters above. While obtaining the real m value is difficult, if at all possible, through template morphing [190], it is possible to obtain smooth interpolations among the different templates. The predicted pdf through template morning interpolation is

$$f_{\text{pred}}(x|m) \approx \sum_{i,j=0}^{n-1} (m - m_0)^j (M^{-1})_{ij} f(x|m_i) \tag{7.10}$$

where $M_{ij} = (m_i - m_0)_j$ defines a $n \times n$ transformation matrix. The parameter m_0 is an arbitrary chosen reference value for the morphing parameter and n is the number of sampled templates (alternative shapes, in this analysis). The prediction is then a weighted combination of the input templates $f(x|m_i)$. Thus, combining this technique with the maximum likelihood calculation allows the adaptation of binned predictions to a given data distribution.

Before probing the existence of new particles, the fitting procedure is validated in different ways, described in the next section. In the final fit procedure the jet mass binning has been reduced from a 2 GeV binning to a 4 GeV binning to reduce the computing time and prevent other technical issues.

7.5 Fit validation

Several tests are performed to validate the multi-dimensional fitting procedure and ensure that the model can reproduce the data and correctly measure signal events, if present. The tests are initially performed in either data control regions, where no signal is expected, or using simulated data. Analyses searching for new particles are performed *blind*, i.e., while defining the analysis methods or selections, the kinematic regions where a new signal might be located are not considered. Once the full procedure has been validated, the analysis is *unblinded* and the signal region or the true data are fitted. This approach reduces the possibility to introduce unintentional biases, eventually leading to false discoveries. The unblinded results are presented in Chapter 8.

Two scenarios are tested: the one corresponding to the lack of signal contribution (*background only*) and the alternative scenario of the presence of a signal (*signal+background*).

Figure 7.21 shows the post-fit pull of each nuisance parameter for a $G_{\text{bulk}} \rightarrow WW$ signal with a mass of 3 TeV, considering both scenarios. The pull is defined as the difference between the post-fit θ and pre-fit θ_{in} nuisance parameter values divided by the pre-fit uncertainty σ_θ. The observed large constraints are expected as a data-driven background estimation is performed, and normalization and shape parameters are effectively estimated during the fit. The nuisance parameters with the label `nonRes` are affecting the QCD non-resonant background; the ones with a label `Top` affects the top background; the ones with a label `Wjets` (`Zjets`) affects the W+jets (Z+jets) background. For QCD background the `PT` (`OPT`) shape nuisance parameters

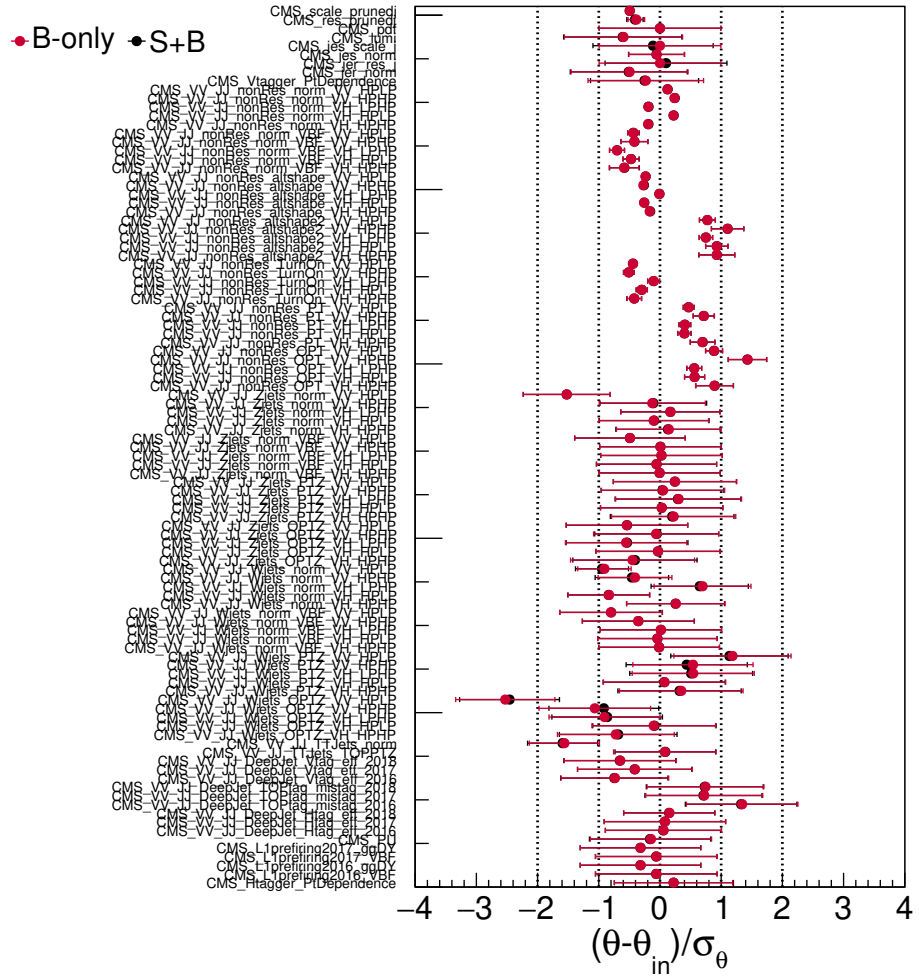


FIGURE 7.21: Pull of each nuisance parameter for a background-only and a signal+background fit, where the signal is the ggF $G_{bulk} \rightarrow WW$ with a mass of 3 TeV.

are derived by simultaneously varying jet masses and m_{jj} by a quantity proportional to m_{jj} $(1/m_{jj})$ and m_{jet} $(1/m_{jet})$ as explained in Section 7.3.4.1. There is an additional nuisance parameter for the non-resonant QCD background labelled as `TurnOn` which takes into account a data/MC discrepancy at high-mjet and low m_{jj} (Sec. 7.3.4.1). As explained in Sec. 7.3.4.1, the non-resonant QCD background also has two additional nuisance parameters that are associated to the shapes that can be obtained by the two different generators and parton shower models (labelled as `altshape`). For both non-resonant QCD and partially resonant V+jets or $t\bar{t}$, the nuisance parameters with the label `norm` affect the normalization of these backgrounds independently. The V+jets background also has two more additional nuisance parameters (labelled as `PT/OPT`) which are derived in a similar manner as for the QCD non-resonant background but they affect only the m_{jj} dimension. Finally, the nuisance parameters associated to the top background shape are labelled as `TOPPTZ` and are derived by varying up/down the top pt weight as described in Section 7.3.4.1. Other important uncertainties are the ones related to the jet mass peak shape (scale and resolution), the V/H tagging uncertainty, the top mistagging uncertainty, L1 prefiring, pileup, and jet energy scale and resolution affecting both normalization of signal/background as well as signal m_{jj} shape.

To further assess the quality of the fit and ensure that the background model can

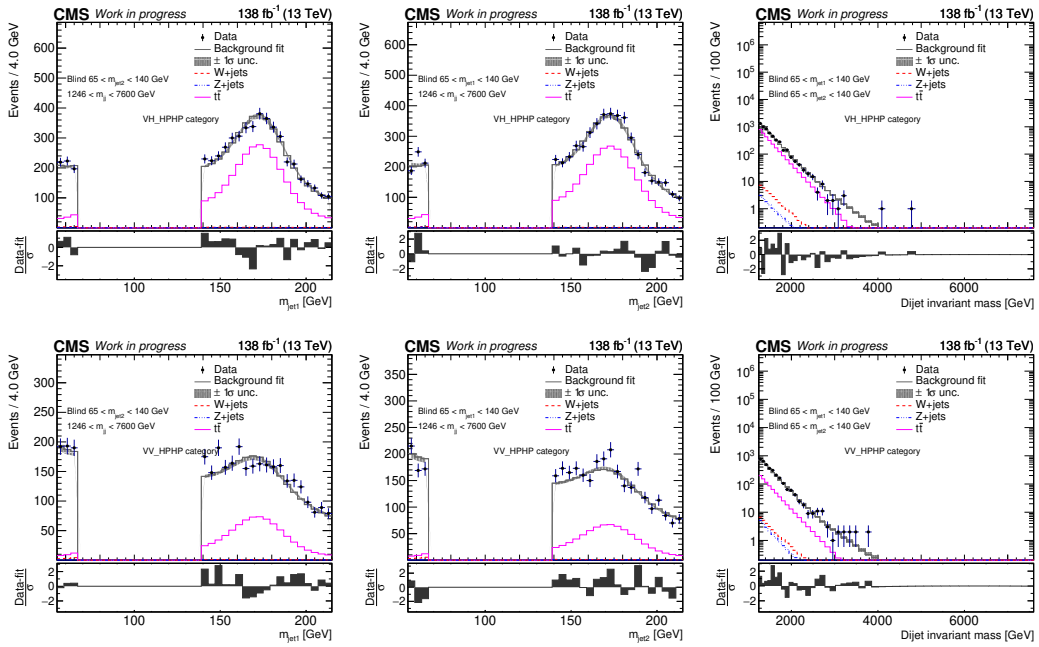


FIGURE 7.22: Projections of data and post-fit distributions onto the m_{jet1} (left), m_{jet2} (middle), and m_{jj} (right) dimensions for the VH HPHP (top) and VV HPHP (bottom) categories.

reproduce the data spectrum, a background-only fit is performed. The agreement is verified in a data control region, defined by excluding events where the mass of both jets is in the range 65–140 GeV where the signal is expected. The post-fit distributions are shown in Fig. 7.22 for the two high purity categories. The other categories can be found in Appendix C. A good agreement is observed in all the categories.

To probe that the model is accurate and able to describe a potential signal, *bias tests* are performed with the injection of signals in simulated data. First, it is verified that if no signal is injected, no fake signal is detected. Then, a signal is injected and the ability of the fit to measure the correct number of signal events is validated. In both tests, for each toy experiment, generated from the background simulation setting the nuisance parameter to their nominal value, a signal+background fit is performed and the number of background and signal events is extracted from the fit. The bias, defined as the difference between fitted and injected signal strength, is divided by the uncertainty on the fitted signal strength to obtain the pull distribution. Ideally, the pull has a Gaussian distribution centred at zero (indicating no bias) with unitary standard deviation (indicating a good estimation of the uncertainty in the fitted signal strength). Figure 7.23 shows the pull distributions obtained by injecting a $G_{\text{bulk}} \rightarrow$ WW signal in the toys, with a strength corresponding to an expected significance of zero (left), two (middle) or five (right) standard deviations. The denominator in the pull definition in these cases is such that the upper or lower uncertainty in the signal strength is taken in the direction of the injected signal strength value, to mitigate possible effects from non-parabolic likelihoods. In most cases, no or small biases are observed, and the standard deviation of the Gaussian, close to 1, indicates that the uncertainty in the fitted signal strength is correctly estimated. A relatively small bias is observed for the lowest mass point (1.4 TeV), up to about 30% of the uncertainty in the fitted signal strength for an injected signal of five standard deviations. For the highest considered mass point, 3 TeV, where the

background contribution is low, the pull distribution in the scenario where no signal is injected deviates from a Gaussian distribution. However, almost no pulls are above 2, indicating that no fake signal is fitted.

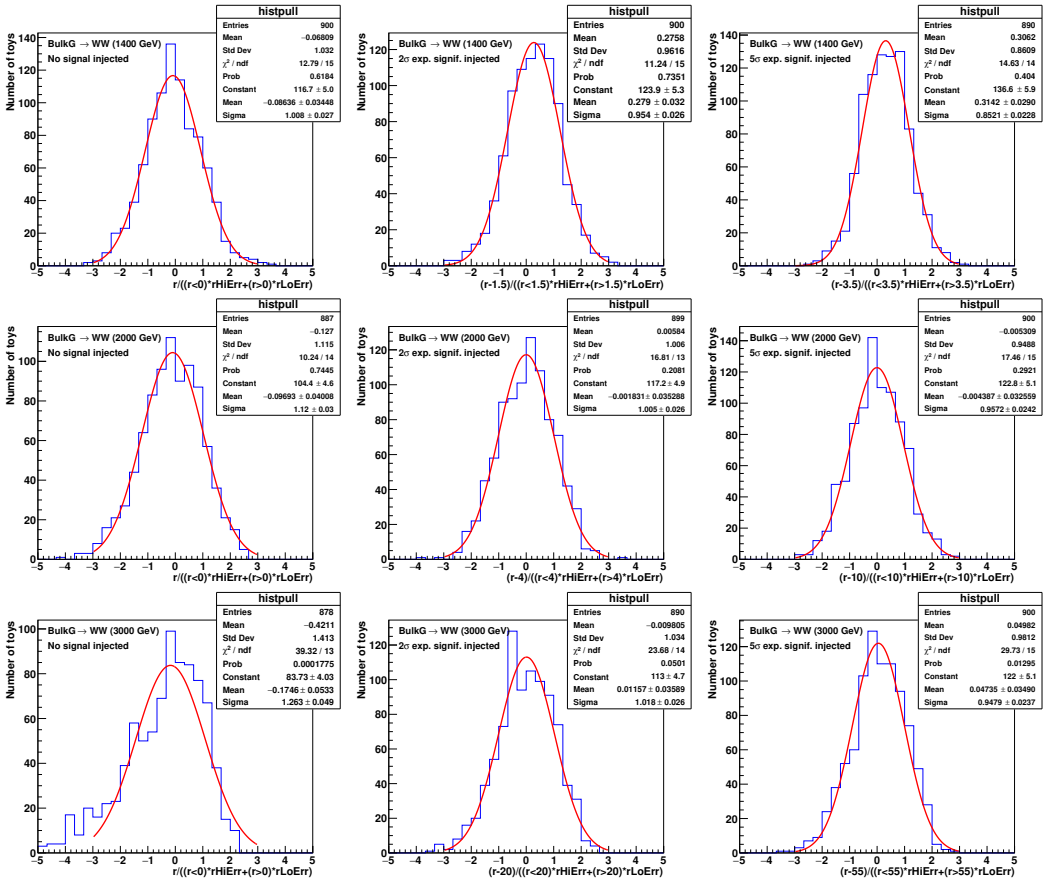


FIGURE 7.23: Pull distribution for toy experiments with a $G_{bulk} \rightarrow WW$ signal injected corresponding to an expected significance of zero (left), two (middle) or five (right) standard deviations, for a signal mass of 1.4 TeV (top), 2 TeV (middle), and 3 TeV (bottom).

8 Results and interpretation

The results presented in this chapter are based on the event selection and reconstruction described in Chapter 6 and on the multidimensional background estimation and signal extraction procedures described in Chapter 7.

Once the fit procedure has been thoroughly validated, the data are *unblinded*, and the signal existence can be probed. First, distributions comparing data to the outcome of the fitting procedure are presented in Sec. 8.1. The methods used to perform the statistical interpretations are assessed in Sec. 8.2. Localized deviations from the SM expectation are observed for some of the probed signal hypotheses. However, once the "look elsewhere effect" is taken into account, as described in Sec. 8.2.1, the results are statistically compatible with the SM expectation. Exclusion limits are presented in Sec. 8.3. The results are summarized in Sec. 8.3.3. Finally, the findings are discussed in Chapter 9.

The work presented in this chapter has been developed by the author of this thesis under the supervision of Dr. Andreas Hinzmann, in collaboration with Dr. Clemens Lange and Dr. Jennifer Ngadiuba. Dr. Gerrit van Onsem has performed the goodness-of-fit test and the calculation of the test full CL_s *exclusion limits. All the other results, including the fit projections, the asymptotic* CL_s *exclusion limits and the significance estimation, were obtained by the author of this thesis*[1].

8.1 Multidimensional fit results

The signal extraction and background estimation procedure is described in Sec. 7.4. It relies on a simultaneous fit to the three-dimensional m_{jj}-m_{jet1}-m_{jet2} data distributions in the 10 event categories of the signal region introduced in Sec. 6.3, five tagging categories for each of the two resonance production categories (ggF/DY and VBF). During the fit, the background normalizations and shapes vary within the uncertainties controlled by the nuisance parameters described in Sec. 8.2.

A goodness-of-fit (gof) test is used to verify how compatible the data are with a hypothesis [192]. The *p-value*, i.e. the probability of obtaining a result as compatible or less with the hypothesis as the one observed, evaluates the agreement level between the predicted (by the model) and the measured number of events in each bin of the relevant distributions. Toy experiments are generated from the model in order to obtain the distribution of a test statistic reflecting the level of agreement between the observed and expected histograms. A likelihood ratio is used as test statistics t,

[1] Given the large amount of probed signals and long running time of the calculations, Dr. Jennifer Ngadiuba helped in the submission of the final asymptotic CL_s and significance calculations to reduce the total running time. During the whole optimization and validation phase the author took care of it.

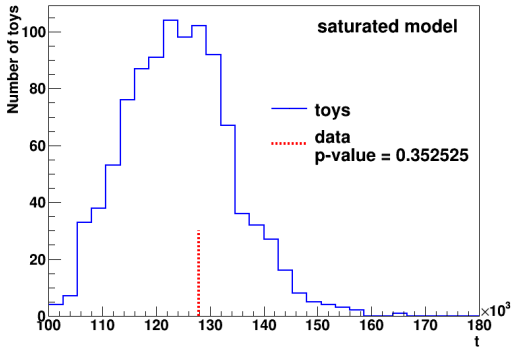


FIGURE 8.1: Goodness-of-fit tests using background-only pseudodata, with the saturated algorithm. The distribution of the test statistic t for the background-only toys is shown in blue, and the value of the test statistic for data is shown in red.

as described in Ref [193]. The nominator is

$$L = \prod_j \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-(d_j - f_j)^2/(2\sigma_j^2)}, \tag{8.1}$$

where $d_j \pm \sigma_j$ is the j-th measured data point with RMS deviation σ_j and f_j is the model prediction. The *saturated model* is introduced to provide an alternative hypothesis, i.e. a denominator to the ratio, where the model prediction is equal to the data at every measured value. Figure 8.1 shows the distribution of the test statistic t of the toys using the saturated algorithm. The signal strength is left floating when fitting the toys so that the measure is independent of the presence or absence of a signal. The value of the data t_0 is indicated with a red dashed vertical line. From the normalized toy distribution $f(t)$ a p-value can be calculated, as the integral $p = \int_{t=t_0}^{+\text{inf}} f(t)dt$. The measured p-value $p = 0.35$ indicates good compatibility between data and the background-only model.

The m_{jet1}, m_{jet2} and m_{jj} spectra in data for the whole phase space and all the 10 event categories are shown in Figures 8.2 and 8.3 for the high purity and low purity tagging categories of the ggF/DY region and in 8.4 for all VBF categories. The solid grey curves represent the results of the maximum likelihood fit to the data. The lower panels show the corresponding pull distributions, quantifying the agreement between the hypothesis of background only and the data. The resonant background components are shown separately. A signal is superimposed onto all three projections corresponding to a signal yield as expected from the theoretical prediction and the analysis selection efficiency, and scaled by the factor reported in the legend. The background yields in the signal region extracted from the background-only fit, together with their post-fit uncertainties, are summarised in Table 8.1 and compared with observations separately for all ten categories. The extracted cross sections of the resonant backgrounds ($t\bar{t}$ and V+jets) are compatible with the SM expectations (see Table 7.1) within 1–2 standard deviation of the post-fit uncertainties in the majority of the categories. In the prefit distributions, the W peak in the VH categories is underestimated and the fitting procedure, performing the data-driven background estimation, compensates it, as shown in Figures 8.2–8.4 where a good agreement is found.

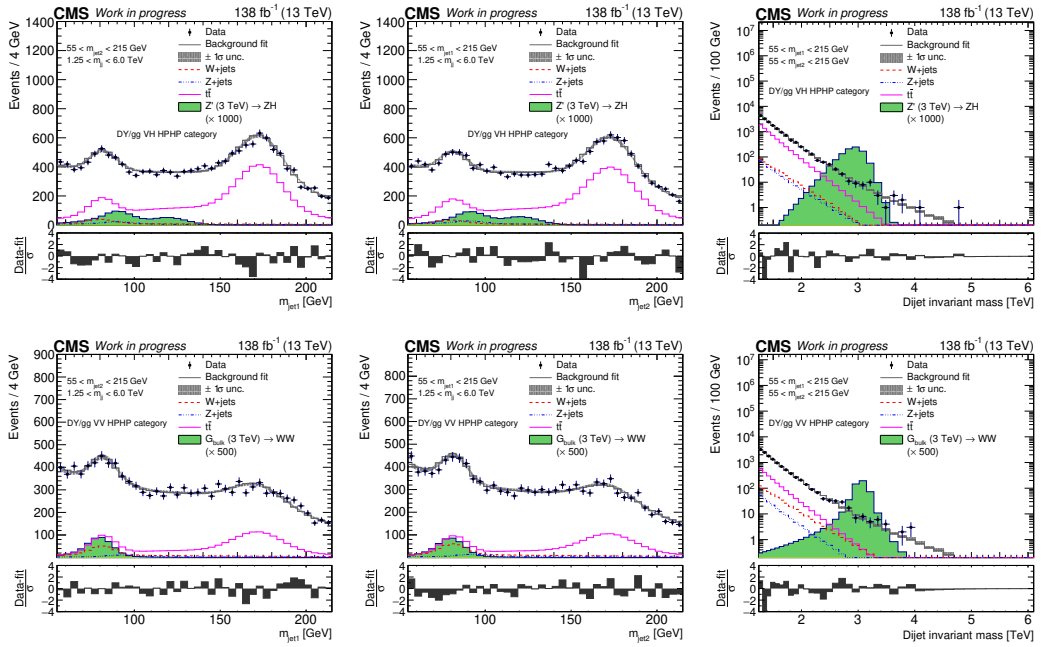


FIGURE 8.2: For the VH HPHP ggF/DY (top) and VV HPHP ggF/DY (bottom) categories: comparison between the background fit and the data distributions of m_{jet1} (left), m_{jet2} (middle) and m_{jj} (right). The background shape uncertainty is shown as a gray shaded band, and the statistical uncertainties of the data are shown as vertical bars. An example of a signal distribution is overlaid, where the number of expected events is scaled by the factor reported in the legend. Shown below each mass plot is the corresponding pull distribution (Data-fit)/σ, where $\sigma = \sqrt{\sigma^2_{\text{data}} - \sigma^2_{\text{fit}}}$ for each bin to ensure a Gaussian pull-distribution, as defined in Ref. [194].

TABLE 8.1: Observed yield and background yields extracted from the background-only fit together with post-fit uncertainties.

Category	Data	Total	QCD	$t\bar{t}$	W+jets	Z+jets
VBF VH HPHP	331	335 ± 22	190 ± 19	138 ± 12	4 ± 2	3 ± 1
VBF VV HPHP	204	211 ± 15	159 ± 14	41 ± 3	8 ± 3	3 ± 1
VBF VH LPHP	847	838 ± 36	656 ± 33	162 ± 15	9 ± 4	11 ± 5
VBF VH HPLP	599	605 ± 28	481 ± 26	111 ± 10	8 ± 4	5 ± 2
VBF VV HPLP	890	897 ± 33	819 ± 31	42 ± 5	25 ± 9	11 ± 4
DY/gg VH HPHP	16366	16400 ± 340	9190 ± 220	6580 ± 200	360 ± 120	280 ± 100
DY/gg VV HPHP	12316	12310 ± 230	9540 ± 170	2030 ± 70	530 ± 110	200 ± 70
DY/gg VH LPHP	40666	40600 ± 570	31330 ± 380	7780 ± 240	760 ± 240	730 ± 250
DY/gg VH HPLP	33629	33690 ± 380	27580 ± 280	5210 ± 170	580 ± 150	320 ± 120
DY/gg VV HPLP	53807	53810 ± 450	49490 ± 340	2180 ± 90	1690 ± 240	450 ± 130

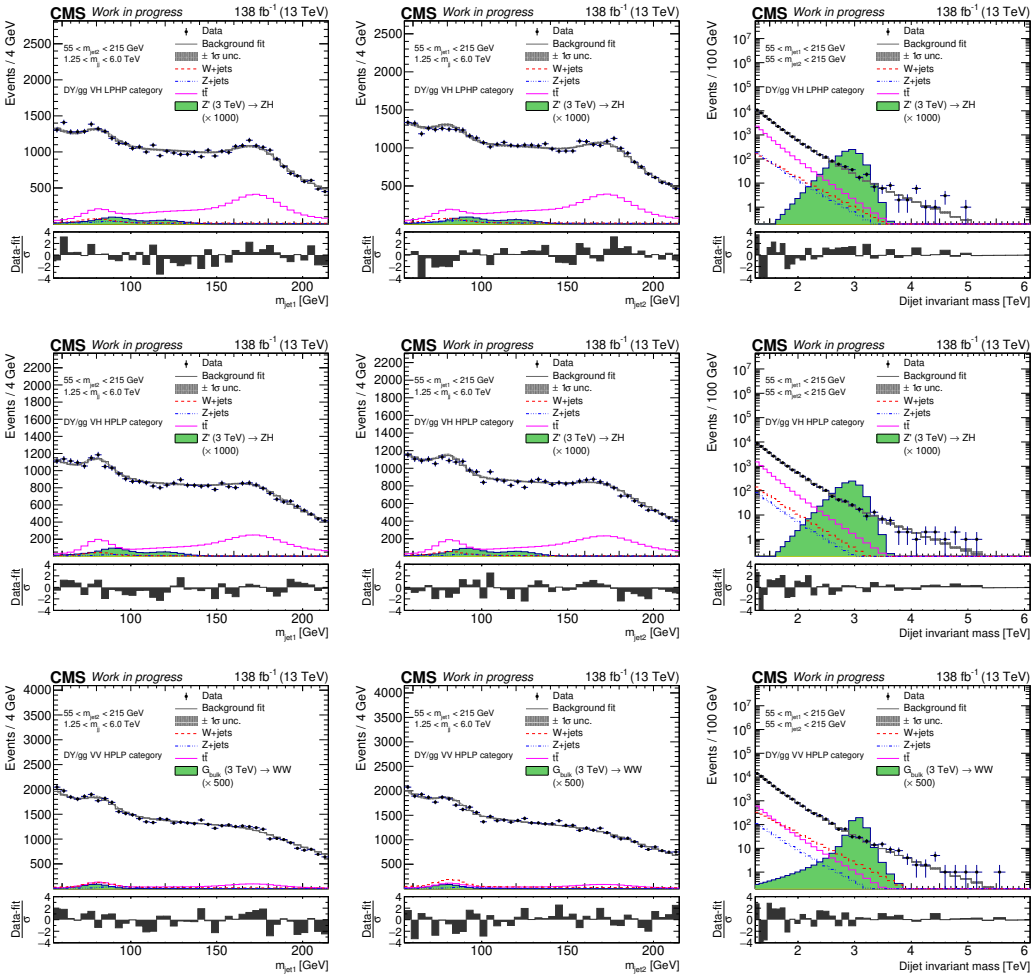


FIGURE 8.3: For the VH LPHP ggF/DY (top), VH HPLP ggF/DY (middle), and VV HPLP ggF/DY (bottom) categories: comparison between the background fit and the data distributions of m_{jet1} (left), m_{jet2} (middle) and m_{jj} (right). The background shape uncertainty is shown as a gray shaded band, and the statistical uncertainties of the data are shown as vertical bars. An example of a signal distribution is overlaid, where the number of expected events is scaled by the factor reported in the legend. Shown below each mass plot is the corresponding pull distribution.

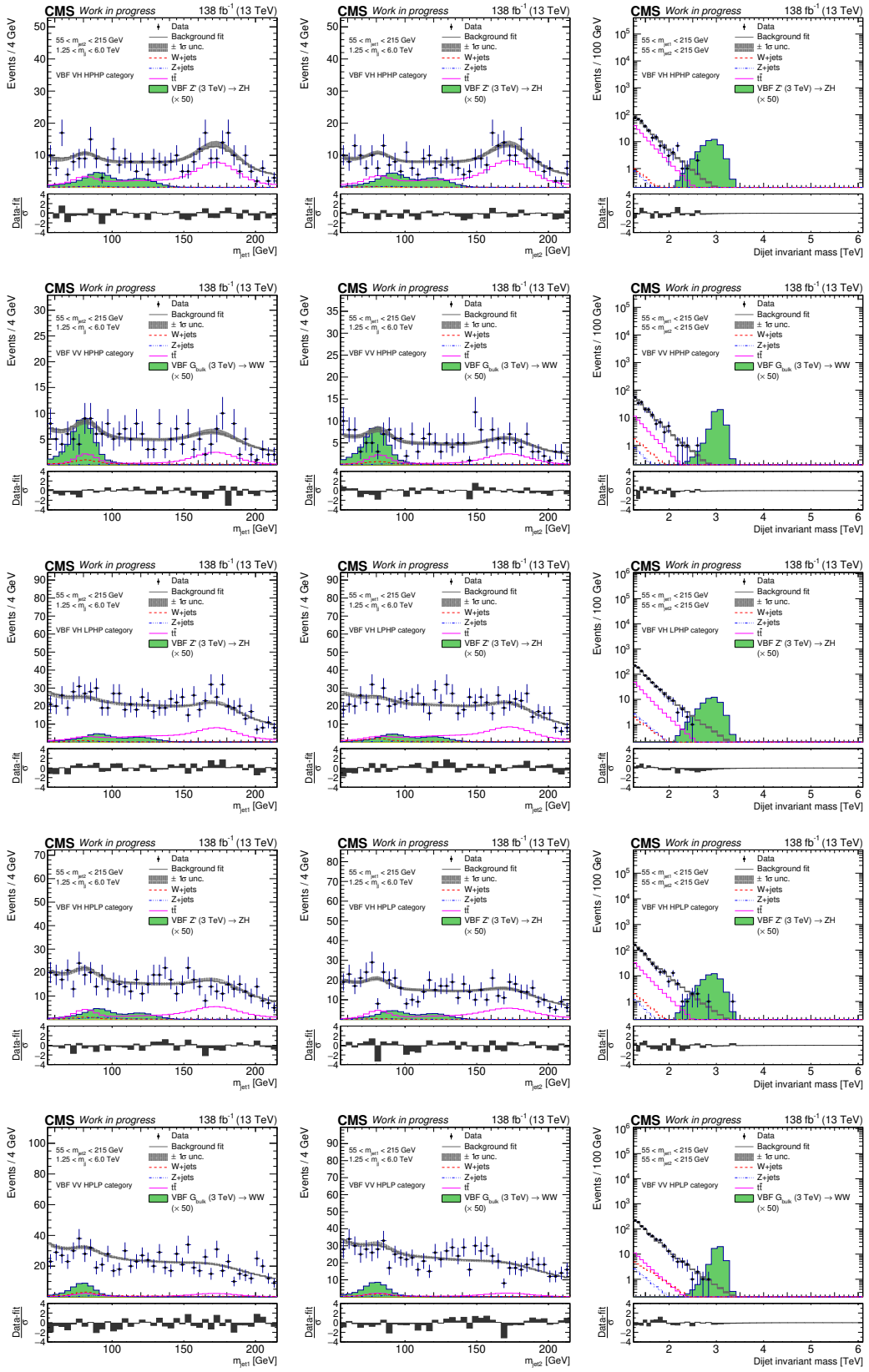


FIGURE 8.4: For, from top to bottom, the VH HPHP VBF, VV HPHP VBF, VH LPHP VBF, VH HPLP VBF, and VV HPLP VBF categories: comparison between the background fit and the data distributions of m_{jet1} (left), m_{jet2} (middle) and m_{jj} (right). The background shape uncertainty is shown as a gray shaded band, and the statistical uncertainties of the data are shown as vertical bars. An example of a signal distribution is overlaid, where the number of expected events is scaled by the factor reported in the legend. Shown below each mass plot is the corresponding pull distribution.

8.2 Statistical interpretation

Two hypotheses are compared: the *null-hypothesis* H_0 corresponding to the lack of signal contribution (*background only*, obtained for signal strength μ = 0) and the H_1 hypothesis corresponding to the alternative scenario of the presence of a signal (*signal+background*).

A *test statistic* q_μ is defined to compare the hypotheses in a quantitative and reproducible way. First, the profile likelihood ratio $\lambda(\mu)$ is introduced as [191]

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}_\mu)}{L(\hat{\mu}, \hat{\hat{\theta}})} \tag{8.2}$$

from the likelihood in Eq. 7.9. The *profile likelihood* $L(\mu, \hat{\hat{\theta}}_\mu)$ is maximized for a certain μ, i.e. is the conditional maximum-likelihood (ML) estimator of $\bar{\theta}$. The term *profile likelihood* refers to the procedure of refitting the nuisance parameters to maximize the likelihood for each possible value of the signal strength μ, the parameter of interest. The denominator, instead, is the maximized (unconditional) likelihood function, with $\hat{\mu}$ and $\hat{\hat{\theta}}$ the global maximum of the likelihood. The dependence on μ reflects the loss of information due to systematic uncertainties. The test statistic [191]

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \text{for } \hat{\mu} \leq \mu \\ 0 & \text{for } \hat{\mu} > \mu \end{cases} \tag{8.3}$$

is especially suited for setting upper limits on the signal strength. Higher values of q_μ correspond to increasing incompatibility between the data and μ. Data with $\hat{\mu} > \mu$ do not represent less compatibility with μ than the data obtained, thus, setting $q_\mu = 0$ for $\hat{\mu} > \mu$, they are not taken as part of the rejection region of the test. Given a μ hypothesis, the test statistic value q_μ^{obs} is measured on data and the nuisance parameters $\hat{\hat{\theta}}_\mu^{\text{obs}}$ and $\hat{\hat{\theta}}_0^{\text{obs}}$ are obtained for the signal+background and the background only hypothesis, respectively. The nuisance parameters are then used to build the probability density function of the test statistic for the signal+background and the background only hypothesis. Such pdfs are used to estimate the p-values associated to the two hypotheses

$$\begin{cases} p_\mu = P(q_\mu \geq q_\mu^{\text{obs}} | \mu s(\hat{\hat{\theta}}_\mu^{\text{obs}}) + b(\hat{\hat{\theta}}_\mu^{\text{obs}})) \equiv \text{CL}_{s+b} \\ p_0 = P(q_\mu \geq q_\mu^{\text{obs}} | b(\hat{\hat{\theta}}_0^{\text{obs}})) \equiv \text{CL}_b \end{cases} . \tag{8.4}$$

Finally, the

$$\text{CL}_s = \frac{\text{CL}_{s+b}}{1 - \text{CL}_b} \tag{8.5}$$

quantity can be introduced to assess how well the two hypotheses can be separated. If CL_s is lower than the a-priori confidence level α, a model with signal strength μ' is excluded at (1-α) confidence level (C.L.). As a convention in high energy physics, the 95% C.L. *observed* upper limit on the theoretical model is set by extracting μ' from the equation $\text{CL}_s = 0.05$. The *expected* upper limit is obtained from background only pseudo experiments. For each of the experiments, the μ' giving the 95% C.L. is calculated. The expected 1σ and 2σ uncertainty bands are the intervals containing 68% and 95% of toys distributed around μ'. Thus, expected limits can be used to estimate

the sensitivity of the designed analysis independently from the statistical fluctuations of the given data. However, the described method of calculating the *expected* values requires an extensive use of computing power, therefore the *asymptotic* calculation proposed in Ref. [191] is used in this analysis. The test statistic is replaced by

$$q_\mu = -2\ln\lambda(\mu) \approx \frac{(\mu-\hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N}) \tag{8.6}$$

where $\hat{\mu}$ follows a Gaussian distribution described by μ' and σ. N represents the data sample size and, thus, the term $\mathcal{O}(1/\sqrt{N})$ can be neglected for large data samples ($N \to \infty$). Furthermore, the evaluation of σ is greatly simplified considering, instead of a large amount of simulated experiment, the so-called *Asimov data set*, where all statistical fluctuations are suppressed and the estimators for all parameters are replaced by their expectation values. The use of the asymptotic variables simplifies the calculation of the upper limits. Upper limits with the full CL_s calculation have been evaluated for one signal hypothesis with masses of 3, 4 and 5 TeV and found to be well within the 1σ uncertainty band of the *asymptotic* calculation. Thus, the asymptotic approach provides a valid result also at dijet invariant masses of several TeV, where the number of events is limited.

8.2.1 Quantifying an excess of events and the look elsewhere effect

A different test statistic can be introduced to probe whether an excess of events is compatible with the discovery of a positive signal and thus with the rejection of the background-only hypothesis corresponding to $\mu = 0$. The test statistic [191]

$$q_0 = \begin{cases} -2\ln\lambda(0) & \text{for } \hat{\mu} \leq 0 \\ 0 & \text{for } \hat{\mu} > 0 \end{cases} \tag{8.7}$$

is used for this purpose, where $\lambda(0)$ is the likelihood ratio introduced in Eq. 8.2 for $\mu = 0$. With this definition, a disagreement between the data and the background-only hypothesis is considered only if $\hat{\mu} = 0$. If the event yield is above the background expectation, $\hat{\mu}$ increases and a larger incompatibility between the data and the background only hypothesis is observed. A p-value p_0 is introduced from the observed value q_0^{obs} of q_0:

$$p_0 = P(q_0 \geq q_0^{\text{obs}} | b(\hat{\hat{\theta}}_0^{\text{obs}})) = \int_{q_0^{\text{obs}}}^{+\infty} f(q_0|0,\hat{\hat{\theta}}_0^{\text{obs}}) dq_0, \tag{8.8}$$

with $f(q_0|0,\hat{\theta}_0^{\text{obs}})$ the pdf of the statistic q_0 under the assumption of the background-only hypothesis. A hypothesis is excluded if its p-value is observed below a specified threshold. In particle physics, the p-value is usually converted in *significance* Z. The significance Z is defined such that a Gaussian distributed variable found Z standard deviations above its mean has an upper-tail probability equal to p [191]

$$Z = \Phi^{-1}(1-p_0), \tag{8.9}$$

where Φ^{-1} is the quantile of the standard Gaussian distribution. Conventionally, a discovery is claimed if a significance of at least 5 standard deviations is found, while $Z = 3$ is the threshold for an evidence.

In searches like the one described in this thesis, a "bump" hunt is performed, i.e. the analysis looks for an excess of events in an unknown location of the phase

space under study. Thus, the signal distribution depends on an unknown location parameter, the resonance mass, while the background distribution is independent of it. The p-value defined in Eq. 8.8 is the *local* p-value for a certain resonance mass hypothesis. Since the resonance can appear anywhere in the probed mass range, the search looks for the largest excess of events above the background. In case of an excess, the *global significance* is estimated from the local significance, correcting for the so-called *look elsewhere effect* (LEE) [195], the probability to observe the same excess anywhere in the whole mass range. The look elsewhere effect can be approximately described in terms of *trial factors*, the ratio between the probability of observing the excess at some fixed mass point, to the probability of observing it anywhere in the range. In Ref. [195] it is shown that the trial factor (trial #) can be estimated as

$$\text{trial \#} = \frac{1}{3}\frac{\Delta_M}{\sigma_M} Z_{\text{local}} \tag{8.10}$$

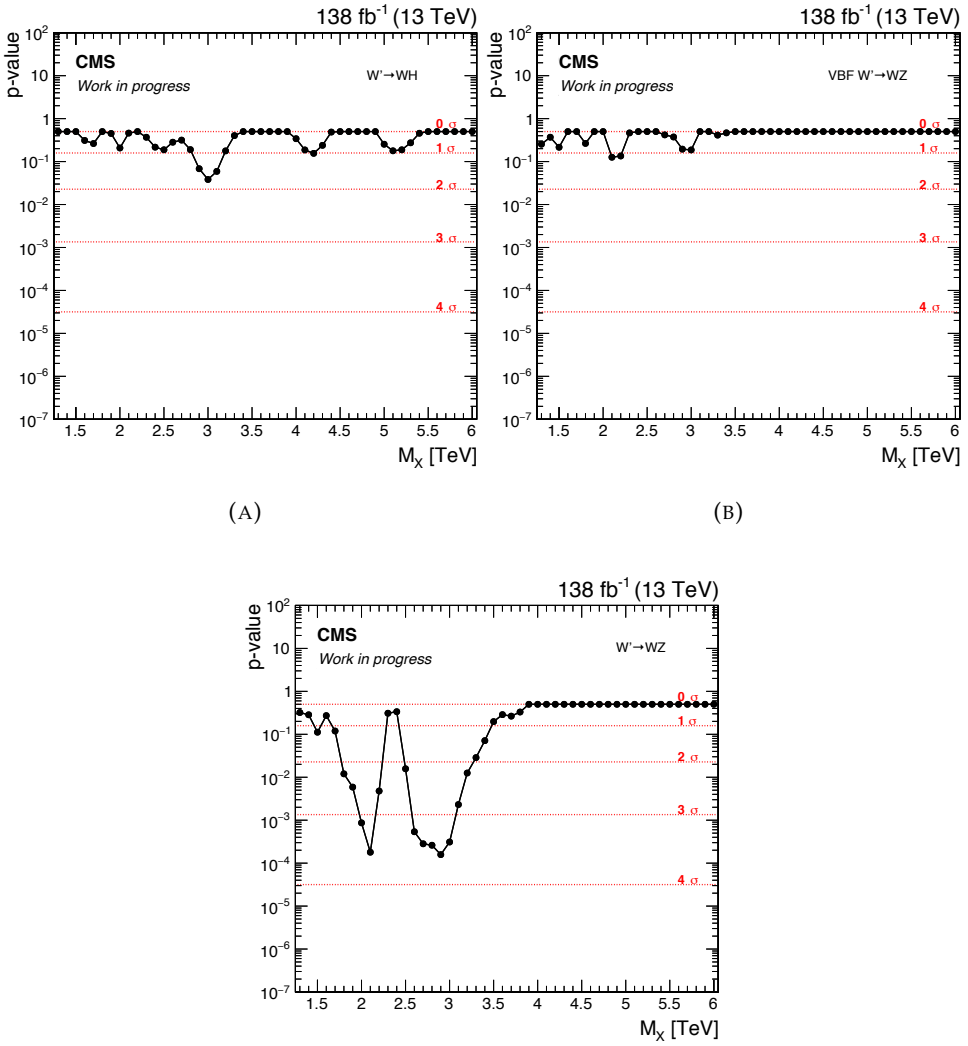
where Δ_M is the probed resonance mass range and σ_M the mass resolution. The global p-value can be roughly estimated as the product of the local p-value and the trial #.

8.3 Test for signal and exclusion limits

8.3.1 Observed significance

The significance of local deviations from the standard model background has been checked over the considered resonance mass range 1.3–6 TeV in 100 GeV intervals, the average mass resolution, for all signal hypotheses. Relevant local p-value examples are shown in Fig. 8.5, while for all other probed signal hypotheses the p-value can be found in Appendix D.1. For resonances decaying in the VH channel (Fig. 8.5a) and for VBF produced resonances (Fig. 8.5b) the significance is below 2 standard deviations (2σ) for all considered mass points. Figure 8.5c shows the local p-value in the VV channel where the two largest observed excesses, up to 3.6σ, are reported: the local significance is above 3 standard deviations at 2.1 TeV and 2.5–3 TeV. Various checks of the background estimation procedure have been performed to investigate the observed features. Using a dijet fit function for the partially resonant background with one parameter less than the one used in Eq. 7.6 would yield better modeling of the tails of the m_{jj} distribution, with a degraded description of the bulk of the spectrum, and thus larger χ^2/N_{dof}. The use of the fit function with reduced number of parameters would change the significance at 2.1 and 2.7 TeV by less than 0.5 standard deviations and of 1 standard deviation, respectively. Thus, for both excesses, a local significance well above 2 standard deviations would still be observed. Further tests have been carried out increasing the shape and normalization uncertainties of the partially resonant backgrounds or increasing the number of nuisance parameters describing the $\text{t}\bar{\text{t}}$ components, yielding a local significance always above 2σ at 2.1 TeV and above 3σ at 2.7 TeV.

Figure 8.6 presents an overview of the pulls between the data distribution and the model prediction in signal enriched regions for all the analysis categories as a function of the dijet invariant mass. The reported excesses can be seen in the upper left plot, where the masses of both final state AK8 jets are required to be in the W and Z bosons mass range. An under fluctuation of event is observed at 1.4 TeV in several categories. A part from the reported features, a good agreement between data and expectations is observed in the other regions.



(A) (B)

(C)

FIGURE 8.5: Local p-value distributions as a function of the resonance mass for the DY/VBF $W' \to WZ/H$ signal hypothesis. The red horizontal lines show the corresponding significance.

Since excesses are observed, the global p-value should be calculated. A first estimation is obtained as the product of the local p-value and the trial #, defined in Eq. 8.10. For this analysis trial # corresponds to 169.2, yielding a global significance of ≈2.3 standard deviations. A more precise calculation of the global p-value was obtained using background-only toy experiments and performing a likelihood scan, as described in the CMS internal Ref. [196]. Also in this case, a global significance of 2.3 standard deviations is found.

Thus, the results are statistically compatible with the SM expectation.

8.3.2 Observed exclusion limits

Exclusion limits are set on narrow resonances decaying to two vector bosons or one Higgs and one vector boson. The CL_s prescription evaluated using the asymptotic approximation, as described in Sec. 8.2, is performed to set 95% C.L. upper limits on the production cross section times branching fraction of BSM resonances.

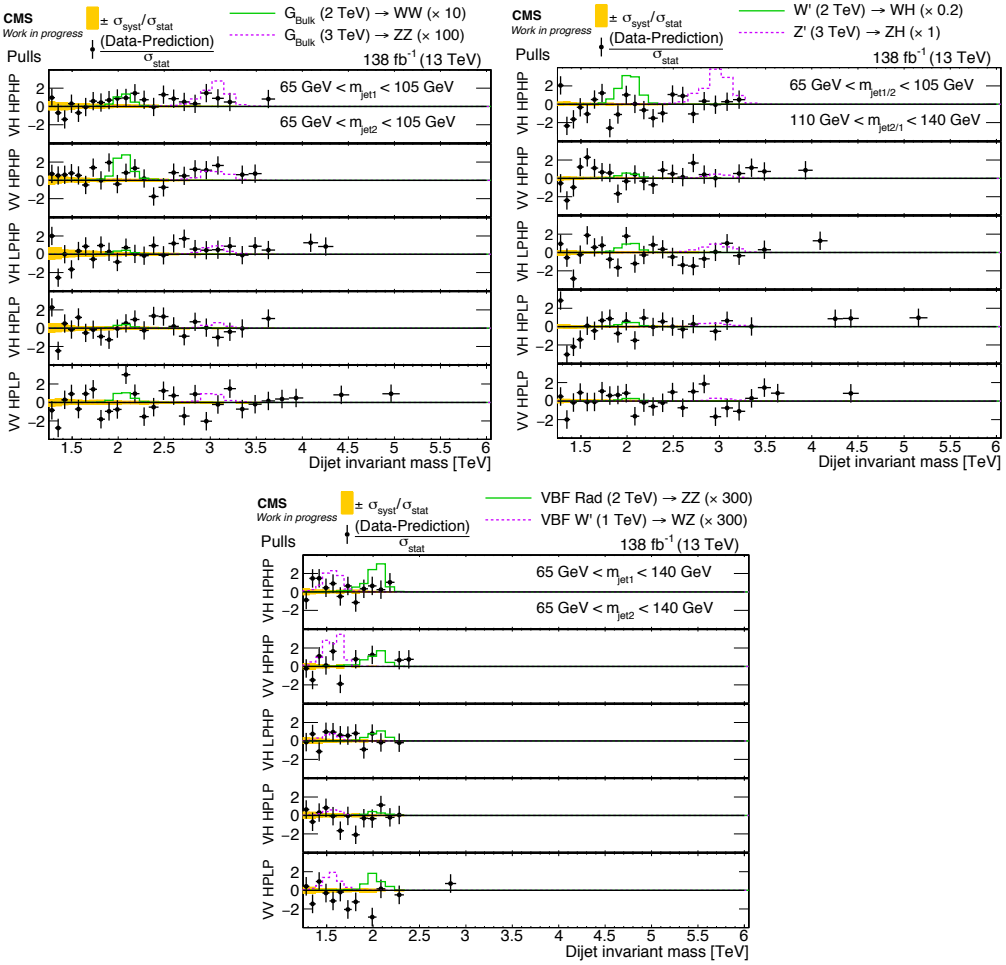


FIGURE 8.6: Pull distribution (Data-fit)/σ_{stat}. Three projections of the three dimensional phase space are shown in regions enriched in signal from (upper left) DY/gg VV ($65 < m_{\text{jet1}} < 105$, $65 < m_{\text{jet2}} < 105$), (upper right) DY/gg VH ($65 < m_{\text{jet1/2}} < 105$, $110 < m_{\text{jet2/1}} < 140$) and (lower) VBF VV/VH ($65 < m_{\text{jet1}} < 140$, $65 < m_{\text{jet2}} < 140$). Post-fit systematic uncertainties are indicated by the shaded bands. Exemplary signal points normalized to the theoretical prediction for the signal production cross section are shown using full and dashed lines.

A total of 16 different signal hypotheses (as listed in Table 6.1) is tested with the three-dimensional fitting approach described in this thesis. The fit is performed in the m_{jj}-m_{jet1}-m_{jet2} plane. Both VV and VH decay modes are covered, as well as different resonance production modes. Furthermore, limits combining VV and VH decay channels of a spin-1 resonance are set. The limits are interpreted in the context of the bulk graviton and radion model (see Sec. 2.3.1) and the HVT model B and C scenarios (see Sec. 2.3.2).

In the following, the observed limits for the different scenarios are compared with the expected cross sections from theoretical calculations. Here the limits combining different decay modes are shown, while limits separated for the 16 different signal hypotheses can be found in Appendix D. The results are finally summarized in Sec. 8.3.3.

Figure 8.7 shows the resulting 95% C.L. observed and expected exclusion limits on the signal cross section as a function of the resonance mass for a Radion$\rightarrow$ VV (top) and a $G_{\text{bulk}} \rightarrow$ VV (bottom) signals combining WW and ZZ decay modes.

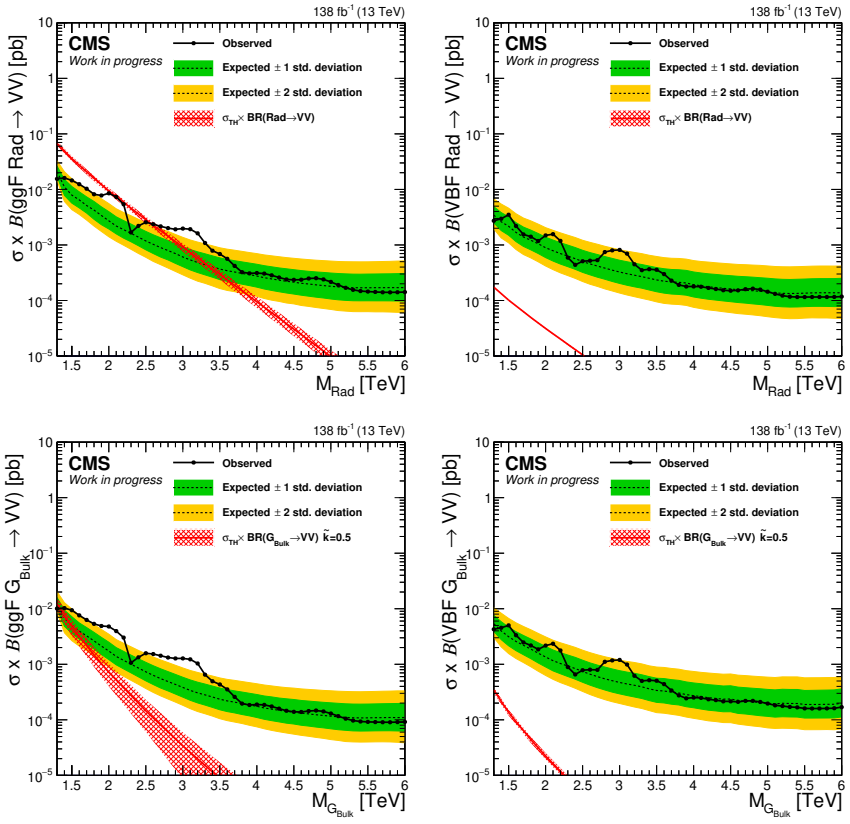


FIGURE 8.7: Observed and expected 95% C.L. upper limits on the product of the production cross section (σ) and the branching fraction B, obtained after combining all categories with 138 fb^{-1} of 13 TeV data for a Radion$\rightarrow$ VV (top) and a $G_{bulk} \rightarrow$ VV (bottom). On the left the ggF production mode is shown, while on the right the VBF one is presented.

For the ggF production modes, the excesses reported above are visible, while for the VBF production mode no excess is reported. For the ggF production mode, the Radion$\rightarrow$ VV ($G_{bulk} \rightarrow$ VV) can be excluded up to resonance masses of 2.7 TeV (1.4 TeV). Concerning the VBF production mode, upper limits on the production cross section are set in the range from 3 fb for a resonance mass of 1.3 TeV to 0.1 fb at 6 TeV for the Radion$\rightarrow$ VV and from 4 fb for a resonance mass of 1.3 TeV to 0.2 fb at 6 TeV for the $G_{bulk} \rightarrow$ VV.

The resulting 95% C.L. observed and expected exclusion limits on the signal cross section as a function of the resonance mass for spin-1 resonances combining different final states are shown in Fig. 8.8. For spin-1 resonances, excesses are observed only in the VV channel of the DY production mode, while for the VH channel or VBF production mode no excess is reported. An under fluctuation is observed at 1.4 TeV in the VH final state. For the DY production mode, resonances decaying to VV (VH) final states can be excluded up to resonance masses of 4.5 TeV (4.2 TeV). Combining the VV and VH final state, masses up to 4.8 TeV are excluded. For resonances produced through VBF, limits on the production cross section are set from 7–10 fb at 1.3 TeV to 0.3–0.4 fb at 6 TeV.

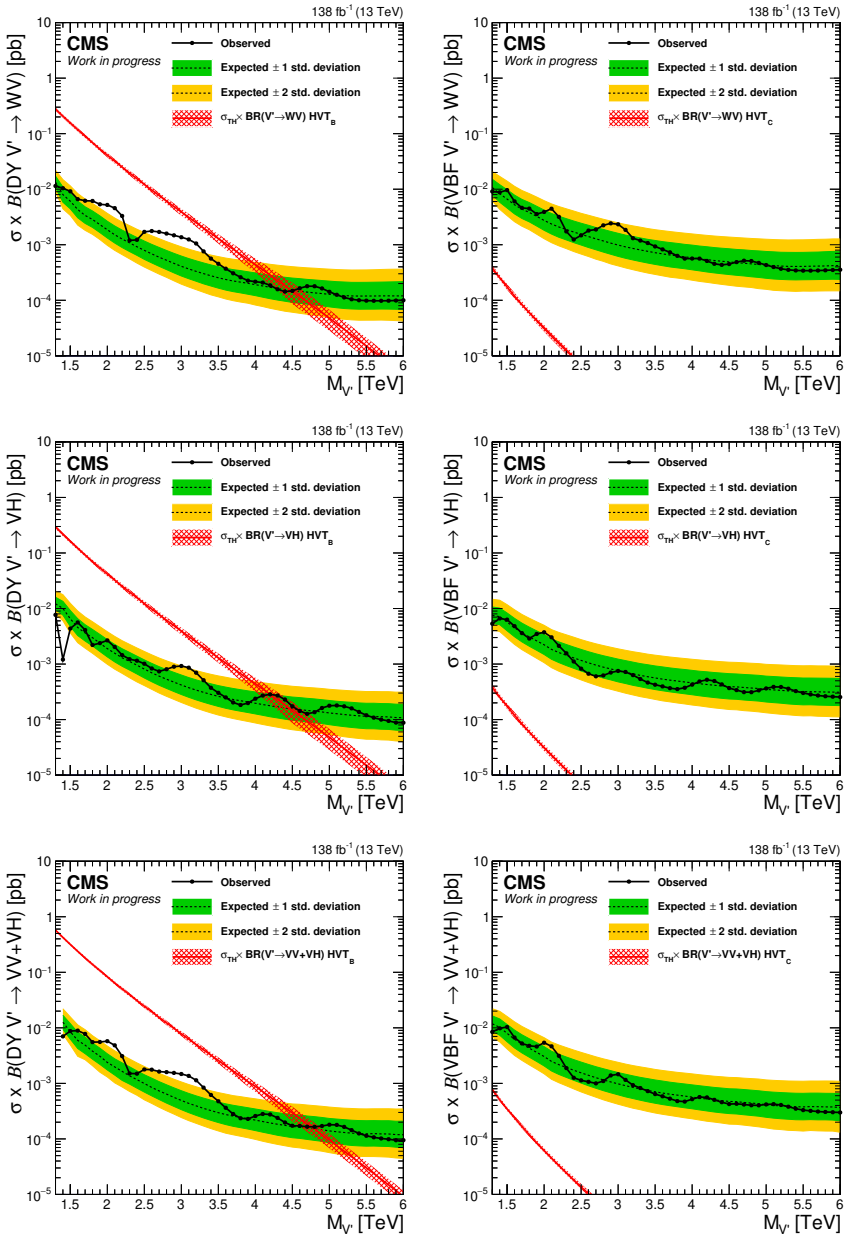


FIGURE 8.8: Observed and expected 95% C.L. upper limits on the product of the production cross section (σ) and the branching fraction B, obtained after combining all categories with 138 fb^{-1} of 13 TeV data for $V' \to WV$ (top), $V' \to VH$ (middle) and $V' \to VV + VH$ combining $V' \to WV$ and $V' \to VH$ (bottom). On the left the DY production mode is shown while on the right the VBF one is presented.

8.3.3 Summary

Limits are set on 16 different signal hypotheses and 10 combined channels.

In Fig. 8.9 the results obtained with the method described in this thesis are compared to the expected limits set in the previous version of this analysis [9, 177, 178] using the same 78 fb^{-1} dataset collected in the 2016–2017 data taking period. For the $G_{bulk} \rightarrow WW$ signal (left), this analysis performed with 78 fb^{-1} improves the sensitivity of more than 30% at higher resonance masses with respect to Refs. [9, 177, 178] thanks to the adoption of new ML-based taggers. A similar sensitivity is instead observed below resonance masses of ≈2 TeV, for the increase in $t\bar{t}$ background that is now picked due to the different tagger choices and extended phase space. Considering the full 138 fb^{-1} dataset, up to 60% improvement in sensitivity is observed with respect to Refs. [9, 177, 178]. On the right the $G_{bulk} \rightarrow ZZ$ signal is shown, where a larger improvement is observed thanks to the adoption of a dedicated $Z \rightarrow b\bar{b}$ tagger.

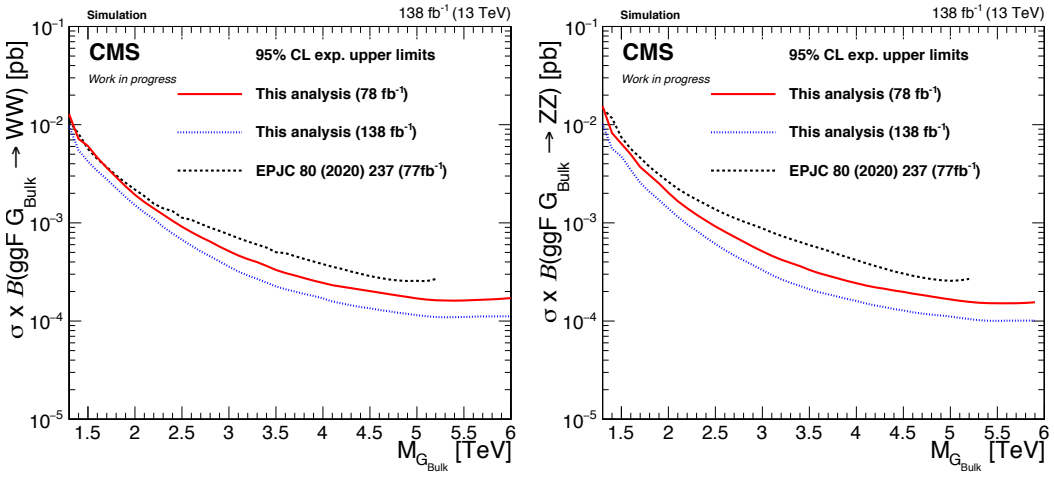


FIGURE 8.9: Expected 95% C.L. upper limits on the product of the production cross section (σ) and the branching fraction B for a $G_{bulk} \rightarrow WW$ (left) and a $G_{bulk} \rightarrow ZZ$ (right) signal with the method presented in this thesis using the same dataset (red solid line) of the previous version of this analysis [9, 177, 178] (black dash-dotted line). The final limit obtained when combining data collected in 2016–2018 is also shown (blue dashed line).

The presented analysis sets most stringent constraints on resonaces with masses above 2–3 TeV in all the combined channels and in signals where a Z or a H boson are considered in the final state. Table 8.2 summarizes the expected and observed lower exclusion limits on the resonance mass. The symbol "-" means no mass exclusion limit is set, but upper limits on the production cross section are established. Upper limits on the cross section of G_{bulk} ($\tilde{\kappa} = 0.5$[2]) produced through VBF combining WW and ZZ are set for the first time, as well as for VBF produced spin-1 resonances in the WH final state and combining different channels. The mass exclusion limit on $G_{bulk} \rightarrow VV$ is extended by 500 GeV with respect to the combination of different VV, VH, and HH decay channels using the 2016 data set. The mass exclusion limit on $V' \rightarrow VH$ is extended by almost 1 TeV compared to the previous result from the CMS collaboration based on 36 fb^{-1}.

[2] In Ref. [61] limits on VBF produced G_{bulk} are set for $\tilde{\kappa} = 1$. For this value of $\tilde{\kappa}$ a mass dependent width is considered and the resulting signal model cannot be compared to the narrow width approximation allowed by $\tilde{\kappa} = 0.5$ used in this thesis.

Model		Expected limit (TeV)		Observed limit (TeV)	
		In literature	This analysis	In literature	This analysis
Radion ggF	WW	3.1 [63]	3.2	3.1 [63]	1.3–1.98 and 2.15–2.46
Radion VBF	WW	- [61, 63]	-	-[61, 63]	-
Radion ggF	ZZ	2.5[62]	2.5	3.0 [62]	1.84
Radion VBF	ZZ	- [61]	-	- [61]	-
Radion ggF	VV	2.9 [61]	3.42	3.2 [61]	2.66
Radion VBF	VV	- [61]	-	- [61]	-
HVT model B W'	WZ / WH	4[61] / 3.8[63]	4.29/4.27	4.3[61] / 4 [63]	4.37/ 4.06
HVT model B Z'	WW / ZH	3.9 [63] / 3.7[64]	3.8/3.83	4 [63] / 3.7[64]	1.3–3.05 and 3.26–3.52 /3.91
HVT model B V'	VV / VH	3.5 [58]/ 3.1[59]	4.49/4.48	3.8[58] / 2.50 and 2.76–3.30 [59]	4.5/4.21
HVT model B V'	VV+VH	4.2 [53]	4.8	4.5 [53]	4.77
HVT model C W'	WZ / WH	- [61, 63] /	-/-	- [61, 63] /	-/-
HVT model C Z'	WW / ZH	- [61, 63] / - [64]	-/-	- [61, 63] / - [64]	-/-
HVT model C V'	VV/ VH	/	-/-	/	-/-
HVT model C V'	VV+VH		-/-		-/-
G_{bulk} $\tilde{\kappa} = 0.5$ ggF	WW	1.7 [63]	-	1.8 [63]	1.35
G_{bulk} $\tilde{\kappa} = 0.5$ VBF	WW	-[63]	-	- [63]	-
G_{bulk} $\tilde{\kappa} = 0.5$ ggF	ZZ	1.1 [62]	-	1.2 [62]	-
G_{bulk} $\tilde{\kappa} = 0.5$ VBF	ZZ	-[62]	-	- [62]	-
G_{bulk} $\tilde{\kappa} = 0.5$ ggF	VV	1.1 [53]	1.5	0.85 [53]	1.35
G_{bulk} $\tilde{\kappa} = 0.5$ VBF	VV		-		-

TABLE 8.2: Summary table comparing the lower limits on the mass of the signal hypotheses described in Sec. 2.3 already existing in literature and the ones set in this thesis. The symbol - indicates that no mass exclusion limit has been placed despite (at least) one search has been performed, placing limits on the production cross section; if more than one search has been executed, the most sensitive in terms of cross section are reported. If no analysis has been carried out in one of the channels, the channel is left empty.

9 | Discussion and Outlook

The first part of this thesis presents a search for resonances with masses above 1.3 TeV decaying to WW, WZ, ZZ, WH and ZH hadronic final states for a total of 16 different signal hypotheses. The data-driven background estimation and signal extraction procedures feature a maximum likelihood multidimensional fit to the m_{jj}-m_{jet1}-m_{jet2} plane, taking advantage of the fact that the signal is resonant in all three dimensions. This approach has been applied to VH final states for the first time in this thesis. Each of the two final state bosons decays in a large radius jet, and the candidate resonance is reconstructed from the resulting dijet system. The study employs machine learning based substructure techniques to distinguish jets produced in the hadronization of single quarks or gluons from jets originating from vector bosons or containing b-quark pairs. The analysis also targets different resonance production modes, with the introduction of categories dedicated to identifying resonances produced in association with two additional jets (vector boson fusion). This work analyzes data corresponding to a total integrated luminosity of 138 fb^{-1} recorded by the CMS detector at a centre-of-mass energy of 13 TeV.

The exclusion limits presented in this search are similar or more stringent with respect to existing results. The framework where the analysis has been developed is especially suited for the combinations of different decay channels, thus ten channel combinations have been performed, for the first time or probing resonances with cross sections up to an order of magnitude smaller than existing results.

In Fig. 9.1 the expected exclusion limits for a $\text{W}' \rightarrow \text{WH}$ are taken as an example to compare the results obtained here with other recent ones. The limit obtained with the method presented in this thesis for a total integrated luminosity of 138 fb^{-1} (red solid line) is compared to the latest result from the CMS collaboration in the all hadronic final state and a total integrated luminosity of 36 fb^{-1} (dot-dashed purple line) [59]. The improvement in sensitivity is up to 40% larger than the one achievable from only increasing the dataset, thanks to the multidimensional fitting method and more performing tagging algorithms. Moreover, up to 70% improvement is obtained with respect to results obtained by the ATLAS Collaboration in a similar search for WH resonances decaying in the all hadronic final state [60] (solid purple line) on the dataset recorded in 2016–2018 by the ATLAS detector. A recent result from the CMS collaboration in the W(lep)H(had) final state for 138 fb^{-1} [63] is also shown (dotted green line). While the semileptonic final state provides the best sensitivity at lower resonance masses, above ≈ 1.8 TeV this thesis's search achieves the best to date expected limits. Moreover, since the search presented in this thesis is orthogonal by construction to the search in Ref. [63], the two can be combined to obtain the best sensitivity over the full mass range.

In 2022, the LHC Run-3 data taking period will start and 300 fb^{-1} of data will be delivered to the experiments, allowing to probe whether the observed excesses are statistical fluctuations or would, instead, become a significant deviation from

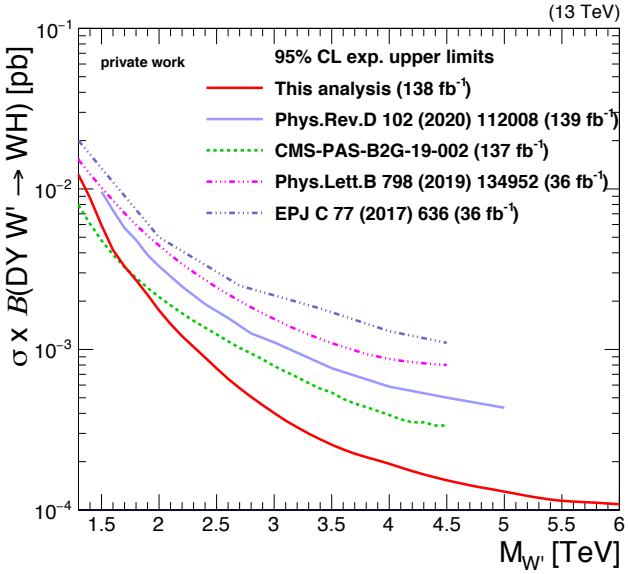


FIGURE 9.1: Expected 95% C.L. upper limits on the product of the production cross section (σ) and the branching fraction B for a $W' \rightarrow WH$ with the method presented in this thesis (red solid line) is compared to the latest result from the CMS collaboration for $W' \rightarrow WH$ in the all hadronic final state and a total integrated luminosity of 36 fb^{-1} (dot-dashed purple line) [59], the combination of different WH final states and a total integrated luminosity of 36 fb^{-1} (dot-dashed magenta line) [53], results obtained by the ATLAS Collaboration in a similar search for WH resonances decaying in the all hadronic final state(solid purple line) [60] and a recent result from the CMS collaboration in the W(lep)H(had) final state (dotted green line) [63].

the SM expectation. For this purpose, a search optimized for resonances decaying to two vector bosons could provide the best setup. Moreover, the larger dataset and the possible rise of the center-of-mass energy to 14 TeV will increase the number of events in the phase space corresponding to high resonance masses.

The approach described in this thesis has the advantage of avoiding the introduction of jet mass categories allowing to look for several decay modes in a single search, greatly simplifying their combination. The inclusion of HH decay mode is, for instance, straightforward and, performing an optimization for such decay mode, it can reach the sensitivity of dedicated searches. The fit to the whole jet mass spectrum enables also the possibility of looking for final states with top quarks or even more exotic possibilities. The resonances could decay to non-SM particles with an unknown mass. The approach described in this thesis could be generalized to probe this scenario. The presented search was carried out under the assumption that the new resonances' width is smaller than the detector energy resolution. The described approach could be further developed to consider also larger signal widths, resulting in a broad resonance. These effects are predicted by an increasing number of theoretical models [197] and are especially relevant when considering non-SM decay channels.

The use of machine learning taggers has greatly improved the sensitivity of the presented search. However, a dedicated decorrelation method has been developed and validated to employ them in this multidimensional fitting approach. New taggers [198, 199] are now becoming available that could simplify this task or yield even better performance.

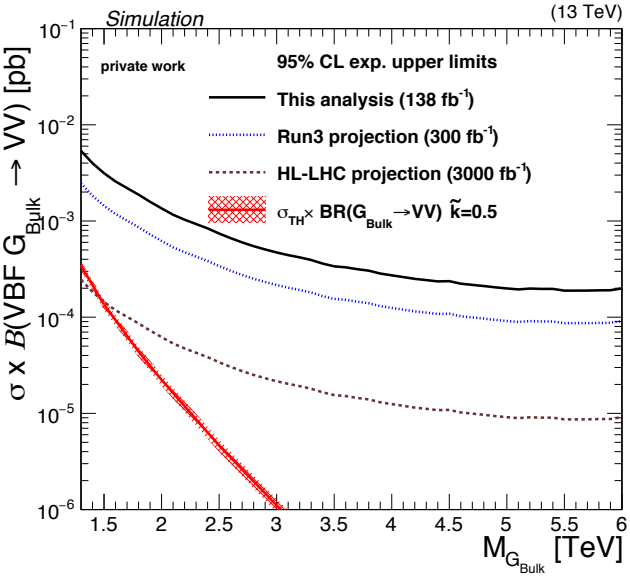


FIGURE 9.2: Expected 95% C.L. upper limits on the product of the production cross section (σ) and the branching fraction B for a VBF produced $G_{bulk} \rightarrow VV$ with the method presented in this thesis for a total integrated luminosity of 138 fb^{-1} (black solid line), rescaled to a luminosity of 300 fb^{-1} expected at the end of the Run-3 data taking (blu dottted line) and rescaled to 3000 fb^{-1}, the minimum luminosity expected at the end of the HL-LHC data taking (dashed brown line). For these projections no further improvements of the analysis strategy and systematic uncertainties were assumed and the center-of-mass energy was kept at 13 TeV.

Furthermore, the LHC will be upgraded to increase the delivered luminosity, High-Lumi LHC, and the experiments will collect up to 4000 fb^{-1} of data in this phase at a centre-of-mass energy of 14 TeV. At such energies, the production cross section of a 3 TeV (6 TeV) graviton is 1.5 (2.5) times higher than at $\sqrt{s} = 13$ TeV [43]. Figure 9.2 shows a comparison of the analysis presented in this thesis to projections obtained by rescaling the expected limits to the expected Run-3 and HL-LHC luminosities. No further improvements of the analysis strategy and systematic uncertainties were assumed and the center-of-mass energy was kept at 13 TeV. However, it shows that the large HL-LHC dataset will allow setting mass exclusion limits on VBF produced resonances.

As described in the second part of this thesis, the detectors will be upgraded to operate at the HL-LHC. The tracker will feature a finer segmentation, allowing better reconstruction of high momentum jets and thus increasing the ability to identify resonances as the one probed in this search.

Moreover, searches for new physics have just started to study VBF produced resonances as they have to face several challenges. For instance, spin-dependent polarization effects cause different kinematic distributions for different signal hypotheses and the selection with a traditional cut based approach leads to inefficiencies for some of them. Thus, while alternative strategies could be explored, a great benefit will come from larger datasets. The characteristic signature of VBF produced resonances is the presence of additional jets in the forward part of the detector. An essential aspect of the upgrade of the CMS detector for the HL-LHC period is the improvement of that detector section. The tracker will cover an extended η region

and several sub-detectors will feature an increased granularity, facilitating the identification of VBF jets.

While new BSM particles have not been discovered (yet), they may still exist at the TeV scale, and the collected or soon available data must be carefully analyzed to probe their existence and perform more precise standard model measurements. New jet tagging algorithms are being developed and pileup mitigation techniques are constantly being improved to increase the reach and the measurement precision beyond the gain given by larger available datasets. Detector technologies are also evolving to ensure high-quality data in future data taking periods. The second part of this thesis is dedicated to the study of new sensors for the upgrade of the CMS detector at HL-LHC.

Part II

Upgrade of the tracking detector for HL-LHC

10 HL-LHC and the upgrade of the CMS detector

At the end of 2024, the LHC will complete its course, after having delivered $\approx$300 fb^{-1} of data to the CMS detector. Several major upgrades are planned for the High Luminosity (HL) - LHC [10], extending the machine's discovery potential, ensuring another decade of operability and delivering a tenfold increase in the integrated luminosity with respect to the current LHC design [200].

In Sec. 10.1, the most important research areas that could be explored at the HL-LHC are explained, then the key aspects of the upgrade to the accelerator complex are outlined (Sec. 10.2). In Sec. 10.3, the upgrade of the CMS detectors is described. The upgrade of the tracker system is especially relevant for this part of the thesis and thus more details are provided in Sec. 10.4.

10.1 Physics motivation

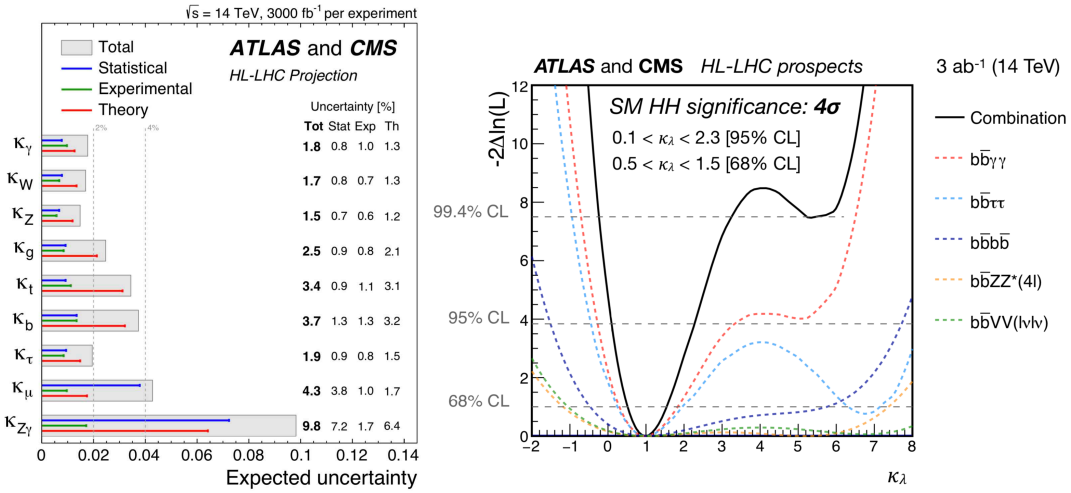
Observations made by experiments at the LHC have largely confirmed the predictions of the SM over a dynamic energy range extending up to the energy of 13 TeV. These early results have opened the way for future studies at the HL-LHC, which will include several precision SM and flavour measurements, extensive searches for new phenomena, and a determination of a myriad of Higgs boson properties [201, 202].

Several analysis efforts will profit from a decrease in statistical uncertainty owing to the 3000 fb^{-1} integrated luminosity expected at the HL-LHC. In particular, searches for rare phenomena and new physics will directly increase their sensitivity as a result of the larger dataset. Analyses whose sensitivity is limited by systematic uncertainty will have the possibility to apply more stringent requirements which lead to purer samples and larger signal-to-background ratios, driving sensitivity beyond the current limits. In the following, a brief overview of the primary objectives of the HL-LHC physics programme is given.

10.1.1 Properties of the Higgs boson

After the discovery in 2012, our knowledge of the Higgs boson has increased and several of the expected decay modes have been observed. The HL-LHC will produce 170 million Higgs bosons and 120 thousand Higgs boson pairs, extending the foreseen LHC dataset by a factor 10, increasing the precision of the Higgs coupling measurements.

So far, the Higgs boson would be the first fundamental scalar particle ever observed. However, it remains possible that the discovered particle is composite rather



(A) (B)

FIGURE 10.1: Left: Expected uncertainties on the coupling modifier parameters κ_i for the combination of ATLAS and CMS extrapolations. For each measurement, the total uncertainty is indicated by a grey box while the statistical, experimental and theory uncertainties are indicated by a blue, green and red line respectively. Right: Minimum negative-log-likelihood as a function of κ_λ, calculated by performing a conditional signal+background fit to the background and SM signal. The coloured dashed lines correspond to the combined ATLAS and CMS results by channel, and the black line to their combination. Both taken from Ref. [203].

than fundamental and several SM extensions are predicting it (see Sec. 2.3.2). A composite Higgs would result in a uniform decrease of all the Higgs couplings. To pin down the true nature of the Higgs boson, a precise framework is used in measuring the Higgs couplings, known as the κ-framework, where a set of κ_i factors linearly modifies the couplings of the Higgs boson to SM particles (i). Namely, all κ_i are equal to unity in the SM. As shown in Fig. 10.1a, percent level precisions on the measurements of the main couplings are expected at the HL-LHC, which will give evidence for the scalar or composite nature of the Higgs boson.

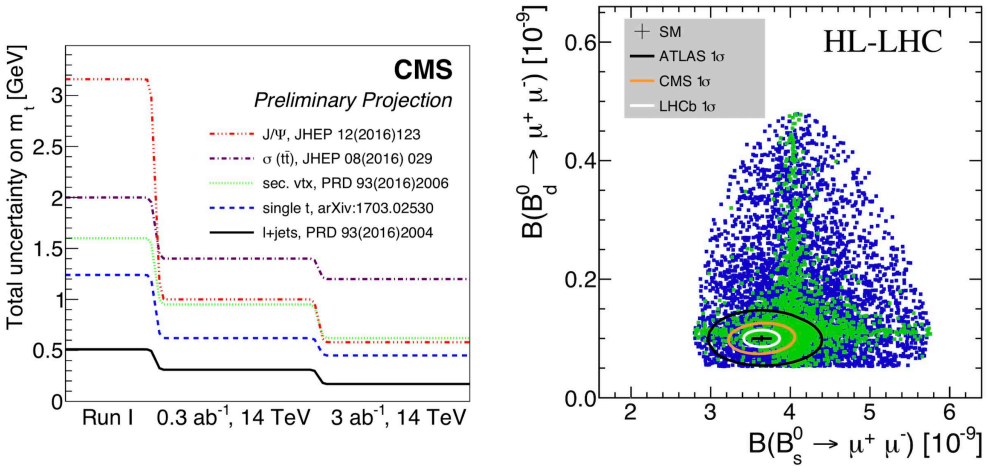
Of special importance is the Higgs self-coupling κ_λ. Measurements of this quantity could indicate the shape of the Higgs potential around its minimum, allowing the verification of the SM electroweak symmetry breaking. A scan of the likelihood as a function of the κ_λ coupling is presented in Fig. 10.1b combining ATLAS and CMS expectations. At the HL-LHC a second minimum of the likelihood would be excluded at 99.4% C.L..

Complementary to precision measurements, direct searches for exotic decays of the SM 125 GeV Higgs boson and of additional neutral or charged Higgs bosons production will enhance the sensitivity to the BSM sector.

10.1.2 Standard Model

At the HL-LHC, thanks to the 3000 fb^{-1} data set, upgraded detectors and improvements in the theoretical understanding, the SM will be probed with great precision at the unprecedented center-of-mass energy of 14 TeV, extending the sensitivity to possible anomalies [204].

In addition to the measurements of the Higgs self-coupling, the EW symmetry



(A) (B)

FIGURE 10.2: Left: The top mass measurement uncertainty for different methods as a function of integrated luminosity. Taken from Ref. [204]. Right: BR($B_s^0 \to \mu^+\mu^-$) vs. BR($B_d^0 \to \mu^+\mu^-$) in the SM (black mark), and in sets of BSM models with FCNC interactions consistent with current data (green points) and randomly generated (blue points). The coloured contours show the expected 1 standard deviation HL-LHC sensitivity of ATLAS, CMS, and LHCb. Taken from Ref. [205].

breaking mechanism can be tested through measurements of pairs or triplets of vector bosons. With the HL-LHC statistics and the detector improvements in the forward region (see Sec. 10.3), this sector will be fully accessible and the presence of anomalous EW couplings could be observed.

The physics program will considerably benefit from enhanced sensitivity to the high-energy tails of kinematic distributions, whose tests are not possible with the existing facilities. These measurements are formulated in the Effective Field Theories (EFTs) framework, which adds dimension-6 operators to the SM Lagrangian, parametrising BSM effects and how the SM processes are modified. Especially relevant, the dimension-6 operators are foreseen to grow quadratically with energy, thus quadratically enhancing a possible signal.

The top quark is the heaviest particle measured so far and has coupling to the Higgs of the order of 1, differently to the other SM fermions. Its mass is of crucial importance in particle physics for its central role in the stability of the SM vacuum and in the EW precision tests. As visible in Fig. 10.2a, the uncertainty on the top mass will be greatly reduced at the HL-LHC in several channels.

Since processes with flavour-changing neutral currents (FCNC) are forbidden at tree level and heavily suppressed in higher orders, searches for top FCNC decays such $t \to c\gamma, cZ, cg$ are typically statistically limited and would greatly benefit from the enhanced HL-LHC dataset, possibly clarifying why the top Yukawa coupling is the largest one or why there are large hierarchies in fermion masses.

10.1.3 Searches for new phenomena

The search for BSM phenomena is one of the crucial reasons to continue hadron collider physics programs [206]. Many theoretical and phenomenological aspects, for instance the hierarchy problem introduced in Sec. 2.3, suggest that new physics is close to the EW scale and the absence of evidence so far may imply that new

phenomena are more difficult to detect or that they are, surprisingly, not where we expect them. New phenomena may manifest at slightly higher masses, with lower couplings or more compressed spectra, at the edge of the discovery potential of the LHC. The HL-LHC will cover a broad range of scenarios of BSM physics, setting benchmarks to be used in a potential future machine.

For instance, the kinematic reach for supersymmetry (SUSY) searches will be extended, improving the measurement for compressed SUSY spectra, theoretically well motivated but barely covered by Run 2 analyses for their challenges. Such scenarios and DM signatures with mono-jets, mono-photons and VBF production will be studied at the HL-LHC with increased precision especially in the forward region. The HL-LHC will also cover a considerable part of the phase space where heavy resonances, suggested by the apparent flavour anomalies in B decays, could be. The detector upgrades that will be discussed in Sec. 10.3 will also improve the reach of searches for long lived particles. As an example, new fast timing detectors will be able to detect displaced photon signatures arising from long lived particles in the $0.1 < c\tau < 300$ cm range.

Furthermore, as it was already explained in Sec. 9, the analysis presented in the first part of this thesis work will notably benefit from the data-taking at the HL-LHC. A larger amount of statistics will increase the precision of the search and the background templates derivation. The higher center-of-mass energy will also populate the tails of the dijet mass distributions extending the reach of the search and increase the resonances production cross section. Additionally, upgraded detectors in the forward region will greatly improve the identification and reconstruction of the VBF jets. Further improvements in the identification of boson originating jets are presented in Chapter 11.

10.1.4 Flavour physics

The successful research program of the LHCb experiment has shown the potential of the LHC for flavour physics. The B-physics capabilities will be further increased by the LHCb, ATLAS and CMS upgrades, reaching unprecedented precisions and new flavour observables [205].

In the SM, different generations quarks are mixed by the weak charged-current transitions, encoded in the Cabibbo-Kobayashi-Maskawa (CKM) matrix (Eq. 2.3), whose elements must be measured experimentally. The relations among the elements, to assure the unitary nature of the CKM matrix and respect SM assumptions, are often expressed graphically in the complex plane as the so-called unitarity triangle. A powerful way to probe for virtual BSM effects, at mass scales complementary or superior to those which can be directly searched for at the HL-LHC, consists in overconstraining the apex of this triangle from tree- and loop-level quark mixing processes. At the HL-LHC, experimental or theoretical systematics will not be any more a limiting factor for a large variety of these tests.

One of the most sensitive probes for new physics is provided by the FCNC transitions $b \to s(d)\ell^+\ell^+$. The combination of the ATLAS, CMS and LHCb results at the HL-LHC will allow for a dramatically more precise measurements of these transitions, and shine a power light on various plausible new physics scenarios. Figure 10.2b shows the expected result of such measurements.

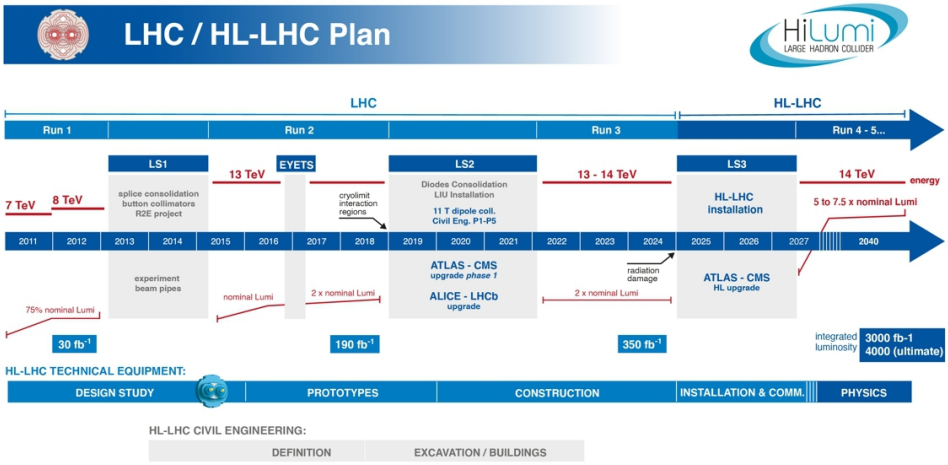


FIGURE 10.3: Schedule of the LHC and HL-LHC operations. Taken from Ref. [207].

10.2 High Luminosity - LHC

To fully exploit the discovery potential of the LHC, the HL-LHC major upgrade is planned in the 2020s, extending the machine operability by another decade and increasing the luminosity by a factor of five beyond its design value.

The planned schedule for the HL-LHC is reported in Fig. 10.3.

The design targets of the HL-LHC are:

- peak luminosity of $5 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ with levelling
- integrated luminosity of $250\,\mathrm{fb}^{-1}$ / year ($3000\,\mathrm{fb}^{-1}$ in ≈ 12 years)
- pileup in the CMS detector to up to 140.

An update of the long-term LHC schedule reduced the running time of the HL-LHC. Nevertheless, since all equipment is being designed with a 50% margin, the nominal goal of $250\,\mathrm{fb}^{-1}$ / year can be exceeded by pushing the HL-LHC targets to the ultimate scenario of:

- peak luminosity of $7\text{–}7.5 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ with levelling
- integrated luminosity of $300\text{–}350\,\mathrm{fb}^{-1}$ / year (total $4000\,\mathrm{fb}^{-1}$)
- pileup in the CMS detector to up to 200.

Exceptional technological advances are needed to achieve these objectives: state-of-the-art 11-12 T superconducting magnets, compact superconducting cavities with precise phase control, new beam collimation technology and high-power superconducting links with almost zero energy dissipation.

These impressive peak luminosities will be achieved thanks to several improvements, for instance to the quadrupole and dipole magnets, and to new radio-frequency crab cavities. The new crab cavities are made of bulk niobium that will be cooled at 2 K to reach the super conducting regime. They have a dedicated shape to tilt the proton bunches in each beam, maximising the overlap at the collision point: Each proton of one bunch will pass through the full length of the opposite bunch, increasing the collision probability.

10.3 Phase-2 upgrade of the CMS detectors

Under the HL-LHC operating conditions event pileup and radiation dose will rise substantially, becoming major challenges for the experiments. To work in such a harsh environment, the detectors of the main LHC experiments will be upgraded. In this section, details of the main challenges for the CMS detectors and key points of the upgrade will be given. In Sec. 10.4 the upgrade of the tracking system is treated in more detail.

The main purpose of the Phase-2 Upgrade is to maintain or improve the performance of the CMS detector, granting excellent efficiency, resolution, and background rejection for all physics objects [208].

10.3.1 Challenges at the HL-LHC

As the number of bunches in each proton beam is fixed, to achieve the impressive target luminosity, the number of interactions in each bunch crossing is increased to up to 200, which is ≈ 5 times higher than the average Run 2 PU presented in Fig. 3.1. The majority of these interactions are *soft*, thus not contributing to the *hard* process relevant for SM or new physics studies. The effect is instead to degrade the triggers and the reconstruction. In the tracking detectors more hits may result in mismeasured or misidentified tracks. In the calorimeters, deposits from PU may add extra energy to the measurements or may be associated to otherwise isolated leptons, deteriorating the sensitivity of the analyses.

The particles emerging from the collisions induce radioactivity in the detector materials and readout electronics, causing radiation damage and potentially progressively degrading the detection performance: It is foreseen that in one single year the dose delivered to the experiments by HL-LHC will be similar to the total dose delivered by the LHC.

Fig. 10.4 presents an overview of the predicted absorbed dose by the different CMS components after the full HL-LHC life time. The main source of radiation are the particles produced in the pp-collisions, therefore detectors close to the interaction point will receive the largest dose. Furthermore, detectors located close to the beam-pipe, as in the forward region, will receive a considerable dose.

In the collisions different types of particles are produced. Charged particles, mostly pions, interact through ionization with the detector material. Photons may produce e^+e^- pair when interacting with the beam pipe or the tracking system, or start electromagnetic cascades in the calorimeters. Particles backscattered from the calorimeters or escaped from the cascades interact with other detector components. Neutrons, for instance, scatter several times, producing more photons and electrons and generating a relatively uniform background in the detector volume. Sub-detectors are differently affected by radiation damage. The main impact on calorimeters is the loss of transmission of scintillation light through the material, reducing the signal of more than 90% in some cases. Radiation damage in silicon tracking systems will be extensively discussed in Sec. 12.5.

10.3.2 The Phase-2 upgrade of the CMS sub-detectors

In order to maintain excellent performance at the HL-LHC, different changes are planned for the Phase-2 upgrade of the CMS sub-detectors. For instance, tracker and endcap calorimeters will be completely replaced while additional chambers will be installed in the muon system. The key features of the sub-systems upgrade are

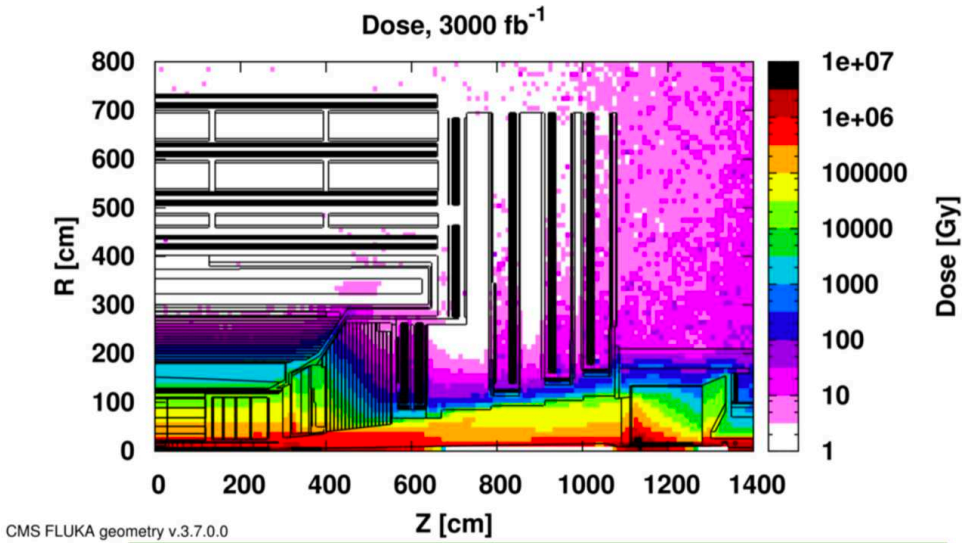


FIGURE 10.4: Predicted absorbed dose in the CMS cavern after an integrated luminosity of 3000 fb^{-1}. Taken from Ref. [208].

summarised in the following, while a description of the various experiment's components and the introduction of the used acronyms is in Chapter 3.

10.3.2.1 Tracker

The tracker system is composed of an inner pixel detector and an outer strip detector. Both systems will have an increased granularity in the Phase-2 upgrade.

- **Inner Tracker** The pixel size will be reduced by a factor 6 with respect to the Phase-1 detector and the sensors will be thinner to improve the two track separation and the impact parameter resolution for better b-tagging and reconstruction of τ hadronic decays in boosted jets. The system coverage will be extended up to $|\eta| < 4$ to match the calorimeter coverage.
- **Outer Tracker** Higher granularity will be achieved by reducing the length of the strips. The detector material budget will be also decreased to lower the rate of γ-conversions and improve the p_{T} resolution. The tracker will also contribute to the L1 trigger, increasing the background rejection at the earlier stage of the event selection process.

Details are provided in Sec. 10.4.

10.3.2.2 Calorimeters

Both the electromagnetic and the hadronic endcap calorimeters will be replaced by the High Granularity Calorimeter (HGCAL) [209]. It will provide excellent three dimensional shower images thanks to the unprecedented transverse and longitudinal segmentation. To withstand the foreseen total fluence of $\approx 10^{16}$ n_{eq} cm^{-2}, silicon sensors will be used as active material in the front sections. Plastic scintillator tiles combined with Silicon Photo Multipliers (SiPM) will be employed in the rear part, where the expected fluence is 8×10^{13} n_{eq} cm^{-2}.

The electromagnetic section covers 25 X_0 and one interaction length λ, while the hadronic part 3.5 λ. The silicon cells size of the electromagnetic and of a large

fraction of the hadronic calorimeters will be $\approx$ 0.5-1 cm^2. Plastic scintillators with $\approx$ 4-30 cm^2 segmentation will cover the remaining part of the hadron calorimeter.

Such high granularity will greatly benefit the reconstruction of physics objects and the particle-flow calorimetry, allowing also pileup subtraction. A further advantage of the high density of the HGCAL is the lateral compactness of the particle cascades allowing good two-shower separation.

The barrel calorimeters will be only partially replaced. The EB front-end electronics and HB off-detector electronics will be substituted while the EB photodetectors and lead tungstate crystals and the HB absorber, active material and front-end electronics will be retained [210].

10.3.2.3 MIP Timing Detector

The MIP Timing Detector (MTD) [211] will give timing information for MIPs with 30-40 ps resolution at the beginning of the HL-LHC operation, degrading slowly due to radiation damage to 50-60 ps by the end of HL-LHC operations. It exploits the fact that the individual interactions within the bunch crossing are distributed over time, with an RMS of 180-200 ps, to assign charged tracks to the correct interaction vertices. Tracks pointing roughly towards a reconstructed vertex but coming at the wrong time can be eliminated from consideration as contributing to that particular collision. The MTD consists of a timing layer placed between the tracker and the barrel calorimeter (crystal scintillators read out with SiPMs) and two disks in front of the electromagnetic calorimeter endcap (silicon Low Gain Avalanche Detectors - LGADs). The time of electromagnetic showers will also be determined to a precision of 30-50 ps, above p_T of a few GeV, in the upgraded calorimeters. Combining this information with the MTD measurement, photons can be associated to the correct charged particle vertex.

10.3.2.4 Muon system

The muon system consists of several subsystems to identify and measure muons at the required precision and accuracy needed to fulfil the CMS physics program. To maintain excellent performance at the LHC, the electronic of the existing DT, CSC, and RPC detectors will be upgraded. Furthermore, four additional stations with higher rate capability will maintain good L1 muon trigger acceptance in the forward region [212].

Gas Electron Multiplier (GEM) chambers with good position resolution are placed in a region where the magnetic field is still reasonably high to improve momentum resolution for the standalone muon trigger and to improve the matching with tracks in the global muon trigger. The two last stations will use low-resistivity RPC with lower granularity but good timing resolution. GEM chambers have been already successfully installed, representing the first completely installed Phase-2 detector upgrade [213, 214].

10.3.2.5 Trigger

The high average 140-200 PU of the HL-LHC poses a serious challenge to the trigger system, thus the Phase-2 upgrade of the L1 system utilizes technological advances to enhance the physics selectivity already at the hardware level of the data acquisition system [215, 216]. To profit from the extended coverage and increased granularity of the upgraded CMS detector, the latency of the system is extended from

the current 3.4 µs (limited by the tracker readout) to 12.5 µs. This will provide sufficient time for the hardware track reconstruction and matching of tracks to muons and calorimeter information, for the first time at L1. Field Programmable Gate Arrays (FPGA) coupled with high-speed optical links are used to retrieve the detector data and provide the systems interconnections. Sophisticated algorithms, including machine learning-based approaches, target the selection of specific final states, maintaining the efficiency of the signal selection to the level of the Phase-1 performance and significantly enhancing, or enabling, the selection of any possible new phenomena that could lead to unconventional signatures.

10.4 Inner and Outer tracker upgrade

10.4.1 Limitations of the present tracker

The tracker system currently in operation in CMS has been described in Sec. 3.3. The strip detector was designed to maintain high efficiency up to an instantaneous luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, an average PU of 20-30 and an integrated luminosity of $500\,\mathrm{fb}^{-1}$. Despite the fact that the instantaneous luminosity during Run 2 already exceeded the design value, the tracker showed very good performance. Nevertheless degradation due to radiation damage is expected above $500\,\mathrm{fb}^{-1}$. The Phase-1 pixel detector is already a replacement of the original one that had to be substituted due to dynamic inefficiencies in the readout chip at high rates. Furthermore, the tracker provides coverage only up to $|\eta| < 2.4$.

The main effects of radiation damage in the strip tracker are the increase of leakage current and full depletion voltage. While the former can be mitigated by the lowering of the operating temperature, the latter is not reduced by it. The major consequences of accumulating radiation dose in the pixel sensors are lower charge collection efficiency and reduced Lorentz angle. The first causes the degradation of the hit efficiency. The Lorentz angle Θ_L is the angle between the drift velocities of the charge carriers in a sensor with and without magnetic field applied. If electric field $\vec{E}$ and magnetic field $\vec{B}$ are perpendicular, $\tan(\Theta_L) = \mu_H B$, with μ_H the Hall mobility. The Lorentz angle determines the lateral drift of charges and its reduction results in decreased charge sharing among adjacent pixels and deteriorated resolution.

As it is foreseen that above $1000\,\mathrm{fb}^{-1}$ the degradation of the detector performance in terms of tracking, b-tagging and impact parameter resolution will be unacceptable, the tracker system will be replaced for operation at HL-LHC.

10.4.2 Requirements of the Phase-2 Tracker

The Phase-2 Tracker will be composed of two silicon based sub-systems. The Outer Tracker (OT) will consist of strip and macro-pixels sensors, while the Inner Tracker (IT) will be made of pixel sensors [11].

The leading requirements to the design of the tracker upgrade are listed below:

- **Radiation hardness** The tracker must be fully efficient up to the target integrated luminosity of $3000\,\mathrm{fb}^{-1}$ with a 50% margin. While the design will grant accessibility to the IT during ordinary shutdowns to exchange modules damaged by radiation, no maintenance intervention is planned for the OT. The

maximum expected fluence for the different tracker components is summarized in Table 10.1, reaching a fluence of 2.3×10^{16} 1 MeV n_{eq} cm^{-2} in the innermost regions of the IT. Variations along z are moderate as the particle fluence depends primarily on r.

TABLE 10.1: Maximum expected fluence for selected detector regions or components of the tracker. Values are for 3000 fb^{-1} of p p collisions at $\sqrt{s}$ = 14 TeV assuming a total cross section σ_{pp} of 80 mb. The given positions in r and z correspond to the location at which the quoted maximum fluence levels for the respective region or component type are reached. Adapted from Ref. [11].

Region or component	Max. fluence [n_{eq} cm^{-2}]	r [mm]	z [mm]
IT barrel layer 1	2.3×10^{16}	30	0
IT barrel layer 2	5.0×10^{15}	61	0
IT barrel layer 3	2.0×10^{15}	104	0
IT barrel layer 4	1.5×10^{15}	146	89
IT forward, ring 1	1.0×10^{16}	51	252
OT PS modules	9.6×10^{14}	218	129
OT 2S modules	3.0×10^{14}	676	2644

- **Increased granularity** To ensure efficient tracking in a high pileup environment, the channel occupancy must be kept at around or below the per cent level in the OT and the per mille level in the IT.
- **Improved two-track separation** Due to hit merging in the pixel detector, the track finding performance in highly energetic jets of the present tracker is limited. Therefore, two-track separation needs to be improved for data taking during high luminosity operation.
- **Reduced material in the tracking volume** As the amount of material impacts the reconstruction in the tracker, the Phase-2 detector will be lighter.
- **Robust pattern recognition** The upgraded tracker is designed to enable fast and efficient track finding, notably at the HLT.
- **Contribution to the level-1 trigger** In high pileup conditions selection algorithms become inefficient and the choice of relevant physics events at the first trigger stage becomes extremely challenging. To preserve and possibly enhance the performance in a wide spectrum of physics channels, tracking information will be used in the L1 event selection.
- **Extended tracking acceptance** The upgraded tracking system will provide efficient tracking up to about $|\eta| = 4$ to improve the physics capabilities in the forward region.

Fig. 10.5 shows one quarter of the Phase-2 tracker layout. The number of layers has been optimized to ensure robust tracking:

- Unaffected performance if one layer is lost.
- Track finding at L1 trigger.
- Pixel based track seeding.

In the next sections details on the two sub-systems will be provided.

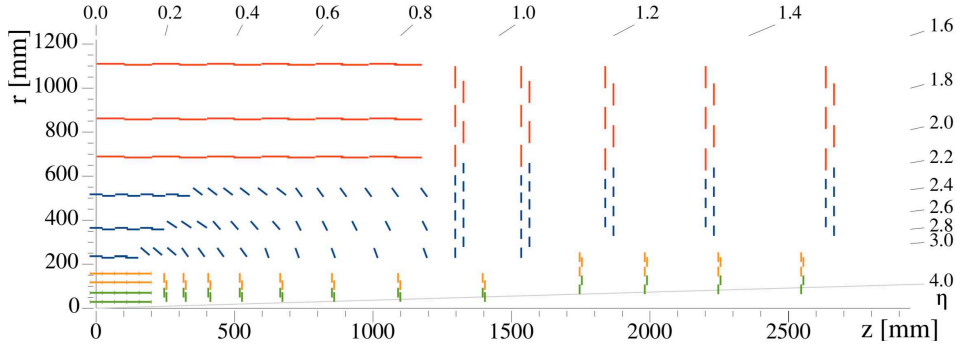


FIGURE 10.5: Sketch of one quarter of the tracker layout in r-z view. In the Inner Tracker the green lines correspond to pixel modules made of two readout chips and the yellow lines to pixel modules with four readout chips. In the Outer Tracker the blue and red lines represent the macro-pixels and strip sensors, respectively. Taken from Ref. [11].

10.4.3 The Outer Tracker

The design of the Outer Tracker has been strongly driven by the necessity to contribute to the L1 event selection. Thus, the OT has to self-select information at every bunch crossing, relying on data being locally reduced by the front-end electronics.

This is accomplished by so called p_T-modules, capable of rejecting signals from particles below a certain p_T threshold. The concept, illustrated in Fig. 10.6, is based on the bending of charged particles in the transverse plane by the 3.8 T field of the CMS solenoid, with angles depending on their p_T. The modules consist of two single-sided closely-spaced sensors read out by a common set of front-end ASICs. The signals in the two sensors can therefore be correlated to select hit pairs ("stubs") compatible with particles above the chosen p_T threshold. Selecting particles with a minimum p_T of $\approx 2\,\text{GeV}$, the data volume is reduced by an order of magnitude, enabling the stubs transmission at 40 MHz.

For the approach to work, the strips of the top and bottom sensors of a module are parallel to each other and parallel to the z axis in the barrel and nearly radial in the endcaps. To be able to measure the z and r coordinates in the barrel and endcaps, respectively, the p_T-modules come in two variants:

- **Pixel-Strip (PS) modules** consist of a strip (PS-s) and a macro-pixel sensor (PS-p). The strip and macro-pixel length is about 2.4 cm and 1.5 mm, respectively. They are used for the three innermost layers of the OT.

- **2 strips (2S) modules** are made of two strip sensors. The strips have a 5 cm length. These modules are used in the outermost three layers, in the radial region above 600 mm.

In the endcaps disks, the modules are arranged in ring structures, with the internal rings equipped with PS modules, while 2S modules are used at larger radii.

For this concept to work, both top and the bottom sensors of a module must be connected to the readout electronics. To achieve a reliable connectivity, the two halves of each module are read out independently. To avoid inefficiencies in reconstructing stubs when particles cross the module near the centre with a large incident angle, the first three barrel layers, where the PS modules are used, have an innovative layout. The modules are progressively tilted to be nearly perpendicular to incident particles over the entire barrel length, as visible in Fig. 10.5.

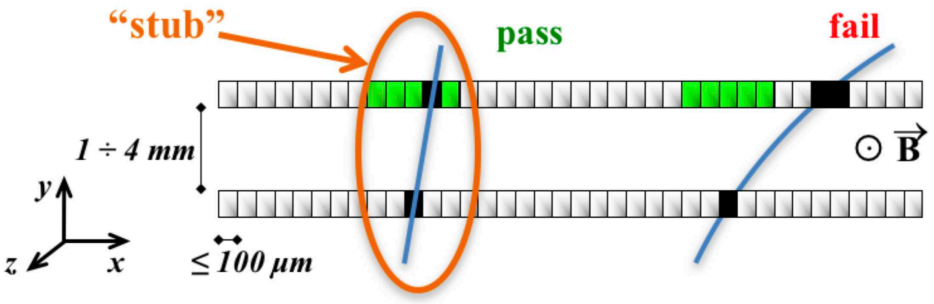


FIGURE 10.6: Illustration of the p_T-module concept. Correlation of signals in closely-spaced sensors enables rejection of low-p_T particles; the channels shown in green represent the selection window to define an accepted stub. Taken from Ref. [11].

10.4.4 The Inner Tracker

For its location close to the interaction point, the Inner Tracker has to withstand the most challenging environment: ionizing radiation dose of up to 1.2 Grad, a hadron fluence of up to 2.3×10^{16} 1 MeV n_{eq} cm^{-2} for 3000 fb^{-1}, a pileup of 140 (or even 200), and hit rates approaching 3 GHz cm^{-2} in the inner layers (for the ultimate scenario). The Phase-2 Inner Tracker (IT) is designed to maintain or improve the tracking and vertexing capabilities under such demanding conditions.

The IT increases the granularity by a factor 6 with respect to the present pixel tracker to limit the occupancy to the per mille level and improve track separation inside jets. The impact parameter resolution will also benefit from a reduced pitch. High radiation tolerance is one of the key requirements. The IT will provide a large forward acceptance and pileup mitigation in the calorimeters thanks to a geometrical coverage up to $|\eta| \approx 4$. To allow the potential replacement of inefficient parts, the design grants a simple installation and removal of the detector. Additionally, it will contribute to the real-time instantaneous luminosity measurement.

Good and relatively stable spatial resolution with respect to radiation damage will be achieved by using n-in-p type silicon pixel sensors with a thickness of 150 µm, segmented in cells of 100×25 µm^2 (with the long side pointing along z in the barrel and along r in the endcaps), or 50×50 µm^2.

Readout chips will be realized in 65 nm CMOS technology to fit the pixel segmentation. In addition, they will feature an architecture where a group of channels (referred to as pixel region) shares digital electronics for buffering, control, and data formatting. The RD53 collaboration [217], a joint ATLAS-CMS effort, is developing such a pixel readout chip (PROC). The PROC active dimensions are 16.4×22.0 mm^2. Two types of modules are foreseen, with two and four chips, arranged as one by two, referred to as 1×2 and 2×2 modules, respectively.

The IT is composed by a 4 layer barrel section (Tracker Barrel Pixel Detector, TBPX) and eight small forward double-disks (Tracker Forward Pixel Detector, TFPX) plus four large endcap disks (Tracker Endcap Pixel Detector, TEPX) at each side of the TBPX. In the TBPX, the pixel modules are arranged in "ladders". They are mounted staggered in radius to obtain an overlap in the r-ϕ direction. As the modules on a ladder do not overlap in z, gaps at $\eta = 0$ are prevented by having an odd number of modules along z and splitting the barrel mechanics in z into slightly asymmetric halfs. In TFPX and TEPX, the modules are arranged in concentric rings. Each double-disc consists of two discs with overlaps in r and r-ϕ. Each disc is split

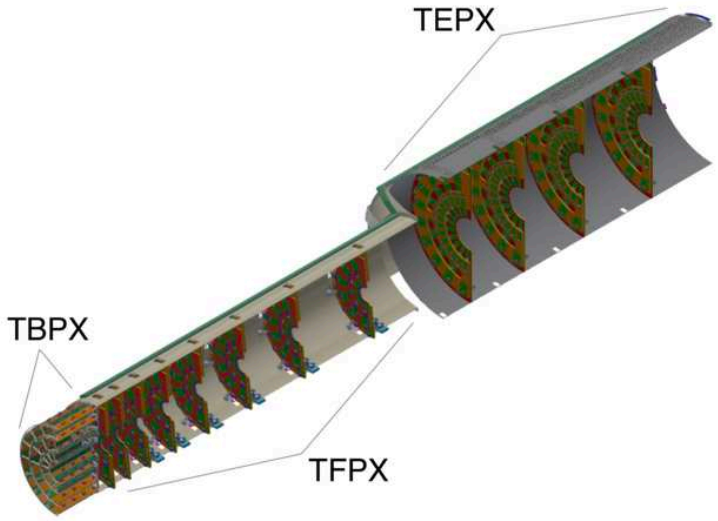


FIGURE 10.7: Perspective view of one quarter of the Inner Tracker, showing the TBPX ladders and TFPX and TEPX dees inside the supporting structures. The pixel modules are shown as orange elements in TBPX and as green elements in TFPX and TBPX. The dees are depicted as red and orange surfaces. Taken from Ref. [11].

into two D-shaped halves, referred to as "dees". The TEPX will provide the required luminosity measurement capability.

In addition to Fig. 10.5, a perspective view of the IT elements is shown in Fig. 10.7.

Planar pixel sensors are the baseline technology for the outer 3 of the 4 TBPX layers, where the irradiation level will reach a neutron equivalent fluence of 5×10^{15} cm^{-2}, while for the innermost TBPX layer and TFPX ring, 3D sensors are also considered.

In the following, first the pixel module concept is summarized, then an overview of the sensor options under consideration is presented, followed by a description of the read-out electronics.

A detailed characterization of the sensors studied for this thesis work will be given in Sec. 12.6.

10.4.4.1 Pixel module

The Phase-2 pixel module, based on the Phase-0 and Phase-1 pixel detectors, is composed of a pixel sensor, two or four read-out chips to which the sensors are bump-bonded, a flex circuit (High-Density Interconnect - HDI), and a mechanical support. The HDI printed circuit board is used to distribute data, commands and power to and from the modules [218]. A layer of thermally conductive carbon foam connects the CO_2 cooling pipes to the module, removing the generated heat. The pixel chips and sensors will be operated at a temperature of about $-20\,^{\circ}$C.

10.4.4.2 Pixel sensors

Pixel sensors optimized for the HL-LHC conditions will be produced on 6" n-in-p wafers. The choice of the n-in-p presents several advantages. For instance, especially relevant in a high radiation environment, collecting electrons, they can be operated under-depleted. Additionally, they are manufactured with single-sided processing, as opposed to n-in-n sensors, limiting the production costs. However, the choice of an n-in-p bulk implies that the high voltage at the periphery of the sensor is close to

the readout chip and spark protections through Parylene coating need to be introduced.

While the sensor thickness in the present tracker is between 270 and 285 µm, thinner sensors are preferred for Phase-2 because the initial advantage of a larger signal in thicker sensors disappears when severe trapping is present at large fluences. Furthermore, thinner sensors need smaller operational voltages and produce less leakage current. Thus, the active thickness of the planar pixel sensors will be 150 µm.

The pixel area is reduced by a factor of six with respect to the present CMS pixel detector. Two options are under consideration: $100\times25\,\mu m^2$ (with the long side pointing along z in the barrel and along r in the endcaps), or $50\times50\,\mu m^2$.

By adopting the $100\times25\,\mu m^2$ variant, the impact parameter resolution in the xy-plane is improved, as well as the p_T resolution. Square pixels lead to a more precise determination of the impact parameter along z. Nonetheless, the relative differences on the track parameters resolution are typically small, with a trade-off between primary vertex discrimination and resolution on the impact parameter. However, the $50\times50\,\mu m^2$ option would result in long clusters at the edges of the barrel layers, especially for the first layer, demanding more stringent requirements on the operational threshold of the chip and a larger bandwidth to read the data out. The rectangular variant is currently considered as the preferred option.

The design of pixel sensors with such reduced cell size presents several challenges and the optimization must take the design rules of silicon sensor foundries into account. Special attention is devoted to the pixel bias schemes to test sensors before bump bonding. Several options are considered, such as punch through biasing common to several adjacent pixels, individual bias dots or no biasing scheme at all. The latter would require additional techniques to access all pixels during the test of the sensor. Also for the pixel isolation, p-stop and p-spray alternatives are being investigated.

Planar sensors are the baseline technology for the outer 3 of the 4 TBPX layers, where the irradiation level will reach a neutron equivalent fluence of $5\times10^{15}\,cm^{-2}$, while for the innermost TBPX layer and TFPX ring, where fluences up to $2.3\times10^{16}\,cm^{-2}$ and $1\times10^{16}\,cm^{-2}$ are expected, respectively, 3D sensors are also considered.

In 3D sensors, columnar electrodes penetrate the bulk, decoupling the distance between electrodes from the sensor thickness. Thus thicker sensors could be used to generate a larger signal while maintaining small depletion voltages and efficient charge collection. Thanks to smaller distances between electrodes, compared to planar sensors, the impact of charge carrier trapping at large fluences is reduced. Therefore 3D sensors can achieve higher radiation hardness at a lower operating voltage, decreasing the risk of edge sparking and lowering power consumption. This comes with a more complex and expensive fabrication, relegating such sensors as an option only for the pixel detector area with the highest radiation dose.

Several sensors from different R&D submissions are being tested at the time of writing:

Planar
- Hamamatsu Photonics (HPK)
- FBK
- L-Foundry

3D
- FBK, Single sided DRIE process
- CNM.

A detailed characterization of the HPK sensors studied for this thesis work are given in Sec. 12.6. A preliminary study, now superseded by more recent publications, performed by the author of this thesis on FBK planar and 3D sensors is summarized in Appendix F.

10.4.4.3 Readout electronics

The development of a custom radiation hard readout chip is a crucial component of the IT readout system. It will be based on the Time over Threshold (ToT) method, in which the measurement of the deposited charge is obtained from the time during which the analogue pulse exceeds a certain threshold. The charge is digitized with 4 bits resolution. The detection threshold is required to be lower than 1200 electrons.

The readout chip will feature calibration pulse injection available for all pixels and extensive monitoring features, for example for the chip temperature.

The readout chip is designed in high density 65 nm CMOS technology to match the small pixel cells, the unprecedented hit rates of $3\,\mathrm{GHz\,cm^{-2}}$ and the data buffering required during the trigger latency. The minimal requirement in terms of radiation tolerance is 500 Mrad, leading to the replacement of the inner barrel layer after five years at full luminosity, while the ultimate goal is 1 Grad.

11 Performance of jet substructure

As described in Sec. 4.3, reconstructed high level objects are the basis for physics analyses. Thus, it is crucial to test the performance of the upgraded detector on reconstructing such objects. In this chapter, the focus will be on jets and their substructure. First the simulation setup is introduced in Sec. 11.1, then the figures of merit in the Phase-2 tracker are provided (Sec. 11.2). Finally, the performance of jet reconstruction and the impact on discrimination of a chosen signal from the background are described in Sec. 11.3 and Sec. 11.4, respectively.

The work described in Sec. 11.3 and Sec. 11.4 has been performed by the author of this thesis, under the supervision of Dr. Andreas Hinzmann. The simulated samples, described in Sec. 11.1, were also produced by the author.

11.1 Simulation setup

Similarly as it was explained in Sec. 4.2, also for this study the MC generators are interfaced with the GEANT4 Toolkit to carry out a detailed simulation of the CMS detector, and the 3.8 T magnetic field taken into account by field maps. The simulation of the electronic response (*digitization*), the emulation of the level-1 trigger, and the offline reconstruction of physics objects are performed as well, in the 200 PU scenario.

The detector parameters are modified to match the ones of the Phase-2 detector. In the Outer Tracker, the sensors are simulated with a thickness of 200 µm. The strips in 2S and PS-s sensors have simulated dimensions of 90 µm $\times$ 50.25 mm and 100 µm $\times$ 23.13 mm, respectively. In the PS-p sensors the simulated pitch (lenght) is 100 µm (1446 µm). In the Inner Tracker the sensor thickness is set to 150 µm and the 100×25 µm^2 pixel pitch variant is considered.

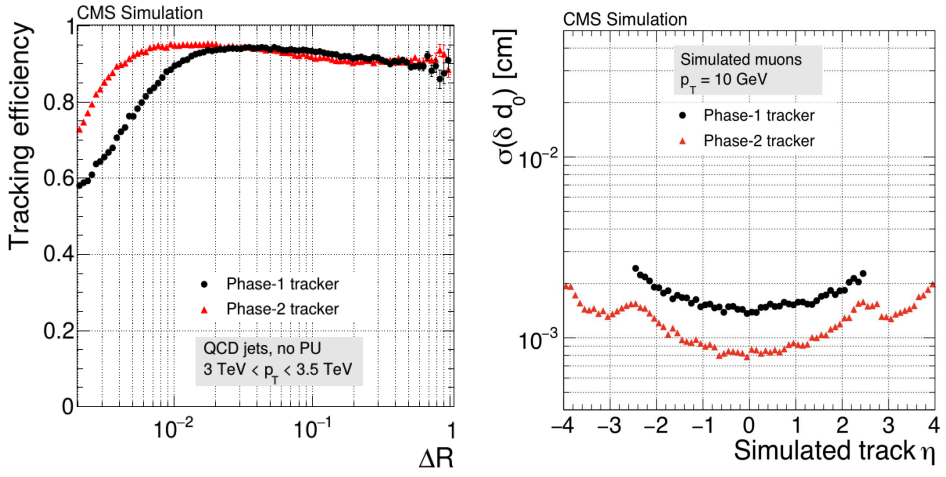
Furthermore, since digitized hits are created when the total charge associated with a given channel exceeds a predefined threshold, such threshold has also been implemented based on the knowledge of the future readout electronics available at time the Phase-2 Tracker TDR [11] study was performed. For pixel modules, the parameters have been chosen to reproduce the expectation of the RD53 chip [217]. Thus, the threshold is set to 1200 e^- and a readout based on the Time over Threshold method with 4 bit charge resolution is implemented. For the Outer Tracker the threshold is 0.4 times the most probable charge from a minimum ionizing particle at normal incidence to ensures a $\approx$ 99.5% efficiency while keeping the noise contribution at an almost negligible level.

The details of the simulation setup for the studies presented in Sec. 11.3 and Sec. 11.4 are listed in Appendix E.

11.2 Tracking performance

The track finding efficiency, the fake rate, and the resolution of the estimated track parameters have been considered to evaluate the performance of the track reconstruction. A reconstructed track is associated to a simulated particle if at least 75% of its hits originate from this simulated particle. If a reconstructed track is not associated with any simulated particle, it is considered as a combination of unrelated hits and designated as a fake track. The tracking efficiency is defined as the fraction of charged particles associated to at least one reconstructed track, while the fake rate is the fraction of fake tracks in the set of all reconstructed tracks. Efficiency and fake rate not only depend on the algorithms and track selections used, which are not fully adapted to the Phase-2 conditions yet, but also on the intrinsic tracker properties, such as its layout, its material budget, and the precision of its measurements, providing a good estimation of the Phase-2 detector performance. For single muons with a transverse momentum of 10 GeV, the efficiency is stable and close to 100% in the entire pseudorapidity range. For tracks generated in $t\bar{t}$ events, passing a certain set of quality requirements, the efficiency is around 90% in the central region while the fake rate is lower than 4% in the entire pseudorapidity range [11].

The ability to reconstruct tracks inside jets is especially relevant for the studies presented in Sec. 11.3. The tracking efficiency inside high momentum jets is shown in Fig. 11.1a as a function of the distance between a simulated track and its nearest neighbour ΔR. The performance of the Phase-1 and Phase-2 trackers are compared in absence of pileup. The Phase-1 reconstruction employs a special algorithm to split clusters and iteration to perform robust tracking in jet cores, which has not been ported to the Phase-2 reconstruction yet. Nevertheless, thanks to the higher granularity of the new detector, a significantly higher efficiency is visible for small values of ΔR and further improvement is expected for large values of ΔR as well after applying a dedicated tuning.



(A) (B)

FIGURE 11.1: Left: Tracking efficiency in cores of jets with $3 < p_T < 3.5$ TeV as a function of the distance ΔR between a simulated track and its nearest neighbour, for the Phase-1 (black) and the Phase-2 (red) tracker, without pileup. Right: Resolution of the transverse impact parameter as a function of the pseudorapidity for the Phase-1 (black dots) and the upgraded (red triangles) tracker, using single isolated muons with a transverse momentum of 10 GeV. Taken from Ref. [11].

The transverse impact parameter d_0 is one of the five observables used to fully describe a particle trajectory, evaluated at the point of closest approach of the track to the beam axis. The resolution of the track parameters is given by the RMS of the residuals differences between the reconstructed and the simulated track parameters. The resolution of the transverse impact parameter d_0 as a function of the pseudo-rapidity is shown in Fig. 11.1b for single muons with a 10 GeV p_T. The transverse impact parameter resolution in the Phase-2 detector is below 10 µm in the central region, improved with respect to the Phase-1 detector, and $\approx$ 20 µm at the edge of the acceptance region, where it cannot be measured with the present detector.

11.3 Jet substructure reconstruction

Jet substructure techniques, as the softdrop jet mass and the N-subjettiness introduced in Sec. 5.3, are largely used to identify highly boosted W, Z, and Higgs bosons, as well as top quarks. Such techniques take advantage from the particle flow approach (see Sec. 4.3 for details), profiting from the best possible momentum and angular resolution for each particle, reconstructed from all sub-detectors.

The tracker provides important information for the PF algorithm. For hadronically decaying W bosons with 2 TeV p_T, from Eq. 5.1, the maximum angular separation between the $q\bar{q}$-pair initiating the shower is 0.08 and the separation among each particle created in the hadronization will be even smaller. Thus, in this case, the high angular resolution demonstrated in Fig. 11.1a, is fundamental for identifying the two quark-originated sub-jets.

The HL-LHC high number of pileup interactions constitutes a challenge for the jet substructure reconstruction, which is addressed not only by adopting a high granularity tracker but also employing algorithms capable of mitigating the contamination from *soft* interactions. In the following the PUPPI technique, described in Sec. 4.4, is adopted.

The increased center-of-mass energy achievable at the HL-LHC allows extending the mass reach of many searches, assuming the reconstruction holds in highly boosted jets. Jet energy composition can be used to test the reconstruction of large momentum jets as no relevant dependence from the object energy is expected from MC simulations. During the hadronization process, pions, the lightest mesons in nature, are created with the highest probability from the kinetic energy of the mother parton. Pions come in three flavour, π^-, π^+ and π^0, and each of them has the same probability to be produced. Each pion type carries $\approx$ 20% of the jet energy. This implies that roughly 65% of the jet energy comes from pions. Charged pions have a decay length bigger than 10 mm and are considered stable while neutral pions decay instantly to photons. The nucleons and kaons carry about 15% each, with the remaining 5% in fragmentation photons, light and strange baryons, leptons and other particles [149].

The jet particle composition can be summarized as follows [149]:

- $\approx$ 60% charged particles
- $\approx$ 25% photons
- $\approx$ 15% neutral hadrons
- $\approx$ 1% neutrinos.

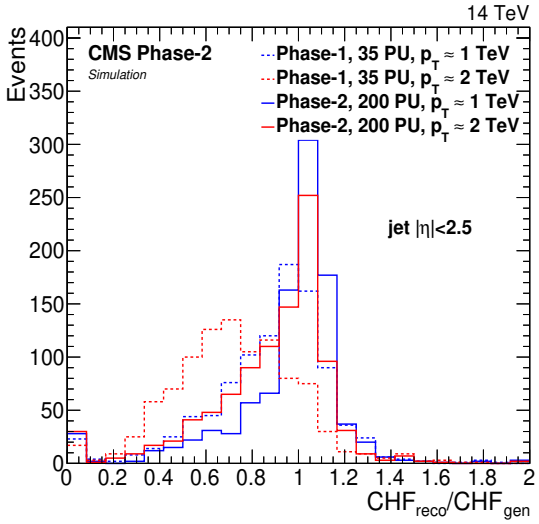


FIGURE 11.2: Ratio between the reconstructed charged hadron energy fraction, CHF_{reco}, and the charged hadron energy fraction at generator particle level, CHF_{gen} for simulated events of pair production of boosted W boson jets with transverse momenta of 1 TeV (blue) and 2 TeV (red). Solid (dashed) lines refer to Phase-2 (Phase-1) simulation. Published in Ref. [11].

The reconstruction of charged particles makes large use of the tracker information, therefore, by measuring the Charged Hadron energy Fraction (CHF) of a jet, the performance of the upgraded tracker can be tested.

The ratio of CHF reconstructed from tracker and calorimeter information and the corresponding fraction at generator particle level in shown in Fig. 11.2 for simulated events of pair production of boosted W boson jets with transverse momenta of 1 TeV and 2 TeV. At such high jet momenta, also the particles inside the jets have large momenta, small track curvature, and are as well very collimated. As it was already demonstrated in Fig. 11.1a, charged hadron reconstruction is degraded for the Phase-1 tracker, resulting in a lower charge hadron energy fraction. The degradation is especially relevant for high momentum tracks, as shown in Fig. 4.4. Instead, charge hadron reconstruction in jets remains efficient up to transverse momenta of 2 TeV for the Phase-2 tracker. This will allow for highly granular jet substructure reconstruction even at higher momenta than achieved with the Phase-1 tracker, as required by the operation at HL-LHC.

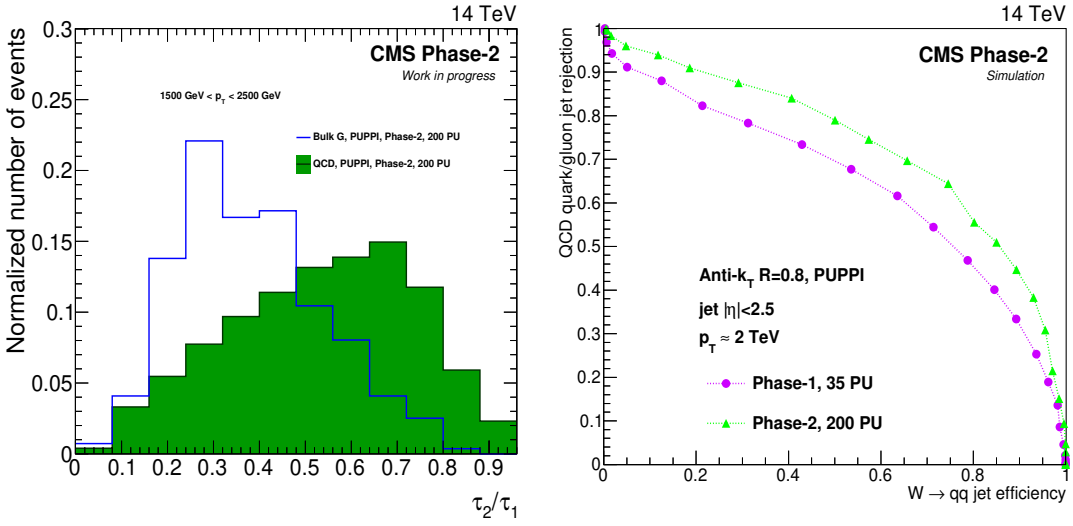
11.4 Signal discrimination

Once verified that the jet reconstruction with the Phase-2 tracker holds to even higher momenta than with the Phase-1 detector, substructure techniques can be used to identify particles and distinguish signals from QCD background.

The target signal used in this study is a Bulk Graviton G with a mass of 4 TeV. It decays into two W bosons. This study considers the case where one W decays hadronically into two quarks and the other one decays leptonically. The purpose of this analysis is to understand how good the jet originated from the hadronically decayed W boson, for a jet p_T of about 2 TeV, can be distinguished from a QCD jet.

The variable used to discriminate the signal jets from the background is the N-subjettiness ratio τ_2/τ_1, introduced in Sec. 5.3.2. τ_N is a jet shape observable which,

given N subjet axes within a jet, sums the angular distances of jet constituents to their nearest subjet axis, to quantify how likely a jet is to have N or more subjets.



(A) (B)

FIGURE 11.3: Left: N-subjettiness ratio τ_2/τ_1, for the QCD multijet background in green and for a jet originated from $W \rightarrow q\bar{q}$ in blue. Right: QCD quark and gluon jet background rejection as a function of boosted W boson jet identification efficiency with the N-subjettiness ratio τ_2/τ_1. The performance of the Phase-1 and Phase-2 simulations is compared. Published in Ref. [11].

The τ_2/τ_1 distribution for the signal and background described above is shown in Fig. 11.3a. It is possible to see how signal and background distributions have the peak position at lower and higher τ_2/τ_1 values, respectively. Thanks to the improved jet reconstruction with the Phase-2 detector, significant gain is expected in the identification of boosted objects with τ_2/τ_1 and can be observed in Fig. 11.3b. Here, the ROC curves comparing the performance of the Phase-1 and the Phase-2 simulation are showed. For a typical background rejection rate of 70%, a 30% relative gain in efficiency for boosted W, Z and H bosons is expected for jets with a p_T of 2 TeV.

Additionally, in all-jets searches, for instance for $G \rightarrow WW \rightarrow q\bar{q}q\bar{q}$ as presented in the first part of this thesis, where the relative gain is squared, a factor 1.7 better signal efficiency and cross section limits are expected at di-boson resonance masses of 4 TeV.

12 Silicon sensors

In the first part of this chapter, the most relevant properties of Silicon for its applications in particle detectors are reviewed. Segmented silicon detectors allow the reconstruction of the incoming particles position and are largely employed at collider experiments like CMS. They provide the high resolution necessary for in jet tracking, crucial, for instance, for the identification of b-quarks from Higgs decays. Pileup and a high radiation environment constitute the main challenges for operation at the LHC. The most important causes and consequences of radiation damage are also addressed.

In the second part, the sensors and the read-out chip used for the studies presented in this work are described.

This chapter's purpose is to provide a summary of the working principles of silicon detectors which are comprehensively addressed in literature [219–222]. Additionally, radiation effects are extensively discussed in Ref. [223, 224]. A review of resolution measurements is presented in Sec. 13.1.

12.1 Silicon properties as a semiconductor

Silicon (Si), a member of group IV in the periodic table, is a tetravalent-metalloid crystalline solid and classified as a semiconductor. In crystals, the periodic lattice establishes allowed energy bands for electrons that exist within that solid. The valence band is the lower energy band and corresponds to those electrons that are part of the covalent bonding that constitute the inter-atomic forces within the crystal. The next higher-lying band is called the conduction band. Electrons in this band are free to migrate through the crystal and contribute to the material's electrical conductivity. Elements are classified as conductors, semiconductors, or insulators, depending on the size of the energy gap that separates valence and conduction bands. For insulators, the band-gap is usually larger than 5 eV, whereas, for semiconductors, the band-gap is considerably less. Among the class of semiconductors, silicon has a band-gap of 1.12 eV at 300 K.

Electrons follow the Pauli principle, and thus the occupancy of energy states is described by the Fermi-Dirac distribution

$$f(E) = \frac{1}{e^{(E-E_F)/k_B T} + 1}, \tag{12.1}$$

where k_B is the Boltzmann constant and T the temperature. The Fermi energy E_F is the highest possible energy of a fermion at $T = 0$ K. At $E = E_F$ exactly half of the available levels are occupied. At low temperatures and in the absence of impurities, the valence band is full, and the conduction band is empty. If a valence electron, normally part of a covalent bond, gains sufficient thermal energy, it can move across the band-gap into the conduction band. The electron can thus leave the specific bonding

site and drift throughout the crystal. The excitation process also leaves a vacancy, called a hole, in the otherwise full valence band, representing a net positive charge which can move in an applied electrical field; holes are treated like particles and follow the Fermi-Dirac statistic. Electrons and holes are the fundamental charge carriers of a solid-state device. The concentrations of free electrons n in the conduction band at the energy E_C and holes p in the valence band at the energy E_V are

$$n = N_C \exp\left(-\frac{E_C - E_F}{k_B T}\right) \tag{12.2}$$

and

$$p = N_V \exp\left(-\frac{E_F - E_V}{k_B T}\right), \tag{12.3}$$

with N_C and N_V the effective density of states. In absence of electrically active impurities, a semiconductor is called intrinsic and has a very low concentration $n_i = n = p$ of free carriers and a high resistivity. In silicon, n_i is $9.65 \times 10^9\,\mathrm{cm}^{-3}$ at 300 K [225].

The insertion of impurities to the material creates additional energy levels in the band-gap, altering the electron and hole carrier concentrations at thermal equilibrium and thus its conductivity. This process of replacing silicon atoms with impurity elements is called *doping*.

The doped material can be of two types depending on the nature of the impurities. The addition of a group V element, e.g. Phosphorus or Arsenic, with one more valence electron than the host silicon atoms, produces extra electrons charge carriers. The extra-electron left out by the covalent bond has energy near the top of the gap. The energy spacing between this energy level and the bottom of the conduction band is sufficiently small to ensure that a significant fraction of electrons are excited. The remaining doping atom becomes a positive ion. Elements like Phosphorus or Arsenic are called donors because they provide extra conduction electrons. At room temperature, all donor states are ionized, and the concentration of electrons equals the concentration of donor atoms. This material is called "n-type" and electrons are majority charge carriers.

On the contrary, by adding group III elements, e.g. Boron, an excess of holes is obtained. The dopant atoms have one less valence electron than the host Silicon atoms, leaving one covalent bond unsaturated where a vacancy similar to a hole but with energy near the bottom of the gap is created. Since the energy difference between this energy level and the top of the valence band is small, a large fraction of thermal electrons will fill it, leaving a hole in their site of origin. This kind of impurities are called acceptors as they catch electrons and the doped material is called "p-type", where holes are then majority charge carriers.

After their formation, electrons and holes take part in a random thermal motion that results in their diffusion away from the point of origin, described by the Einstein relationship

$$D = \frac{k_B T}{e}\mu. \tag{12.4}$$

where D is the diffusion coefficient. Typical mobility μ values at 300 K are $\mu_n = 1350\,\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ for electrons and $\mu_p = 450\,\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ for holes, respectively [220]. If an electric field is applied, placing electrodes at the boundaries of the silicon volume, electrons and holes will tend to move in opposite directions. The motion is a combination of a random thermal velocity and a net drift velocity parallel to the

direction of the applied electric field. At low to moderate values of electric field intensities, the drift velocity, for both electrons v_n and holes v_p, is proportional to the applied field E according to $v_i = \mu_i E$ ($i = n, p$). As a consequence of the different mobility values, for the same electric field intensity, electrons are faster than holes. At higher electric field intensities, the number of random collisions per unit time increases and the drift velocity reaches saturation. While drifting, the charge carriers are also diffusing.

The motion of either electrons or holes constitutes a current that will persist until those carriers are collected at the electrodes. The total measured current I is the sum of the two separate contributions. Therefore, given a semiconductor with bulk thickness t and surface area A, the current that will flow when a voltage V is applied across the thickness is

$$I = I_n + I_p = \mathrm{A}e(nv_n + pv_p) = \mathrm{A}e\frac{\mathrm{V}}{t}(n\mu_n + p\mu_p).$$

The resistivity of a semiconductor is defined as

$$\rho = \frac{1}{e(n\mu_n + p\mu_p)} \tag{12.5}$$

and it is an important parameter for silicon sensors as it relates the applied voltage to the flowing current

$$\rho = \frac{\mathrm{AV}}{It}. \tag{12.6}$$

12.2 The p-n junction

The contact of two regions with different doping in the same crystal, one p-type and the other n-type, forms a p-n junction. Regardless of the technological procedures leading to the realization of this type of junction, it is a thin area of fast transition from the p-doped region, rich in positive charge carriers, to the n region, doped with donors and rich in negative charge carriers. Thus in the junction, a strong concentration gradient of charge carriers is present, which causes the motion of carriers from regions of high concentration to regions of low concentration. As a consequence of the diffusion of electrons from n-type material into p-type material, and vice versa, holes from p-type to n-type material, the charge carriers recombine, forming a space-charge region (SCR) depleted of free charge carriers. This depleted region extends on the two sides of the junction. As a result, one side hosts electron acceptor sites not compensated by free holes (p-type), and the other contains positively charged empty donor sites not compensated by electrons (n-type). The accumulated space charge creates an electric field that, at equilibrium (i.e. without any external voltage), prevents further diffusion across the junction: a steady-state charge distribution is established and is described by the mass action law

$$np = n_i^2 = N_C N_V \mathrm{e}^{-\frac{E_C - E_V}{k_B T}}. \tag{12.7}$$

The region over which the charge imbalance exists is the depletion region, as there are no free charge carriers. If the concentrations of donors on the n side and acceptors on the p side are equal, the depletion region extends at equal distances to both sides. If the donor concentration in the n-type material is much higher than that of acceptor atoms in the p-type (p-n^+ junction), the electron diffusing across the

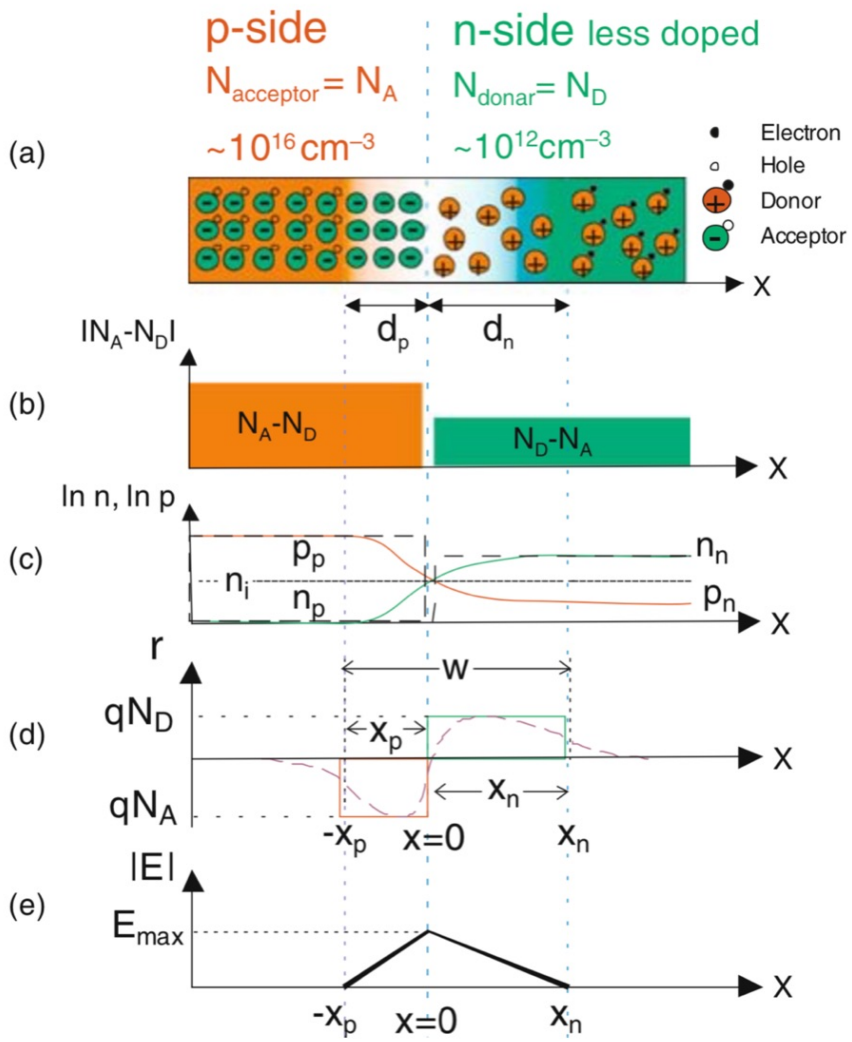


FIGURE 12.1: The diagrams display (a) a simple visualization of the atomic and charge configuration, (b) the doping profile, (c) the mobile charge density, (d) the space charge density and (e) the electrical field configuration. All states are depicting the equilibrium state, without any external voltage applied. Taken from Ref. [219].

junction will tend to travel a greater distance into the p-type material before recombination with holes, so that the depletion region would extend farther into the p side. In the opposite case (p^+-n junction), vice versa, the depleted region would extend farther into the n side. A schematic representation of the atomic and charge configuration, doping profile, mobile charge density, space charge density, and electrical field configuration across the p-n junction is shown in Fig. 12.1

By equipping both the p and the n region of the junction with an electrical connection, a "two terminals electronic device" is obtained, called a junction diode whose main feature is to have a highly non-linear current-voltage relation. If the diode is in a state of thermodynamic equilibrium, i.e. has a uniform temperature and is not connected to energy sources, the two terminals, supposed made of the same metal (e.g.

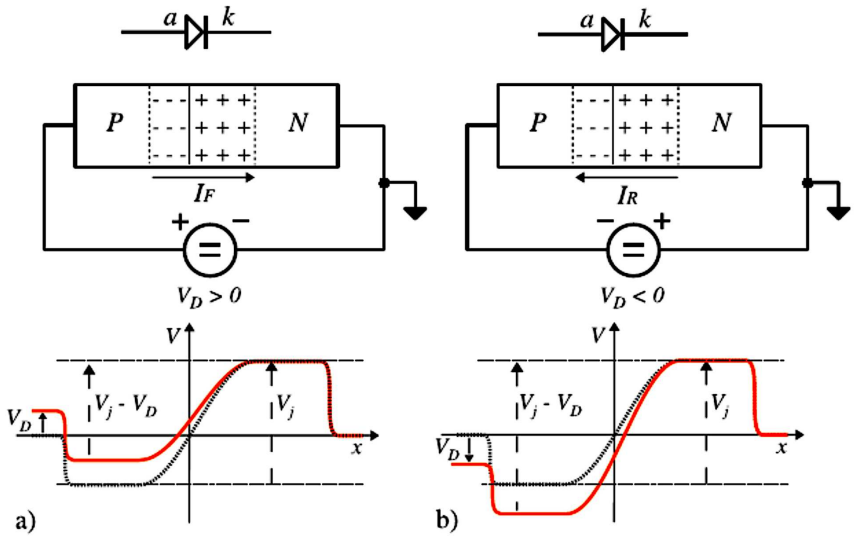


FIGURE 12.2: Direct (left) and reverse (right) polarization of a *p-n* junction. On the top (bottom) the electrical connections (potential distributions) are shown. V_j (black dashed line) is the the junction potential built-in voltage in absence of a voltage source while the red solid line represents the potential when the source is connected. Taken from Ref. [226].

copper), will be at the same electric potential. The electric potential of the diode, the junction potential built-in voltage V_j, is of the order of a few millivolts and is represented by the black dashed line in Fig. 12.2.

By connecting the diode junction to a voltage source V_D, two cases are possible depending on the sign of the applied voltage, illustrated in Fig. 12.2. On the left, the voltage source, applying a positive voltage to the *p* side with respect to the *n* side, constrains the potential barrier between the two regions to a lower value than that corresponding to the junction potential V_j, causing the diffusion of holes from the *p* region to the *n* region and of electrons in the opposite direction. These flows correspond to a current I_F (forward or direct current) in the direction indicated by the arrow. This current increases exponentially with the increase of V_D. In this condition, the diode is in forward bias. On the right, the potential barrier is forced to a higher value with respect to the V_j equilibrium potential, inducing a further removal of electrons and holes from the transition region, extending it. The diode is now in the reverse bias condition. In this configuration, only a weak current I_R (reverse or leakage current) flows, six to ten orders of magnitude lower than I_F. The reverse bias configuration is used to detect particles: By reducing the number of charge carriers in the silicon, and expanding the depletion region, it allows the detection of the e-h pairs generated by a particle passing through the material. The interaction of particles with silicon is summarized in Sec. 12.3.

12.2.1 Reverse bias configuration

In the reverse bias configuration, the width of the depletion zone *d*, the electric field, and the potential φ as a function of the applied voltage can be calculated by solving the one-dimensional Poisson's equation:

$$\frac{d^2\varphi}{dx^2} = -\frac{\varrho(x)}{\epsilon} \tag{12.8}$$

where ϵ is the dielectric constant of the material and ϱ the charge density. For a *p-n* junction with constant doping concentrations on both sides, the total width of the depletion region is

$$d = x_n + x_p = \sqrt{\frac{2\epsilon}{e}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)(V + V_j)} \tag{12.9}$$

where x_n and x_p are the depletion lengths on the n and p side with concentrations N_D and N_A, respectively, V the externally applied voltage. For a p^+-n junction ($N_A > 10^{18}\,\mathrm{cm}^{-3}$, $N_D \sim 10^{12}\,\mathrm{cm}^{-3}$), the term $1/N_A$ in Eq. 12.9 can be neglected, meaning that the space charge region is extending much deeper into the lower doped side of the junction. Furthermore, the built-in voltage is small compared to typical operation voltages and can therefore be neglected. With these assumptions

$$d \simeq x_n \simeq \sqrt{\frac{2\epsilon}{e}\frac{1}{N_D}V}. \tag{12.10}$$

Many semiconductor detectors are operated with sufficient reverse bias voltage so that the depletion region extends through the total thickness, creating a fully depleted detector. This configuration is used since charges created in the space charge region can be collected. In contrast, charges created in the non-depleted zone recombine with free majority carriers and are lost. The full depletion voltage V_{dep} is obtained replacing d in Eq. 12.10 with the detector thickness t:

$$V_{\mathrm{dep}} = \frac{eN}{2\epsilon}t^2 \tag{12.11}$$

where N is the dopant concentration on the side of the junction that has lower dopant level. The full depletion voltage is an essential parameter of the sensors. If the applied bias exceeds the full depletion voltage, the device is said to be overdepleted.

12.3 Interaction of particles with silicon

Silicon can be employed for the detection of different kinds of radiation, here focus is given to the interaction of charged particles in the context of high energy physics.

When a charged particle passes through a material, e.g., a semiconductor, the overall significant effect is the production through ionization of many electron-hole pairs along its trajectory. The energy deposition always leads to the creation of equal numbers of holes and electrons, whether the semiconductor is intrinsic or doped. The number of pairs n_{eh} depends on the total energy loss E_{loss} and the average energy expended by the charged particle to produce one pair ε, so that

$$n_{eh} = \frac{E_{\mathrm{loss}}}{\varepsilon}.$$

The average energy loss dE of a charged particle traversing a length dx in a medium

(usually called the *stopping power* $\left\langle -\frac{\mathrm{d}E}{\mathrm{d}x}\right\rangle$ of the material) is described by the *Bethe-Bloch formula*:

$$\left\langle -\frac{\mathrm{d}E}{\mathrm{d}x}\right\rangle = 4\pi N_A r_e^2 m_e c^2 \rho z^2 \frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\frac{2m_e c^2\beta^2\gamma^2 W_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} - \frac{C}{Z}\right] \quad (12.12)$$

where the absorber medium is described by

- I the mean excitation energy,
- Z the atomic number (Z=14 for Si),
- A the atomic mass,
- ρ the density (2.3296 g cm^{-3} for Si) ,
- δ the density effect correction,
- C the shell corrections

and the incident particle by

- z the charge of the incident particle,
- $\beta = v/c$, with v the incident particle velocity
- $\gamma = (1-\beta^2)^{-\frac{1}{2}}$,
- W_{max} the maximum kinetic energy which can be imparted to a free electron in a single collision.

Additionally the first factor in Eq. 12.12, $K = 4\pi N_A r_e^2 m_e c^2 = 0.307\,\mathrm{MeV\,cm^2\,g^{-1}}$, contains the following fundamental constants:

- N_A is the Avogadro's number,
- r_e the classical electron radius,
- m_e the electron mass,
- c the speed of light.

Eq. 12.12 is valid if the mass m of the incident particle is m $\gg m_e$.

Fig. 12.3 shows the stopping power as a function of $\beta\gamma = p/Mc$ for positive muons in copper over nine orders of magnitude in momentum. For the energy of about ten MeV, the $1/\beta^2$ term in Eq. 12.12 is dominant, and the stopping power decreases with increasing energy. At $\beta\gamma \approx 3$, a broad minimum is reached. At higher energies, the logarithmic term leads to a slow rise again, which is eventually cancelled by the density correction. A particle with an energy loss in the minimum of Eq. 12.12 is called a Minimum Ionizing Particle (MIP). All ionizing detectors are based on this process. The sensor design must enable the detection of MIPs with a significant signal/noise (S/N) ratio.

It is important to mention that the energy loss in a finite medium is subject to statistical fluctuations. The number of collisions in a finite medium, as well as the energy transfer per scattering, varies. The first effect can be described by a Poisson distribution, while the latter is described by a function first deduced by Landau. The

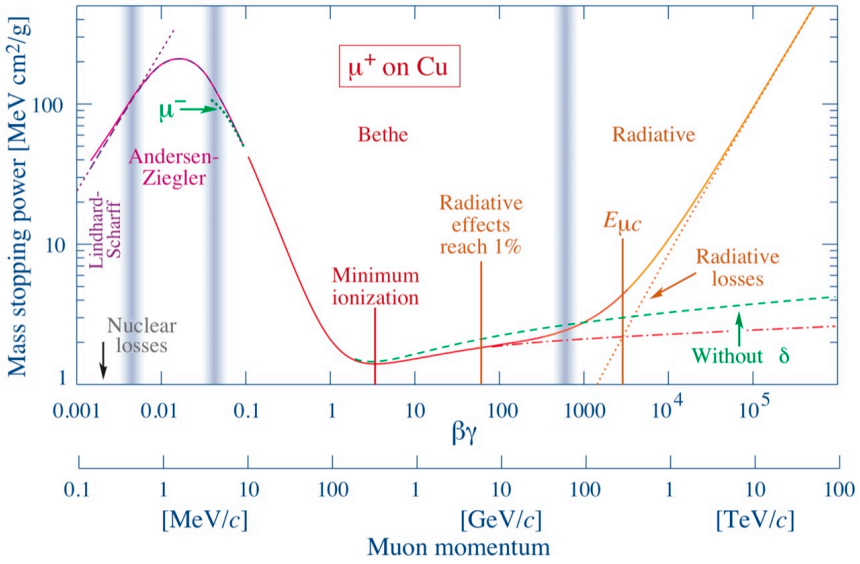


FIGURE 12.3: Mass stopping power ($\left\langle -\frac{dE}{dx} \right\rangle$) for positive muons in copper as a function of $\beta\gamma = p/Mc$ over nine orders of magnitude in momentum (12 orders of magnitude in kinetic energy). Solid curves indicate the total stopping power. Vertical bands indicate boundaries between different approximations discussed in the text. Taken from Ref. [15].

Landau distribution of the charge spectra is shown in Fig. 12.4. In rarer cases, the transferred energy is large, and the so called δ-electrons are created. They are responsible for the long asymmetric tail towards high-charge deposits. The most probable value (MPV) of energy transfer is significantly lower than the average value. For silicon, the average energy used for the creation of one electron-hole pair is 3.6 eV, about three times larger than the band-gap of 1.12 eV, because part of the deposited energy is used for phonon creation. For a MIP, since

$$\left\langle \frac{dE}{dx} \right\rangle \Big|_{min} = 1.664 \text{ MeV g}^{-1} \text{ cm}^2,$$

the most probable number of electron-hole pairs generated in 1 µm of silicon is 76, while the average is 108. This means that a MIP generates a signal of about 11000 electron-hole pairs in sensors of 150 µm thickness. In a volume of $1 \times 1 \times 0.15$ cm^3 of an intrinsic Silicon substrate at room temperature, there are about 2.25×10^8 free charge carriers. The signal created by the 1.1×10^4 electron-hole pairs generated by a MIP is negligible in comparison. The free charge carriers must be reduced by several orders of magnitude to detect the transit of the particle. This can be achieved by cooling to very low temperatures or by depleting the silicon volume of free charge carriers using p- and n-type silicon in a reverse-biased p-n junction configuration described above and is the typical use of silicon as a particle detector.

If the incident particle is an electron, the assumption m $\gg m_e$ is not valid and Eq. 12.12 must be modified accordingly, considering the scattering of two identical and undistinguishable particles. At low energies electrons and positrons primarily lose energy by ionization. While ionization loss rates rise logarithmically with energy, bremsstrahlung (photon emission by an electron accelerated in the nucleus Coulomb field) losses rise nearly linearly, and dominates above a critical energy of a few tens of MeV in most materials. Nevertheless stopping powers for e^-,

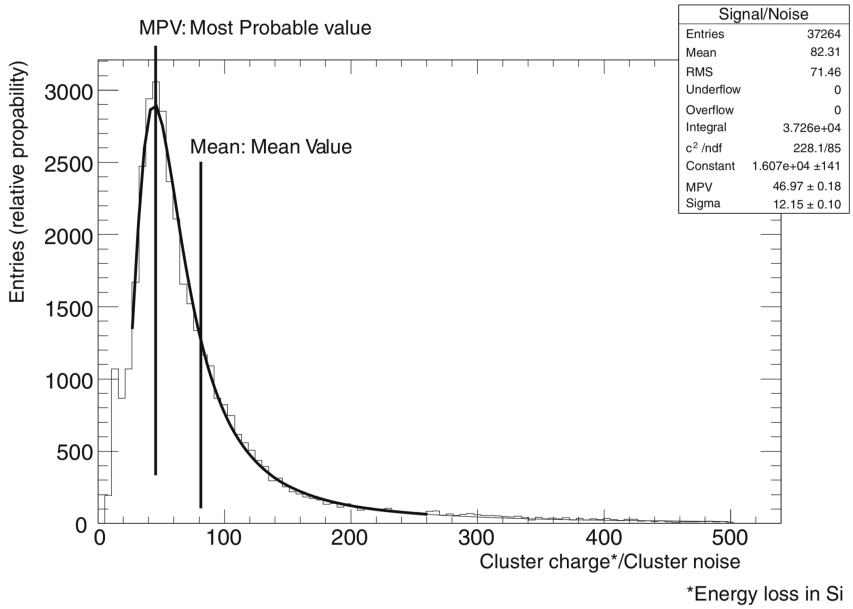


FIGURE 12.4: A Landau distribution. The distribution displays ionized charge over noise in 500 µm silicon from cosmic particles (≈ MIP) arriving at low incident angle in a 3.8 T field. The distribution shows the difference between Mean Value (MV) and Most Probable Value (MPV), also the fairly un-symmetric tail of the distribution from δ-electrons is visible. Taken from Ref. [219].

e^+, and heavy particles in silicon are similar: The minimum value for electrons is 1.50 MeVg^{-1} (at γ = 3.3); for positrons, 1.46 MeVg^{-1} (at γ = 3.7), and for muons, 1.66 MeVg^{-1} (at γ = 3.58) [15].

12.4 Silicon for planar pixel detectors

Silicon has many uses for radiation detection and is the ideal material for tracking applications. This is due to its relatively low atomic number that results in a relatively low energy loss of highly energetic particles (Eq. 12.12). The working principle of silicon based detectors is illustrated in Fig. 12.5, where a charged particle traverses the silicon material creating electron-hole pairs in the depleted volume that drift inside the lattice according to the applied electric field and are collected at the respective electrodes. The drift of the produced charges in the electric field creates the electric signal used to detect the particle, which is already generated when the e-h pairs start to move. The instantaneous current i is given by Ramo's theorem [227]:

$$i = eE_{\mathrm{W}}v \tag{12.13}$$

where v is the drift velocity and E_{W} the *weighting field*, obtained by applying unit potential to the electrode under consideration and zero potential to all others. The weighting field of a pad detector, away from the device edges and assuming the lateral dimensions of the collecting electrode $\gg$ the thickness, is a linear function of the depth. Therefore, the induced charge is the same for any part in the drift path and for e-h pairs created in the middle of the detector, both charge carriers induce the same signal on one electrode.

One of the advantages of using silicon for tracking is the possibility to build

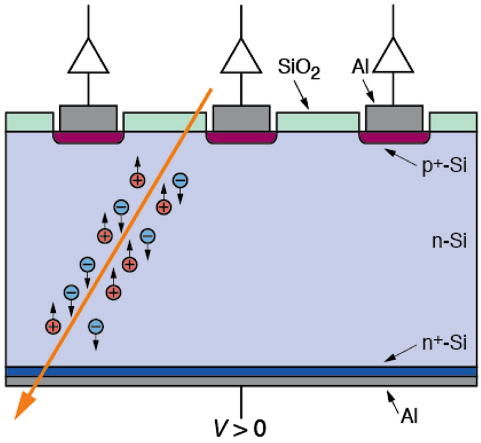


FIGURE 12.5: Cross sectional view of a silicon detector traversed by a charged particle creating electron-hole pairs in the depleted volume that drift inside the lattice according to the applied electric field and are collected at the electrodes. Taken from Ref. [228].

highly granular detectors granting excellent position resolution. In fact, silicon sensors can be segmented in order to achieve information on the particle position when passing through the detector. To obtain unambiguous two-dimensional information, a charge-sensitive detector must provide fine segmentation in both dimensions. The detector electrodes can be segmented to form strips or pixels. In a strip detector, only one spatial coordinate per layer is measured, therefore to obtain both coordinates from the same detector, either both sides have to be segmented, or information from multiple planes has to be combined together. Pixel detectors, in contrast, measure both spatial coordinates on the same side of the sensor so that segmentation is necessary only on one side. In a segmented planar detector, the highly doped electrodes are introduced by implantation or diffusion through a mask to form the strip or pixel electrodes, each of them forming a *p-n* junction. In hybrid detectors, like the ones used in the studies presented in this thesis, the produced silicon sensor has to be connected to the readout chip. The readout chip has the size of the sensor with the same channel pattern. It is is placed on top of the sensor, bump bonded to it.

Due to segmentation, the read-out electrodes dimensions are usually smaller than the thickness and the weighting field differs from the pad detector case. In pixel detectors the last part of the charges drift induces the majority of the signal and charges drifting toward the read-out electrodes provide most of the signal while the ones drifting toward the backplane do not contribute significantly.

Pixel detectors are able to simultaneously meet two fundamental requirements of tracking at modern particle accelerators:

- The possibility of studying short-lived particles: Such particles decay after a few picoseconds in daughter particles which have to be measured close to the interaction point with an accuracy of $\approx 0.1\ c\tau \leqslant 0.3$ mm for a lifetime τ of 1 ps.
- The capability of coping with the increasing interaction rates and energies (and therefore the number of particles produced per collision): accuracy is not the sole relevant parameter. The many other particles produced in the collision complicate the reconstruction and the ability to associate hits to the correct track. Additionally, if the granularity is not fine enough, multiple particles could end up in the same channel.

In HEP experiments, pixel detectors, as reported in Ref. [229], are used for:

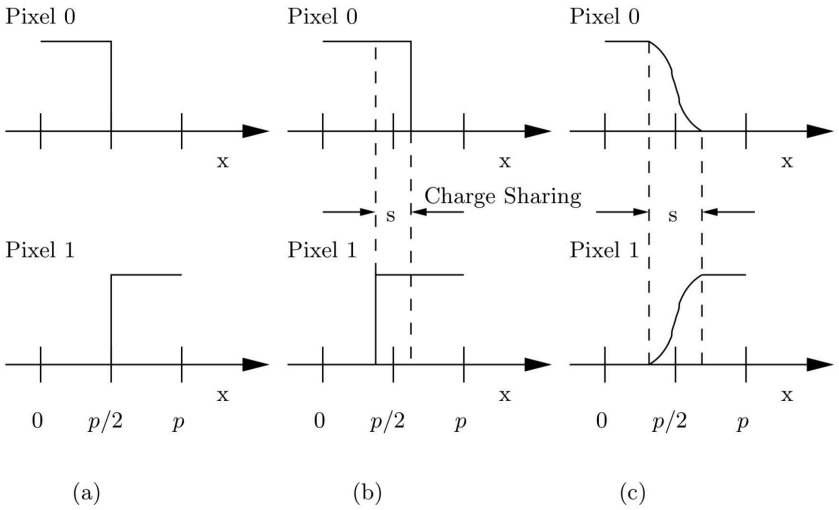


FIGURE 12.6: Signals in two adjacent pixels as function of the impact position x. Detector with binary readout without charge sharing (a), with binary readout with charge sharing (b), and with analog readout (c). Taken from Ref. [221].

- Pattern recognition and identification of particle tracks at large background and pileup levels
- Measurement of primary and secondary vertices;
- Multi-track separation and vertex identification in the core of (boosted) jets;
- Momentum measurement of particles (together with other detectors, like strip detectors);
- Measurement of specific ionization.

12.4.1 Spatial resolution

Among the many factors determining the spatial resolution of a pixel detector, such as the read out mode or the reconstruction algorithm, the pixel pitch is the most significant. In this section the theoretical expectation for the resolution given the pitch size is described while a comparison of measurements obtained from pixel sensors with different segmentations is presented in Sec. 13.1.

The three different configurations in Fig. 12.6 are explained in the following. In all three cases a pixel pitch p is considered centred at the origin of the x axis. The detector is hit by a uniform density of particles.

In Fig. 12.6 (a), the case of a detector with binary readout is shown. In this case no charge sharing among pixels is considered and thus only particles hitting the detector between $\pm p/2$ trigger a signal in Pixel 0. With these assumptions, the average difference between the real impact position and the measured impact position, corresponding to pixel center, is given by the square root of the variance of the uniform distribution: $\sigma = p/\sqrt{12}$.

Examining the case in Fig. 12.6 (b), binary readout in presence of charge sharing, particles hitting the detector in the area indicated by s generate signal in more than one pixels. For the events in which two pixels show a signal from the same

particle the expected resolution is $s/\sqrt{12}$ while if only one pixel is triggered $\sigma = (p-s)/\sqrt{12}$. The resolution is thus improved by charge sharing. In the considered binary readout examples, a pixel could have been only in two states: with or without signal. In Fig. 12.6 (c) the analog readout is considered and pixels show a signal proportional to the collected charge. In the region where more pixels are sharing the signal, different algorithms can be used to reconstruct the particle impact point while when the signal is present in only one pixel the resolution is still limited to $\sigma = (p-s)/\sqrt{12}$. The group of pixels showing a signal from the same particle is called cluster.

The center-of gravity (CoG) algorithm can be applied to reconstruct the position x_{rec} for any number of pixels in the cluster:

$$x_{rec} = \frac{\sum Q_i \cdot x_i}{\sum Q_i} \tag{12.14}$$

where Q_i and x_i are the signal and the center of the i-th pixel in the cluster, respectively. Other reconstruction algorithms are mentioned in Chapter 13.

12.5 Radiation Damage

Silicon sensors are used to detect particles but the prolonged exposition to a source of radiation can damage them. This is especially relevant for pixel detectors that for their high granularity are placed in the innermost region of collider experiments, exposed to high radiation fluences.

It is common practice to divide radiation-induced effects into surface and bulk damage. The pixel modules will be affected by both surface damage, mostly caused by Ionizing Energy Loss in the readout chip, and bulk damages, due to Non Ionizing Energy Loss (NIEL) in the sensor substrate.

12.5.1 Bulk damage

Bulk damage is due to the interaction of incident particles with the silicon lattice atoms. While particles traversing the detector most commonly ionize the material, they also interact with the nuclei via electromagnetic and strong forces. Such interactions are mostly not reversible: Atoms are displaced, and many kinds of defects can originate. Figure 12.7 presents a diagram explaining the most common bulk damage defect types that will be briefly described in the following.

To remove a silicon atom from the lattice, a 25 eV minimum recoil energy is required [230], meaning an electron must have at least 260 keV of energy. At the same time, the heavier protons and neutrons only need 190 eV.

The displacement of an atom of the semiconductor material from its normal lattice site is generated by charged particles interacting with the lattice through the Coulomb force, transferring small energy amounts and thus creating these short distance defects. The vacancy left behind, together with the original atom now at an interstitial position, constitute the Frenkel pair. The void becomes a trapping site for usual charge carriers. These are called point defects to distinguish them from more complex clusters of crystalline damage that are formed along the track of a primary "knock-on" atom if sufficient energy is transferred (about 2 KeV). Typically, electrons and protons, as they interact electromagnetically with the nucleus, cause point defects, while neutrons, tend to create clusters. The different effects of charged and neutral particles can be observed in Fig. 12.8.

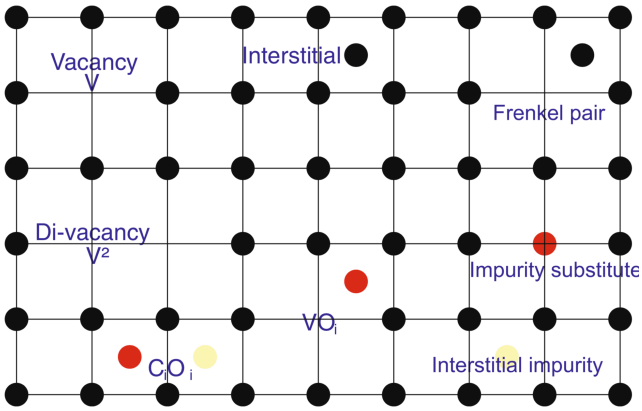


FIGURE 12.7: The figure shows a diagram of the most important typologies of atomic displacements in the lattice after collision with traversing particles. As abbreviation, vacancies are labelled V, interstitials I, di-vacancies V_2. Impurities are labelled with their atomic sign, their index defines their position as substitute or interstitial. Taken from Ref. [219].

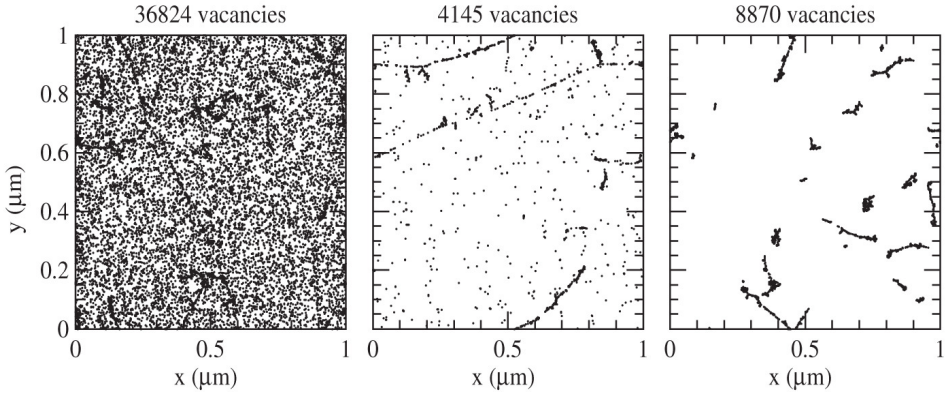


FIGURE 12.8: Simulation of vacancy formation, produced by 10 MeV protons (left), 24 GeV protons (middle) and 1 MeV neutrons (right). The plots are projections over 1 µm of depth and correspond to a fluence of 10^{14} cm^{-2}. Taken from Ref. [231].

The NIEL hypothesis allows to compare the damage caused by the different types of particles with different energies by normalizing the radiation damage. The NIEL is expressed by the displacement damage cross section $D(E)$ defined by [223]

$$D(E) = \sum_i \sigma_i(E_{\text{kin}}) \int_{E_d}^{E_{\text{R,max}}} f_i(E_{\text{kin}}, E_{\text{R}}) P(E_{\text{R}})\, dE_{\text{R}} \tag{12.15}$$

where the sum is over all the possible interactions. σ_i is the cross section of the process i and $f_i(E_{\text{kin}}, E_{\text{R}})$ is the probability of having a collision of a particle with E_{kin}, transferring a recoil energy E_{R}. E_d is the minimum energy required for dislocation damage ($\approx$ 25 eV [224, 231]). Finally, $P(E_{\text{R}})$ is the Lindhard partition function [224], describing the fraction of energy going into silicon atom displacement. For instance, $P(E_{\text{R}}) \sim 50\%$ for 10 MeV protons while $P(E_{\text{R}}) \sim 42\%$ for 24 GeV protons. Also $P(E_{\text{R}}) \sim 43\%$ for 1 MeV neutrons [231].

As standard, the 1 MeV neutron equivalent fluence [$n_{\text{eq}}/\text{cm}^2$] is used and abbreviated with Φ_{eq}. It corresponds to $D_{\text{neutron}}(1\text{MeV})/\text{cm}^2 = 95\,\text{MeVmb}/\text{cm}^2$. This

convention allows to make comparison among radiation damage from different particles and different energies by using a numerical scale factor κ defined as the ratio of the individual damage particle factor for a given energy and 1 MeV neutrons

$$\kappa = \frac{\int D(E)\varphi(E)\,dE}{95\ \text{MeVmb}\ \Phi} = \frac{\Phi_{\text{eq}}}{\Phi}. \tag{12.16}$$

Vacancies, interstitials, and complex clusters create new levels in the band gap of a semiconductor and therefore change its elementary properties. The damage to the lattice created by traversing particles and the following defects diffusion processes must be considered to operate an irradiated sensor. Defects diffusion happens because interstitials and vacancies have high mobility at temperatures $T > 150$ K. As a consequence, Frenkel pairs could recombine, and complex defects could appear. The first phenomenon is short-range and therefore happens with a shorter time constant, while the latter happens with a longer time constant. This temperature-dependent process is called annealing and has a beneficial part that reduces the damage and a degrading one, the reverse annealing part, whose effect is evident also for macroscopic sensor properties as an increase in depletion voltage and leakage current. Beneficial annealing is a short-term process that mitigates the harmful effects of irradiation. Reverse annealing instead takes place for a longer time, further degrading the silicon. As already mentioned, the formation of complex defects produces new energy states in the band-gap, changing the electric properties of the material. There are mainly three macroscopic manifestations of the defects in reverse biased silicon detectors:

- increase of the leakage current
- changes in effective doping concentration and in full depletion voltage
- charge trapping.

12.5.1.1 Leakage current

The formation of mid-gap states facilitates the transition of electrons from the valence to the conduction band leading to an increase of the leakage current in the depletion region. Fig. 12.9 shows the linear dependence of the of leakage current I from the fluence Φ_{eq} received by the silicon sensor. The current variation ΔI normalized to the sensitive volume V is strictly proportional to the irradiation fluence according to [232, 233]:

$$\frac{\Delta I}{V} = \alpha \Phi_{eq}, \tag{12.17}$$

where α, the current-related damage rate, is independent of material type and resistivity. Thanks to the linear increase of the current over several orders of magnitude, diodes can be used to determine the particle fluence by the increase of current.

The leakage current is also strongly dependent on the temperature T according to:

$$I \propto T^2 \mathrm{e}^{-\frac{E_a}{2k_B T}} \tag{12.18}$$

where E_a is the activation energy (1.19–1.21 eV [234]). Thus, to keep the leakage current low, irradiated detectors should be operated and stored at low temperatures.

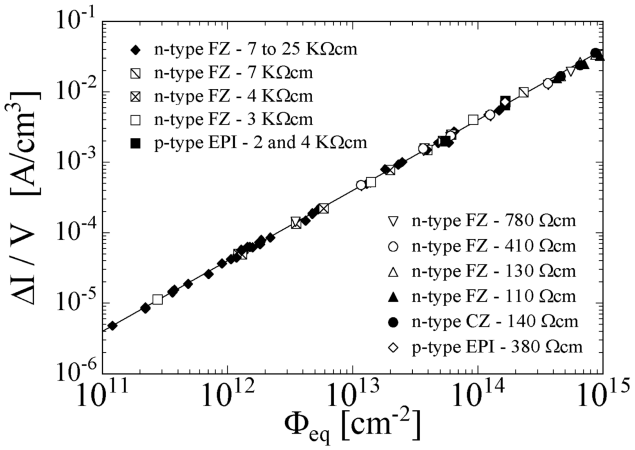
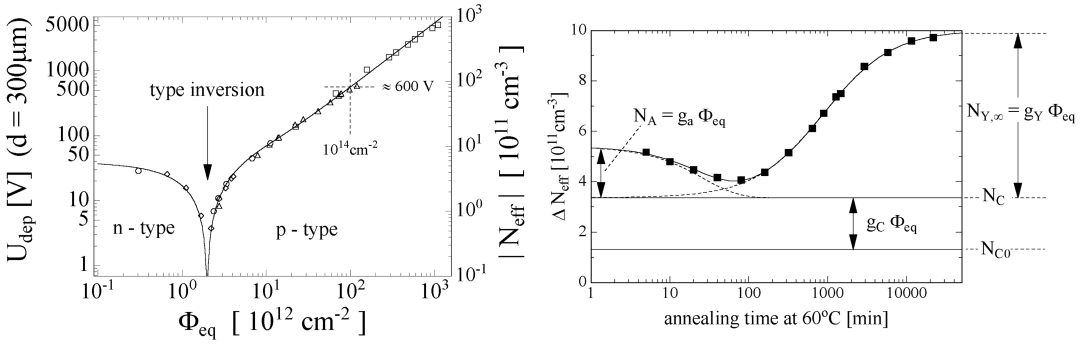


FIGURE 12.9: Fluence dependence of leakage current for silicon detectors produced by various process technologies from different silicon materials. Taken from Ref. [223].



(A) (B)

FIGURE 12.10: Left: Change in the depletion voltage respectively the absolute effective doping concentration as measured immediately after irradiationas a function of fluence. Taken from Ref. [235]. Right: Annealing behaviour of the radiation induced charge in the effective doping concentration ΔN_{eff} at 60 °C for a fluence of $1.4 \times 10^{13}\text{cm}^{-2}$. Taken from Ref. [223].

12.5.1.2 Effect on doping concentration and bias voltage

As a consequence of irradiation, the doping concentration of the sensor may vary. For instance, donors could be removed and acceptor-like defects with a negative space charge could be generated throughout the bulk, leading to variations in the effective doping concentration $N_{eff} = |N_D - N_A|$.

Figure 12.10a shows the fluence dependence of the effective doping concentration and the full depletion voltage. Starting with an *n*-doped material, the effective doping concentration decreases until the number of donors and acceptors are equal and a type-inversion occurs. At the inversion point, the effective doping concentration is zero and silicon behaves as if it were intrinsic. With further irradiation, the acceptors begin to dominate, and the bulk material is now effectively of *p*-type. Following the sign inversion of the space charge, the *p*-*n* junction moves from the p^+side of the sensor to the n^+side and the depletion zone grows from there.

As it was shown in Eq. 12.11, the full depletion voltage depends on the doping concentration: following the changes of N_{eff} in Fig. 12.10a, initially, the depletion voltage decreases to theoretically zero at the inversion point, and then rises again with the effective bulk doping concentration. Therefore, the bias voltage must be adjusted during the irradiation process to ensure full depletion. The fluence at which the inversion takes place depends on the initial doping concentration. For instance, high-resistivity sensors have a low initial donor density and reach the inversion point with less fluence than those of low resistivity.

Figure 12.10b displays a complex annealing behaviour, commonly described by the *Hamburg model* [223, 236, 237]:

$$\begin{aligned} N_{eff} &= N_{eff,\Phi=0} - \Delta N_{eff}(\Phi, T_a, t) \\ &= N_{eff,\Phi=0} - [N_C(\Phi) + N_a(\Phi, T_a, t) + N_Y(\Phi, T_a, t)] \end{aligned} \tag{12.19}$$

where N_C describes the fluence dependence of the effective doping, and the other two terms describe the change of the effective doping after the irradiation and are therefore also dependent on the storage temperature T_a and time t. After irradiation, as shown in Fig. 12.10b, the effective doping concentration decreases to a minimum, reached after about one week at room temperature (beneficial annealing), and slowly increases afterwards (reverse annealing). In order to slow down this increase, irradiated detectors have to be kept cool also outside running periods.

12.5.1.3 Charge trapping effect

The additional energy levels caused by the crystal defects can be filled by carriers produced by traversing particles. If these carriers are trapped for a time longer than the collection time, they do not contribute to the signal and are lost, leading to a degradation of the Charge Collection Efficiency (CCE). The CCE is defined as the ratio of the total charge observed to the actual deposited charge. A high concentration of trapping centres also reduces the CCE. The trapping time τ_i (with i = n, p for electrons and holes, respectively) characteristic of the trapping process related to the defect type t is given by [238]

$$\frac{1}{\tau_i} = c_i(1 - f_t)N_t,$$

where c_i is the capture coefficient, f_t is the defect occupancy and N_t the defect concentration. Trapping can be reduced by collecting electrons instead of holes because they have higher mobility, resulting in a higher velocity, and are thus less prone to trapping. In addition, the charge collection efficiency can be partially restored by applying a higher bias voltage. So, while the efficiency curve of non-irradiated detectors reaches its plateau at the full depletion voltage, irradiated sensors need considerable "over biasing" beyond this voltage. However, increasing the bias might reduce the charge sharing between neighbouring pixels and thus reduce the accuracy of position measurements.

12.5.1.4 Surface Damage

The term surface damage summarizes all defects in the covering dielectrics, such as the Silicon Dioxide and the Silicon-dielectric interface. Since the crystal structure of SiO_2 is highly irregular, displacement of single atoms due to irradiation does not lead to macroscopic changes. Ionization in the oxide, however, is not fully reversible

and can cause steady changes in the interface properties. One consequence is the growth of a positive fixed oxide charge, which saturates after some kGy. This oxide charge changes the electric field in the silicon bulk close to the surface and may lead to an electric breakdown. A further effect of radiation is the generation of interface states leading to a surface-generated current when the space charge region reaches the surface.

12.6 Sensor prototypes

The sensors probed for this study [239] are planar n^+p sensors with a p^+ backside implant produced by Hamamatsu Photonics K.K. (HPK) [240]. They have an active thickness of 150 µm and a pixel size of $100\times25\,\mu m^2$; both p-stop and p-spray technology were tested for the pixel isolation. The depletion voltage is 75 V. Several design variants have been produced to identify the ones satisfying the Phase-2 requirements. More studies focusing on fundamental sensor properties like hit efficiency and signal-to-noise ratio have also been performed and the results are presented in Ref. [241–243].

All sensors used in this study have been bump-bonded to PSI ROC4Sens readout chips [244]. The ROC4Sens chip, developed at the Paul Scherrer Institute in Switzerland based on the PSI46 chip [96] (IBM 250 nm), allows for detailed charge collection studies due to its nonsparsified readout. The details of the ROC4Sens chip are presented in Sec. 12.6.1.

Modules constituted of the sensor bump-bonded to the readout chip are glued to a 1.56 mm thick Printed Circuit Board (PCB). To reduce the material, an open window is present in the PCB behind non-irradiated sensors. For irradiated modules instead, 50 µm of Copper are deposited on the back of the PCB to ensure the thermal contact with the cooling elements.

The readout chip's staggered $50\times50\,\mu m^2$ bump-bond pattern matches the designs in Fig. 12.11, where five top views of 2×2 pixel cells of the designs used in the resolution measurements presented in this study are shown. In Figures 12.11a and 12.11b (12.11d and 12.11e) the variants have p-stop (p-spray) isolation. The designs in Figs. 12.11b and 12.11e have structures intended to mimic the routing of the RD53A [245] chip, the prototype for the final CMS readout chip (not covered in this study, more studies with the RD53A demonstrator chip can be found in Ref. [243]). The design in Fig. 12.11c features a common punch-through, to simultaneously bias the 4 shown pixel cells.

The thin pixel sensors used in this work are produced with the two following substrate options

- Physical thinned (FTH): A sensor with standard thickness is thinned to 150 µm;
- Si-Si direct bonded (FDB): A high resistivity float zone wafer and a low resistivity handle wafer are bonded together. The float zone (handle) wafer is thinned down to an active thickness of 150 µm (50 µm), resulting in a total thickness of 200 µm.

Sensors with the five variants of Fig. 12.11 were used in groups of three sensors to compose the *Dreimaster*, as described in the next chapters. Table 12.1 summarizes the sensor combinations occurred in the studies presented in this thesis.

The proton irradiation to $\phi_{eq} = 2.1 \times 10^{15}\,cm^{-2}$ was performed at the PS-IRRAD Proton Facility at CERN [246] with a beam momentum of 24 GeV. The neutron irradiation to $\phi_{eq} = 3.6 \times 10^{15}\,cm^{-2}$ was performed in the TRIGA Mark II reactor in

upstream reference		DUT		Irradiation	downstream reference		Resolution	Measurement	Runs (Optimal angle)
Nr.	Design	Nr.	Design	$[\phi_{eq}/10^{15}\,\mathrm{cm}^{-2}]$	Nr.	Design	[µm]		
148	P-stop RD53A routing (FDB)	150	P-stop basic (FDB)	0	163	P-spray RD53A routing (FDB)	3.18 ± 0.02	"small" angle scan	2775-2781 (2779)
109	P-stop basic (FTH)	148	P-stop RD53A routing (FDB)	0	110	P-stop common punch-through (FTH)	3.23 ± 0.02	Angle scan	2735 - 2760 (2743)
148	P-stop RD53A routing (FDB)	146	P-spray basic (FDB)	0	163	P-spray RD53A routing (FDB)	3.22 ± 0.03	"small" angle scan	2769-2773 (2773)
148	P-stop RD53A routing (FDB)	163	P-spray RD53A routing (FDB)	0	150	P-stop basic (FDB)	3.26 ± 0.03	Momentum scan	2832 -2840 (2832)
148	P-stop RD53A routing (FDB)	120	P-stop basic (FTH)	2.1, proton	163	P-spray RD53A routing (FDB)	5.00 ± 0.02	Angle scan	2793-2817 (2801)
148	P-stop RD53A routing (FDB)	130	P-stop basic (FDB)	2.1, proton	109	P-stop basic (FTH)	5.6 ± 0.1	Momentum scan	1800, 1810-1821 (1820)
148	P-stop RD53A routing (FDB)	194	P-stop basic (FDB)	3.6, neutron	150	P-stop basic (FDB)	5.00 ± 0.02	Angle and bias scans	3763-3839 (3839)

TABLE 12.1: The table lists the spatial resolution of the design variants described in Fig. 12.11 at the optimal angle for resolution. Each sensor is uniquely identified by the listed number. The beam momentum was 5.6 GeV for all the measurements reported in the table. All non-irradiated sensors were operated at 120 V. "Small" angle scan refers to an angle scan performed around the optimal angle for resolution, to identify it.

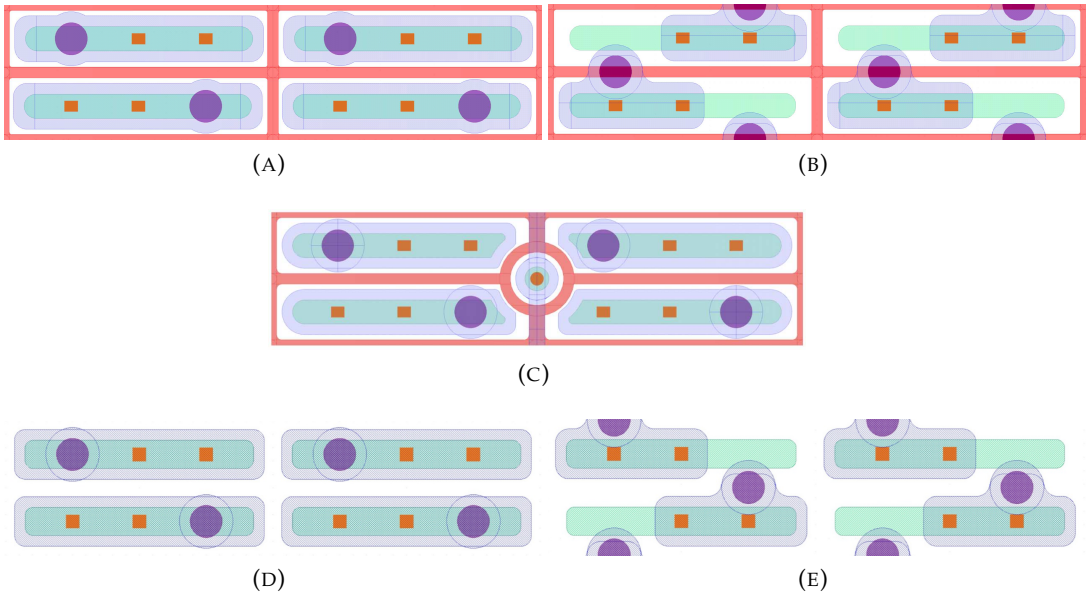


FIGURE 12.11: Top views of 2×2 pixel cells of the designs used for the resolution measurements. The bump-bond pads (purple circles) are arranged in a staggered pattern matching the ROC4Sens readout chip pattern. Red areas represent the p-stop, green dashed areas are n^+ implants and the purple dashed areas are made of metal. The orange squares identify the connection between the metal and the implant. The schemes in Fig. 12.11a and Fig. 12.11d represent a basic design, with p-stop and p-spray isolations, respectively. The designs Fig. 12.11b and Fig. 12.11e have structures intended to mimic the routing of the RD53A chip, the prototype for the final CMS readout chip, with p-stop and p-spray isolations, respectively. Fig. 12.11c has a common punch-through.

Ljubljana [247]. In both cases the readout chip was irradiated together with the sensor to which it was bump-bonded.

To prevent annealing, the sensors are stored cold, except for irradiation, transport and handling. To limit the leakage current during data taking, irradiated sensors were cooled to $\approx -24\,^{\circ}\mathrm{C}$.

12.6.1 ROC4Sens readout chip

All sensors used in this study have been bump bonded to PSI ROC4Sens read-out chips [244]. The ROC4Sens chip, developed at the Paul Scherrer Institute in Switzerland, allows for detailed studies of the sensor properties and of sensor irradiation tests, thanks to its high radiation tolerance.

It covers an area of 7.848 mm $\times$ 9.778 mm and was designed in the same 250 nm CMOS technology employed also for its predecessor, the PSI46 chip, which proved a good radiation tolerance [248]. The pixel array comprises 155 pixels along the short edge (columns) and 160 pixels along the long edge (rows), for a total of 24800 pixels. The ROC4Sens chip has a thickness of 700 µm. The pitch is $50 \times 50\,\mu\mathrm{m}^2$ and each pixel hosts an octagonal bump bond pad with a 15 µm width. The staggered bump bond pads allows also sensors with a $100 \times 25\,\mu\mathrm{m}^2$ pitch to be bump-bonded to it. The main characteristics of the ROC4Sens chip are the absence of on-pixel zero-suppression, to allow for unbiased studies of the charge collection on sensors, and an analogue measurement of the pulse height. Moreover, it is equipped with a fast pixel front end amplifier and a mechanism to inject calibration pulses.

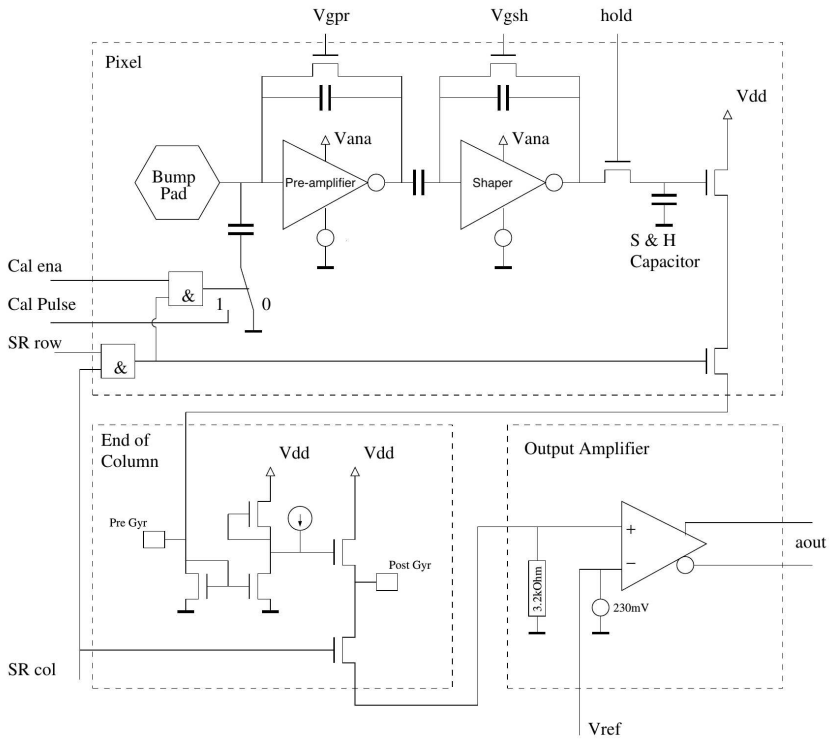


FIGURE 12.12: The ROC4Sens pixel readout scheme. The main elements that can be observed are the signal pre-amplifier and shaper, the sample-and-hold capacitor and the shift registers for the pixel selection. The place of injection of the calibration pulse is also shown. Taken from Ref. [244].

As it is designed for sensor testing, it has no on-chip programming interface to minimise the circuit blocks to be recalibrated after irradiation and a readout targeting a frame rate of 1 kHz to collect large event samples in beam tests. A schematic drawing of the ROC4Sens pixel is presented in Fig. 12.12. On each pixel, the signal is processed by a pre-amplifier and a shaper, ensuring a fast pulse shaping, necessary for the conditions at the LHC. The full signal can be stored for readout in the sample-and-hold (S & H) capacitor. The storage of a hit is triggered from outside the chip and the trigger signal is distributed to all pixels simultaneously. Once a hit is being stored, the pixel cannot accept new incoming hits.

Two separate Shift Registers (SR row and SR col) are employed to select the pixels for read out and calibration.

All pixels of the same column are connected to the same column bus and amplifier at the end of the column.

Vana and Vdig are the supply voltages of the analogue circuit and the digital logic on each pixel, respectively. In idle, at the nominal voltage Vdig $\approx$ 2.2 V the current consumption of the digital logic is 35 mA. For the nominal voltage Vana $\approx$ 1.9 V, the current consumption of the analogue circuit is 150 mA, corresponding to a current consumption of about 6 µA per pixel, evenly split between the shaper and the pre-amplifier.

13 Spatial resolution measurements

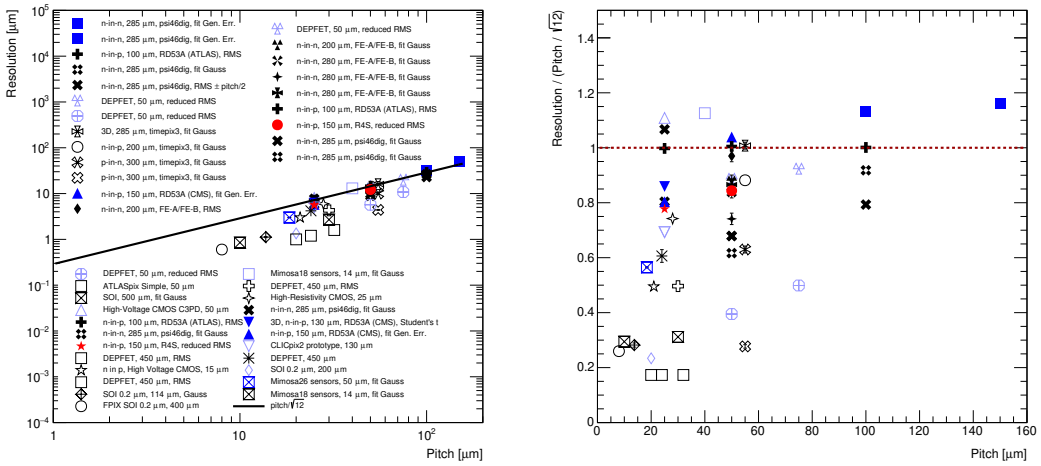
Pixelated silicon detectors are state-of-the-art technology to achieve precise tracking and vertexing at collider experiments, designed to accurately measure the hit position of incoming particles in high rate and radiation environments. In the last decades, the pixel pitch has constantly been reduced to cope with the experiments' needs of achieving higher position resolution and maintaining low pixel occupancy per channel. Thus, when designing and testing sensor prototypes, their spatial resolution is one of the key quantities to establish their suitability for the future employment in the target experiment. The sensor spatial resolution has not been uniquely defined in literature and different techniques can be adopted for its measurement. In Sec. 13.1 an overview of resolution measurements in literature is introduced and the limiting factors to the precision of the position measurement in silicon sensors are presented. Then, the method used in this thesis work is described in Sec. 13.2. The optimization of the track selection criteria and the definition of spatial resolution is then discussed. Finally, the final choices applied to the results in Chapter 14 are summarized.

The content of this and the next chapter has been collected and adapted by the author of this thesis to result in the public note in Ref. [249] and has been accepted by the Nuclear Instruments and Methods in Physics Research (Section A).

The author of this work took part in the data taking, performed the presented data analysis and the reported literature review. The work was performed under the supervision of Prof. Dr. Erika Garutti, in collaboration with Dr. Andreas Hinzmann and Dr. Daniel Pitzl. Finn Feindt developed the online acquisition software and some of the data corrections applied in this work. Dr. Jörn Schwandt designed the sensors. Dr. Georg Steinbrück organized the test beam periods. Dr. Aliakbar Ebrahimi, Finn Feindt, Prof. Dr. Erika Garutti, Dr. Paolo Gunnellini, Caroline Niemeyer, Dr. Daniel Pitzl, Dr. Jörn Schwandt, Dr. Georg Steinbrück and the author of this thesis contributed to the data taking. Prof Emer. Dr. Robert Klanner contributed to the discussion about validity and limitations of the employed method.

13.1 Measurements in literature

The precision of the particle's position measurement is determined by the segmentation of the detector used. Figure 13.1 presents spatial resolution measurements collected from a selection of the available literature performed using pixel detectors of different designs and dimensions, ordered by decreasing pitch sizes. The beam was perpendicular to the sensors for all measurements. It shows that reducing the pixel size greatly improves the resolution, which in many cases is better than the binary resolution of pitch $/\sqrt{12}$. Full markers refer to hybrid detectors, where the pixel sensor is bump-bonded to a readout chip. This constitutes a limitation on the dimension of the pixel size: each channel has to be individually connected through bump-bonding to the corresponding channel of the readout chip. The reduction of



(A) (B)

FIGURE 13.1: Left: Selection of spatial resolution measurements from a selection of the available literature performed using non-irradiated silicon pixel sensors of different designs and dimensions. The measurements are ordered by decreasing pitch sizes. The legend reports the bulk type, the thickness, the readout chip (when applicable), and the residual width definition. When in a paper more measurements are available, e.g. different cluster sizes or algorithms used, the most simple and general one is reported here. Full (empty) markers refer to measurements performed with hybrid (monolithic) detectors. Black markers refer to hadronic high momentum beams, blue to low momentum leptonic beams, and light blue to unknown beam conditions. The red star (circle) refers to the measurement presented in this thesis (obtained on samples of the same submission but different pitch). The solid black line is the binary resolution of pitch $/\sqrt{12}$. More details can be found in Table 13.1. Right: Same measurements, with the resolution divided by the binary resolution (pitch $/\sqrt{12}$) to enhance the differences between measurements. Published in Ref. [249].

the pixel cell area is therefore limited by the achievable logic density of the electronic needed to amplify, discriminate, and process the hit information [229]. The 65 nm CMOS technology achieved a pixel size of $50 \times 50\,\mu m^2$ in the readout chip, allowing the realization of $50 \times 50\,\mu m^2$ and $100 \times 25\,\mu m^2$ sensor pixel cells, whose area is six times smaller than the one used in the current CMS pixel detector [98]. The ultimate limit to the size reduction due to bump bonding is considered to be at 5-10 µm [250]. These challenges can be avoided by employing monolithic detectors (empty markers in Fig. 13.1) that are able to allocate high density CMOS circuitry in smaller pixel cells and therefore achieve an outstanding spatial resolution. The radiation hardness of such devices has been recently improved [251].

Besides the size of the readout electronics, additional effects limit the reduction of the pixel pitch and the consequent improvement of the spatial resolution. Pixel pitches much smaller than the distance in which charges spread due to diffusion (4–8 µm for typical 150–200 µm thickness) would result in excessive charge sharing [229]. Furthermore, simulation studies [252] on $2 \times 2 \times 10\,\mu m^3$ pixel cells show that the resolution becomes worse than pitch $/\sqrt{12}$ as the probability of a minimum ionizing particle traversing the pixel without any charge deposition is non-negligible with such small pixels, reducing the cluster size. The hit position resolution is additionally limited to approximately 1 µm by the presence of delta electrons [253]. Submicron precision is expected to be achieved, despite delta electrons, with stacked

sensors [254, 255] in applications where multiple scattering in the detector material does not affect the measurements.

Legend	Ref.	Pixel cell [μm^2]	Res. [µm]	Def.	Cluster size	Sensor type	Thickness [µm]	Bias Voltage [V]	Threshold [ke]	S/N	Hit Algorithm	Unc. [µm]	Beam	TB Setup	ROC	T [°C]
■	[256]	100×**150**	50.2	Generalized error function	all	n^+-in-n, planar	285	200	2.0	-	CoG (‡)	0.08 (stat), 0.28 (syst) (*)	DESY, e, 5.2 GeV	EUDET, 3.4 µm [257]	PSI46digV2.1-r	17
■	[256]	**100**×150	32.7	Generalized error function	all	n^+-in-n, planar	285	200	2.0	-	CoG (‡)	0.09 (stat), 0.21 (syst) (*)	DESY, e, 5.2 GeV	EUDET, 3.4 µm [257]	PSI46digV2.1-r	17
✚	[258]	25×**100**	∼ 28.9	RMS	1	n-in-p, planar	150	50	1.5	-	Pixel center	-	CERN, π, k and p, 120 GeV	EUDET, 6.8µm [257]	RD53A (LIN FE)	-
❖	[259]	**100**×150	26.5	Gauss	all	n-in-n, planar	285	150	2.4	-	CoG	-	FNAL, p, 120 GeV	Pixel telescope [260], ∼ 8 µm	PSI46digV2.1-r	RT
✖	[261]	**100**×150	22.9	RMS in ± 1/2 pitch	all	n-in-n, planar	285	150	∼ 1.5	-	CoG	-	FNAL, p, 120 GeV	Pixel telescope [260], ∼ 6 µm	PSI46digV2.1-r	RT
△	[262]	**75**×50	20	Reduced RMS, N=2	1	DEPFET	50	-	- (⋄)	≈ 40	CoG +η	-	(§)	EUDET/AIDA [263]	-	-
⊕	[262]	**75**×50	10.8	Reduced RMS, N=2	2	DEPFET	50	-	- (⋄)	≈ 40	CoG +η	-	(§)	EUDET/AIDA [263]	-	-
✠	[264]	**55**×**55**	16	Gauss	all	n, 3D	285	10	1.52	-	CoG+η	0.2	CERN (π 120 GeV)	(⊖)	Timepix3 ASIC (ToT)	RT-
○	[265]	**55**×**55**	14	Gauss	-	n-on-p, planar	200	200	-	-	-	SPS, CERN		(⊖)	Timepix3 ASIC (ToT)	-
❋	[264]	**55**×**55**	10	Gauss	all	p^+-on-n, planar	300	100	1.52	-	CoG+η	0.1	CERN (π 120 GeV)	(⊖)	Timepix3 ASIC (ToT)	-
◇	[264]	**55**×**55**	4.4	Gauss	all	p^+-on-n, planar	300	10	1.52	-	CoG+η	0.2	CERN (π 120 GeV)	(⊖)	Timepix3 ASIC (ToT)	-
▲	[243]	**50**×**50**	15	Generalized error function	all	n-in-p, planar	150	120	0.7	-	CoG	-	DESY, e, 5.2 GeV	EUDET, 3.4 µm [257]	RD53A	RT
♦	[266]	**50**×400	14 (⊞)	RMS	all	n-in-n, planar (SSGb)	200	150	∼ 3.0	-	CoG	0.3 (stat)	CERN, π, 180 GeV	microstrip 3-5 µm	FE-A/FE-B	-9
△	[262]	75 or 50×**50**	12.8	Reduced RMS, N=2	1	DEPFET	50	-	- (⋄)	≈ 40	CoG +η	-	(§)	EUDET/AIDA [263]	-	-
▲	[266]	**50**×400	12.7 (⊞)	Gauss	all	n-in-n, planar (SSGb)	200	150	∼ 3.0	-	CoG	0.3 (stat)	CERN, π, 180 GeV	microstrip 3-5 µm	FE-A/FE-B	-9
❖	[266]	**50**×400	12.1 (⊞)	Gauss	all	n-in-n, planar (ST2)	280	150	∼ 3.0	-	CoG	0.3 (stat)	CERN, π, 180 GeV	microstrip 3-5 µm	FE-A/FE-B	-9
✦	[266]	**50**×400	10.7 (⊞)	Gauss	all	n-in-n, planar (ST1)	280	150	∼ 3.0	-	CoG	0.3 (stat)	CERN, π, 180 GeV	microstrip 3-5 µm	FE-A/FE-B	-9
✠	[266]	**50**×400	10.5 (⊞)	Gauss	all	n-in-n, planar (SSG)	280	150	∼ 3.0	-	CoG	0.3 (stat)	CERN, π, 180 GeV	microstrip 3-5 µm	FE-A/FE-B	-9
✚	[258]	**50**×**50**	∼ 14.5	RMS	1	n-in-p, planar	100	50	1.0	-	Pixel center	-	CERN, π, k and p, 120 GeV	EUDET, 6.7µm [257]	RD53A (LIN FE)	-
●	[242]	**50**×**50**	12.18	Reduced RMS, N=6	all	n-in-p, planar	150	120	∼ 0.4	100	CoG	0.04 (stat)	DESY, e, 5.2 GeV	EUDET [257]	ROC4Sens	RT
✖	[261]	**50**×300	9.8	Gauss	all	n-in-n, planar	285	150	∼ 1.5	-	CoG	-	FNAL, p, 120 GeV	Pixel telescope [260], ∼ 6 µm	PSI46digV2.1-r	RT
❖	[259]	**50**×300	8.9	Gauss	all	n-in-n, planar	285	150	2.4	-	CoG	-	FNAL, p, 120 GeV	Pixel telescope [260], ∼ 8 µm	PSI46digV2.1-r	RT
⊕	[262]	75 or 50×**50**	5.7	Reduced RMS, N=2	2	DEPFET	50	-	- (⋄)	≈ 40	CoG +η	-	(§)	EUDET/AIDA [263]	-	-
□	[267]	130×**40**	13	-	-	(⊘)	-	-	-	-	-	-	-	-	ATLASpix_Simple	-
□	[268]	**32**×24	1.6	RMS	2×2	DEPFET	450	200	-	120	CoG +η	0.1	CERN, π, 120 GeV	6 sensors telescope	-	RT
✚	[269] (⊛)	**30**×**30**	4.3	Gauss	all	SOI sensors	500	130	(●)	≈ 100	CoG	0.1	CERN, π, 120 GeV	(⊖)	-	RT
⊠	[270]	**30**×**30**	2.7	Gauss	all	MimoTel	14 (⊙)	-	-	-	-	-	CERN, π, 120 GeV	EUDET	-	RT
✧	[271]	**28**×**28**	∼ 6	RMS In range ± 20 µm	all	High-Resistivity CMOS	-	6	0.05 (⊛)	-	CoG + η	-	CERN π, k and p, 120 GeV	(⊖)	Investigator chip	-
△	[267]	**25**×**25**	8	-	-	High-Voltage CMOS C3PD	50 -	-	-	-	-	-	-	-	-	-
✖	[261]	**25**×600	7.7	Gauss	all	n-in-n, planar	285	150	∼ 2.0	-	CoG	-	FNAL, p, 120 GeV	Pixel telescope [260], ∼ 6 µm	PSI46digV2.1-r	RT
✚	[258]	**25**×100	∼ 7.2	RMS	1	n-in-p, planar	150	50	1.5	-	Pixel center	-	CERN, π, k and p, 120 GeV	EUDET, 6.8µm [257]	RD53A (LIN FE)	-
▼	[272]	**25**×100	∼ 6.2	Student-t	all	p DRIE column, 3D	130	30	0.9	-	CoG	-	DESY, e, 5.2 GeV	EUDET [257], 3.8-6.2 µm	RD53A	RT
❖	[259]	**25**×600	5.8	Gauss	all	n-in-n, planar	285	150	2.4	-	CoG	-	FNAL, p, 120 GeV	Pixel telescope [260], ∼ 8 µm	PSI46digV2.1-r	RT
▲	[243]	**25**×100	5.8	Generalized error function	all	n-in-p, planar	150	120	1.0	-	CoG	-	DESY, e, 5.2 GeV	EUDET, 3.4 µm [257]	RD53A	RT
★	This thesis	**25**×100	5.61	Reduced RMS, N=6	all	n-in-p, planar	150	120	∼ 0.4	77 ± 6	CoG	0.03 (stat)	DESY, e, 5.6 GeV	Dreimaster	ROC4Sens	RT
▽	[273]	**25**×**25**	∼ 5	-	-	CLICpix2 prototype	130	-	-	-	-	-	-	-	(÷)	-
✳	[274]	**24**×**24**	4.2	-	-	DEPFET	450	200	-	110	CoG (†)	0.16	CERN, π, 180 GeV	5 sensors	-	-
□	[268]	32×**24**	1.2	RMS	2×2	DEPFET	450	200	-	120	CoG +η	0.1	CERN, π, 120 GeV	6 sensors telescope	-	RT
◊	[275]	**21**×**21**	∼ 3	Gauss	-	(◁)	15	60-100	-	27	-	-	CERN SPS	Eudet telescope 2.3 [257]	-	-
☆	[276]	**20**×**20**	1.35	-	5×5	SOI 0.2 µm	500	-	-	300	-	-	-	-	-	-
□	[268]	**20**×**20**	1	RMS	3×3	DEPFET	450	200	-	200	CoG +η	0.1	CERN, π, 120 GeV	6 sensors telescope	-	RT
⊠	[257]	**18.4**×**18.4**	∼ 2-3	Gauss	all	(⊡)	50	-	(#)	-	-	0.01 (stat), 0.08 (syst)	DESY, e, 6 GeV	EUDET DATURA [257]	-	18
⊕	[277]	**13.75**×**13.75**	1.12	Gauss	>1	SOI 0.2 µm	114±6	60	($)	(⋆)	CoG+η	0.03 (stat+syst)	CERN, π, 200 GeV	3 SOImager-2 sensors	SOImager-2	RT
⊠	[270]	**10**×**10**	0.85	Gauss	all	Mimosa18	14 (⊙)	-	-	-	-	-	CERN, π, 120 GeV	EUDET	-	RT
○	[278]	**8**×**8**	0.6	Gauss	-	FPIX SOI 0.2 µm	400	70	-	310	CoG	0.01	FNAL, proton, 120 GeV	3 FPIX telescope	-	-

TABLE 13.1: Spatial resolution measurements at vertical incidence for non-irradiated silicon pixel sensors. When in a paper more measurements are available, e.g. different cluster sizes or algorithms used, the most simple and general one is reported here. Bold in the pitch column highlights the direction of the measurement. The track fitting algorithm has been omitted as the procedures could not be shortly summarized in the table and we refer to the primary papers for this information.

(‡) Measurement using cluster skewness as parametrization are also described and improve the position resolution significantly; (*) For measurement at 27.1°. These measurements were not optimized for data taking at vertical incidence, but at specific incidence angles matching the conditions of the target experiment; (§) CERN (π, 120 GeV) / DESY (e, 4 GeV); (⊖) Timepix3 Telescope, 2.3 µm [279]; (⊞) Telescope extrapolation uncertainty not subtracted; (⊛) Later publication [280] with generalization of the standard η-correction adapted for arbitrary cluster sizes obtains 1.5 µm for the FZ-n wafer and about 3.0 µm for the Double SOI Czochralski type p wafer; (⊘) Monolithic High-Voltage CMOS; (†) Measurements with η corrections and laser calibration are also present. Spatial resolution has been further improved, in a later publication 2D correction and energy split correction are added [281]; (●) Two Seed Method (TSM). (⊛) Analysis threshold, triggering threshold is 0.15 ke; (÷) 5-bit time-over-threshold (ToT); (◁) n in p, High Voltage CMOS; (⊡) AMS 350 nm CMOS technology; (#) Charge in a single pixel ≥ 6×noise; ($) Double Threshold method. (⋆) seed 5 + neighbour 3; (⋄) The seed pixel threshold is indicated as a multiple of the pixel noise, the threshold on the neighbouring pixels is 12 times the noise; (⊙) Active thickness. The total thickness is 700 µm. Published in Ref. [249].

13.2 Spatial resolution measurement method

This part of the thesis work aims to characterize the position resolution along the 25 µm pitch of $100 \times 25\,\mu m^2$ pixel sensor candidates for the Inner Tracker (IT) Phase-2 upgrade of the CMS detector. The sensors are introduced in Sec. 12.6.

13.2.1 Motivation

The spatial resolution benefits from a reduced pixel size but it deteriorates with radiation damage. The pixel modules will be affected by both surface damage, mostly caused by ionizing energy loss in the readout chip, and bulk damage, due to non-ionizing energy loss in the sensor substrate. The creation of bulk defects changes the sensor macroscopic properties and reduces the charge collection efficiency. To choose suitable candidates for the Phase-2 upgrade, it is mandatory to evaluate the performance of irradiated sensors. Since radiation damage varies for different particle types, both proton and neutron irradiation have been performed. The challenge of these measurements is to provide high resolution tracking and, at the same time, the possibility to operate the irradiated sensors at low temperature ($\approx -30\,^{\circ}C$) as in the experiment.

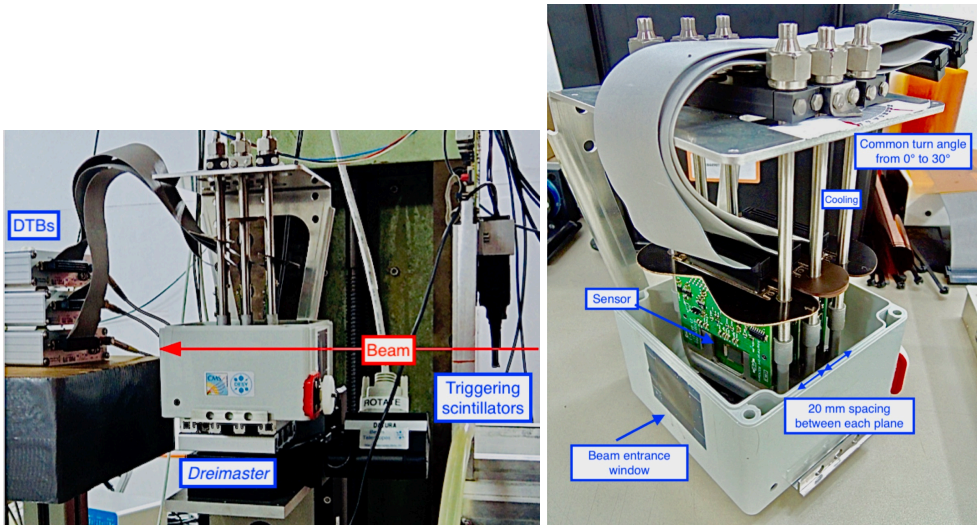
The measurements have been performed in the DESY II test beam facility [282] using an electron beam with momenta between 1 and 6 GeV, a constant absolute momentum spread σ_p of (158 ± 6) MeV over the full momentum range and divergence $d_\alpha = 1$ mrad. The EUDET-type beam telescope [257] available at the facility has a remarkable track resolution that allows measurements of the spatial resolution along the 25 µm direction of non-irradiated sensors [243, 272]. When measuring the resolution of irradiated sensors, the Detector Under Test (DUT) material increases due to the additional cooling equipment. As a consequence, the effective track resolution is degraded, especially when rotating the DUT to perform measurements at different beam incidence angles. In the rotated DUT configuration the material increases and a wider separation of the upstream and downstream telescope stations is required. The approach presented in this thesis does not rely on an external reference tracking detector and allows for the measurement of the spatial resolution along the 25 µm pitch of non-irradiated and irradiated sensors. The setup can be rotated simultaneously with respect to the beam axis to perform measurements at different beam incidence angles.

13.2.2 Experimental setup

Three parallel equidistant planes of sensors (referred to as *Dreimaster*[1]) have been mounted along the beam axis, as shown in Fig. 13.2a. Figure 13.2b is a close up on the *Dreimaster*, where the 20 mm spacing between planes, the adjustable common turn angle and the 25 µm thick Kapton entrance window are visible. The metal tubes allow the circulation of a coolant liquid from an ethanol-based chiller to control the temperature of the irradiated DUTs. For thermal isolation and to prevent condensation, the plastic box can be closed, wrapped with ArmaFlex insulation and flushed with dry air. The cooling and insulation elements are shown in Fig. 13.3.

The material in front of each sensor is minimized to reduce the effect of multiple Coulomb scattering (MS) of the low energy electron beam as depicted in Fig. 13.4, showing a sketch of the setup with the naming convention for the three planes. Thus,

[1] *Dreimaster* is the German word for three-master, a ship with three masts.



(A) (B)

FIGURE 13.2: Left: Image of the *Dreimaster* in the DESY test beam area. The incoming beam is indicated by the red arrow. Elements of the data taking setup described in Sec. 13.2.3 are also indicated. Right: Close up on the *Dreimaster*. The three parallel equidistant planes of sensors are visible. The beam entrance window and the adjustable common turn angle are also indicated. Published in Ref. [249].

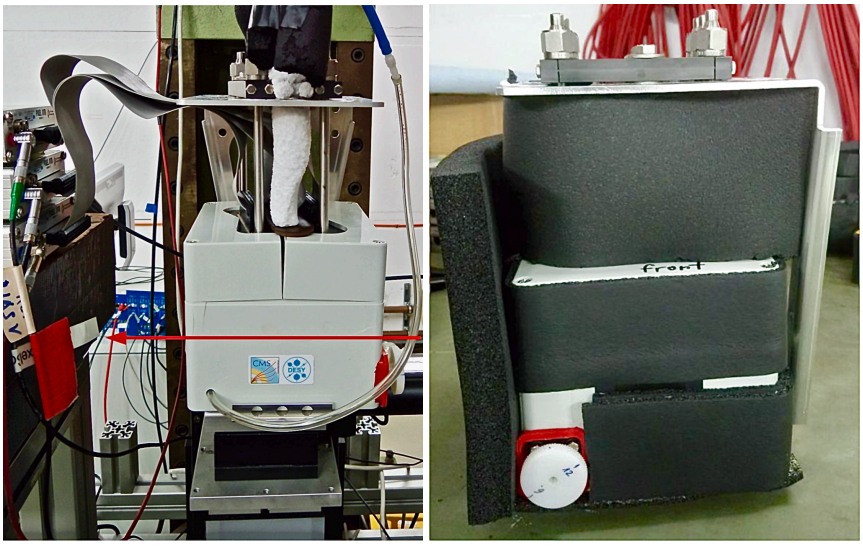
this method allows performing spatial resolution measurements with a pixel pitch of 25 µm at different beam incidence angles and while cooling irradiated devices.

The two external planes serve as reference to reconstruct the tracks of the incoming particles, while the central plane is the DUT.

In the following, the method adopted to perform the resolution measurements is addressed. First, the data taking procedure is explained (Sec. 13.2.3), then the event reconstruction and the alignment process (Sec. 13.2.4) are described. In Sec. 13.2.5 the spatial resolution is defined, and finally in Sec. 13.2.6 the offline analysis is presented. Definitions and applied selections are summarized in Sec. 13.2.7.

13.2.3 Data taking

The PSI ROC4Sens readout chip, to which the sensors are bump-bonded, allows for data taking without zero suppression, making it a powerful tool for sensor studies. After receiving an external trigger, the analog signal of each pixel cell can be stored. An event rate around 150 Hz is reached, limited by the USB2.0 connection between the Digital Test Board (DTB) and the data acquisition computer. The DTB is a dedicated FPGA board and features a digitizer with 12 bit resolution. To save disk space, the digitized signals are stored for a 7×7 pixel cell region-of-interest, centred around a pixel above threshold. Such seed threshold is defined for each pixel as 4 times the noise, given by the Root Mean Square (RMS) of the pixel response in the absence of particles. This procedure is described in detail in Ref. [242]. The signal is corrected for baseline oscillations (pedestal) by subtracting the average pulse-height of the first and last pixel in the corresponding column of the region-of-interest. The external trigger is provided for all the three sensors by scintillators upstream of the *Dreimaster*. Each of the chips to which the sensors are bump-bonded is read-out through a DTB that sends out a BUSY signal while the sensors' analog signals are read-out. To ensure data integrity, a new acquisition can start only when none of



(A) (B)

FIGURE 13.3: Left: Image of the *Dreimaster* in the DESY test beam area. The incoming beam is indicated by the red arrow. The middle plane is cooled through the circulation of liquid from an ethanol-based chiller, as visible from the condensed air around the cooling tube. The dry air tube is inserted in the dedicated entrance at the bottom of the device. Right: The *Dreimaster* is fully closed and wrapped with ArmaFlex insulation.

the three DTBs is in BUSY state. For each configuration, 90k events were collected. Higher statistics was achieved in specific cases.

13.2.4 Reconstruction and Alignment

The track reconstruction is performed in several steps. First, contiguous pixels with a signal above the offline threshold (see Sec. 13.2.6) are clustered together. The cluster charge is the sum of the charge collected by each pixel in the cluster. The cluster position is determined with the Center-of-Gravity (CoG) method from the row and column coordinates of the pixels in the cluster. The η-algorithm [283, 284] would yield better performance especially for small incidence angles, while the head-tail algorithm [285] improves the resolution at large incidence angles. For this work the CoG method was chosen and used consistently for all the analyses. The absolute results may therefore still be improved with optimized algorithms. Here emphasis is given to the relative variation of the resolution after irradiation and as a function of the pixel threshold around the optimal angle.

In Fig. 13.4, a sketch of the setup shows the naming convention for the three planes that will be used in the following.

For the reconstruction of the tracks of the incoming particles, in each plane the cluster local (row, column) coordinates are translated into the global (x, y) coordinate system. The origin of the global coordinate system coincides with the center of sensor B, with the x-axis defined along the 25 µm pitch:

$$x = \text{row} * \text{pitch}_x - l_x/2$$
$$y = \text{column} * \text{pitch}_y - l_y/2,$$

with $\text{pitch}_x = 25\,\mu\text{m}$, $\text{pitch}_y = 100\,\mu\text{m}$, and $l_x = 8\,\text{mm}$ and $l_y = 7.8\,\text{mm}$ being the sensor's dimensions.

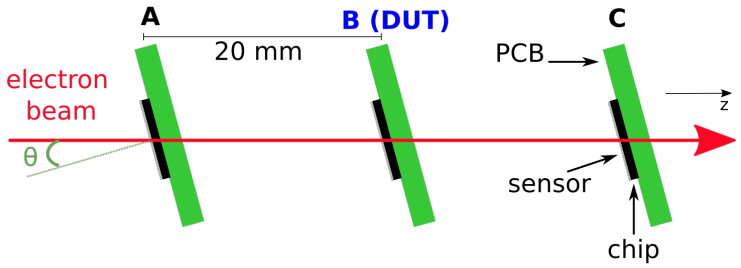


FIGURE 13.4: Sketch of the setup for the resolution measurements (not to scale). Three parallel equidistant planes of sensors (*Dreimaster*) are mounted along the beam axis with a 20 mm spacing between planes and an adjustable common turn angle. The two external planes (A and C) are used as reference to reconstruct the tracks of the incoming particles, while the central plane (B) is the DUT. The thin gray area represents the sensor, first element to be encountered by the beam (red arrow) in each module, followed by the read-out chip (black area) and the PCB support (green area). Published in Ref. [249].

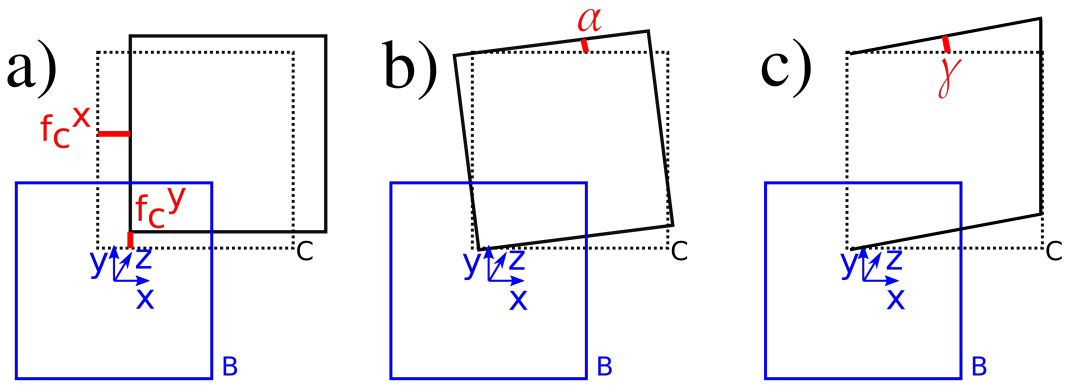


FIGURE 13.5: Scheme of the alignment parameters (in red) between planes B (in blue) and C (in black). The solid black line show the current position of C while the dashed line show the ideal position of C once aligned with B. The parameters with respect to B and A are similarly defined. In a), the alignment parameters f_C^x and f_C^y taking into account the relative position in x and y of B and C are shown. In b), α_C, parametrizing rotations around the z axis, is indicated. In c), the correction γ accounting for rotation around the y-axis is shown. Published in Ref. [249].

The global coordinate system takes into account also the differences between the assumed and the actual location of the detectors, corrected through an iterative alignment procedure. In the alignment procedure the three *Dreimaster* planes are assumed to be parallel, the position of the central plane B is fixed and the two external planes A and C are aligned with respect to B. The alignment parameters described in the following are drawn in red in Fig. 13.5, where for clarity only planes B and C are shown.

Four parameters (f_i^x, f_i^y with i=A, C) take into account the relative position in x and y of the central and external planes (Fig. 13.5, a)). The f_i^x (f_i^y) parameters are obtained by measuring the residual difference between the x (y) coordinate in the i and B plane. For the x coordinate, f_i^x corresponds to the mean of the residual distribution extracted through a Gaussian fit while for the y coordinate the histogram mean is used. A different method is used due to the different pitch size in the two directions. In Fig. 13.6 the residual distributions from which the f_C^x and f_C^y parameters are extracted are shown for a non-irradiated sensor at the optimal angle.

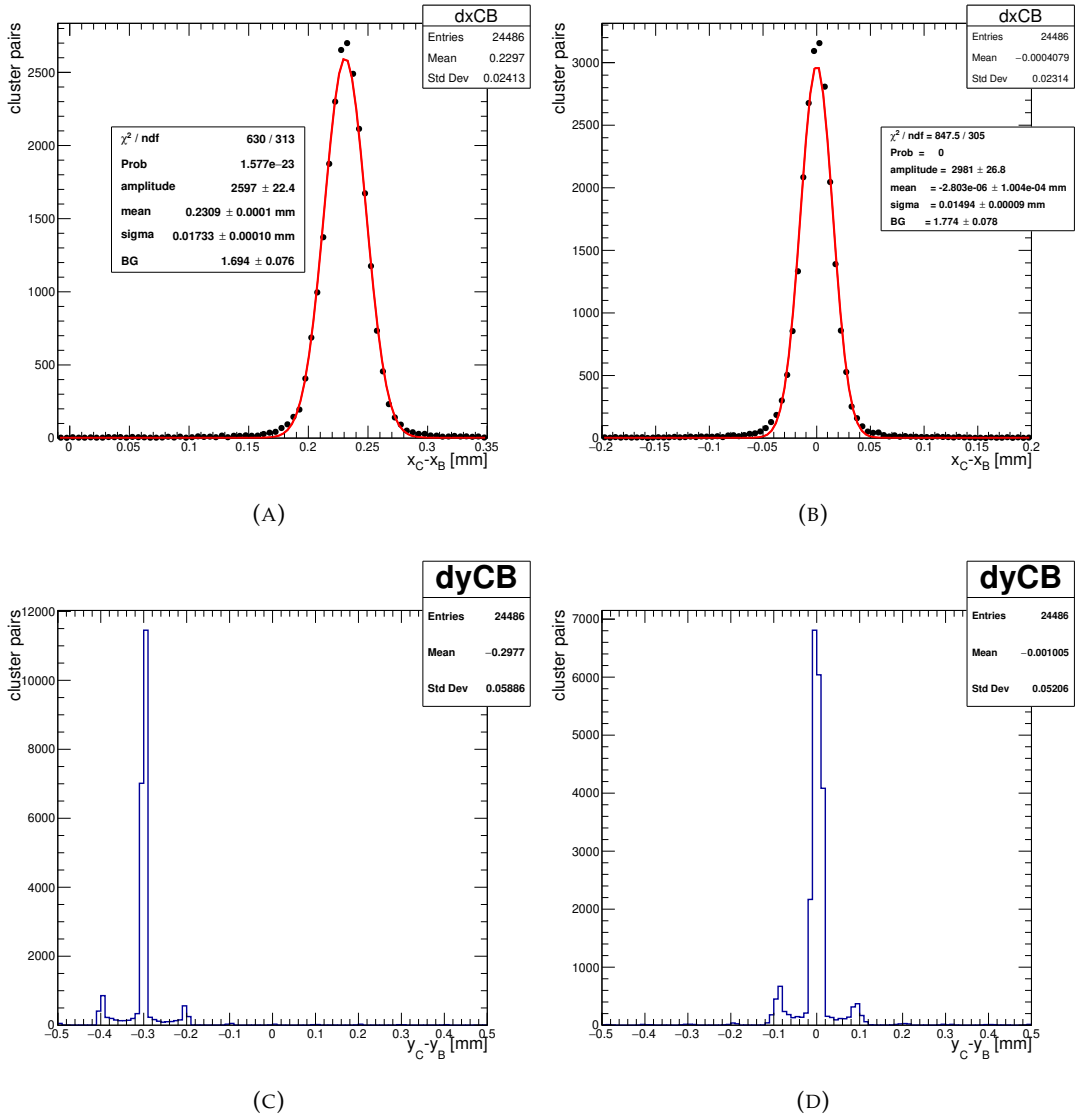


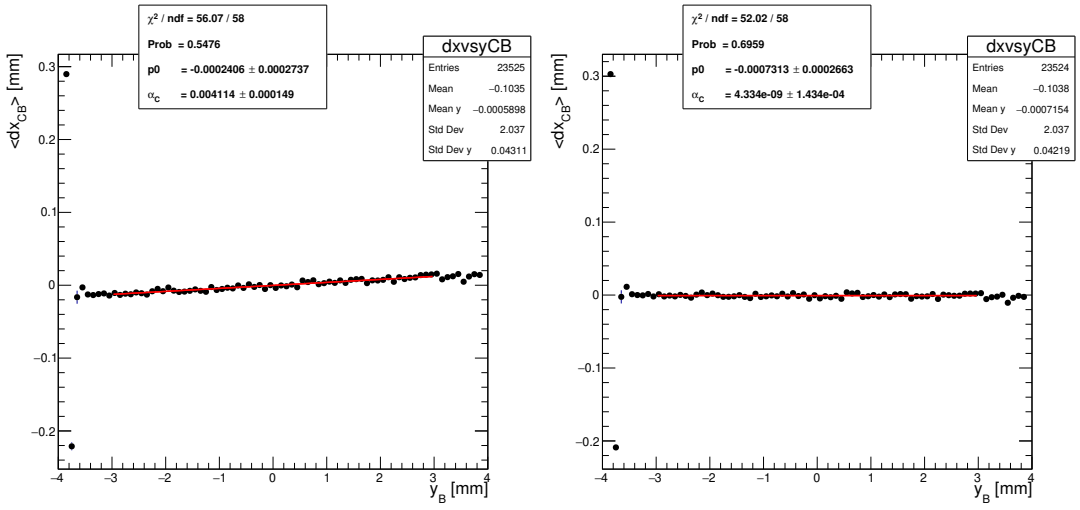
FIGURE 13.6: Residual distributions from which the parameters f_C^x (top) and f_C^y (bottom) are extracted. The distributions on the left (right) are the residuals at the beginning (end) of the alignment procedure. The Gaussian fit on the x residuals is also shown.

The coordinates are updated as:

$$x_i \to x_i - f_i^x$$
$$y_i \to y_i - f_i^y.$$

Two parameters (α_A and α_C) allow for rotations around the beam axis z (Fig. 13.5, b)). The α_i parameters correspond to the slope of the linear function fitted to the distribution of the residual difference of the x coordinate on the i and B planes (with i=A, C) as a function of the y_B coordinate. The fitting procedure is shown in Fig. 13.7. The angular correction is applied through a rotation matrix as:

$$x_i \to x_i \cos(\alpha_i) - y_i \sin(\alpha_i)$$
$$y_i \to y_i \sin(\alpha_i) - y_i \cos(\alpha_i).$$



(A) (B)

FIGURE 13.7: Distribution of the residual difference of the x coordinate on the C and B planes as a function of the y_B coordinate. The red curve is the linear fit from which the parameter α_C is extracted. The distribution on the left (right) is at the beginning (end) of the alignment procedure.

One additional parameter γ (Fig. 13.5, c)) is introduced once a preliminary alignment is in place to correct the x coordinate for rotations around the y-axis, initially excluded by the assumption of parallel planes: $x_i \to (1+\gamma)x_i$, with i=A, C. Also the γ parameter corresponds to the slope of a linear fit (Fig. 13.8). The fit is performed on the distribution of $dx_3 = x_B - (x_A + x_C)/2$ as a function of the x_B coordinate.

In Fig. 13.9 the values assumed by a selection of the alignment parameters at each iteration are shown. The initial value of all parameters is zero. At the first step only f_i^x and f_i^y are corrected (Fig. 13.9a). Subsequently, α_i is taken into account (Fig. 13.9c). Last, the γ correction is estimated and applied (Fig. 13.9d). Fig. 13.9b shows the small adjustments of f_i^x due to the corrections introduced by the angular parameters.

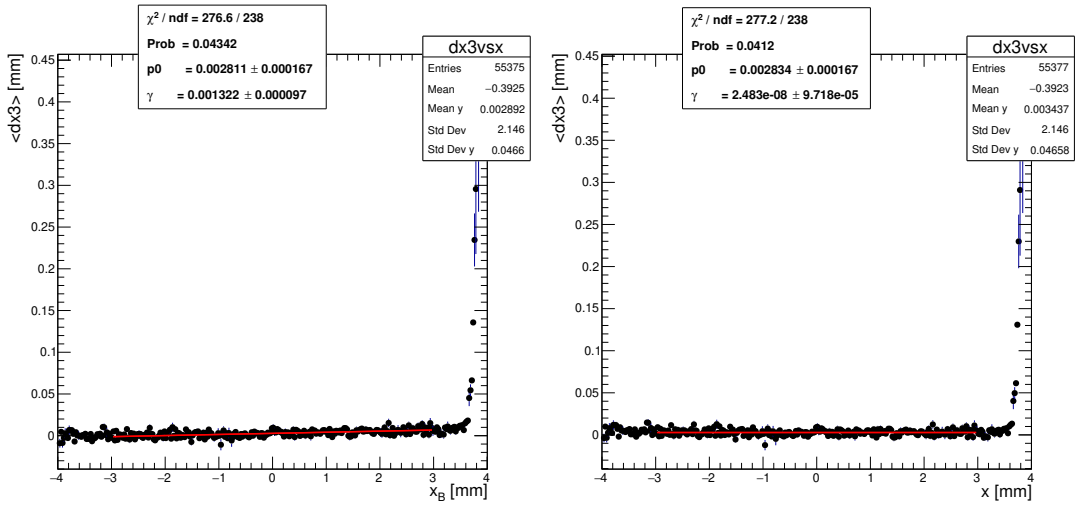
Once all transformations and corrections are applied to the coordinates, the resulting coordinate system is skewed and non-Cartesian.

The alignment procedure is repeated when sensors in the *Dreimaster* are exchanged or when the beam incidence angle θ is varied, as this is achieved through mechanical movements of the setup. Each time the procedure is iterated until the variation of the alignment parameters with respect to the preceding iteration is stable and consistent with zero, with sub-micron and sub-mrad precision for f_i^x and f_i^y and the rotation parameters, respectively. In Sec. 13.2.6.1 the effect of different thresholds on the alignment parameters is reported.

A problem related to the use of the CoG algorithm with threshold cuts can be observed in Figs. 13.6d and 13.10. Since the resolution is measured along the x direction, the explanation focuses on Fig. 13.10 and the cluster size. The cluster size, here and in the following, is defined as the number of rows in a cluster. The residual

$$\Delta x = x_B - \frac{x_A + x_C}{2} \tag{13.1}$$

distribution shown in Fig. 13.10 has a 2-peak structure for events with cluster size 1 in the central plane. The peaks are separated by $\text{pitch}_x/2$ = 12.5 µm. They are due



(A) (B)

FIGURE 13.8: Distribution of the residual difference $dx_3 = x_B - (x_A + x_C)/2$ as a function of the x_B coordinate. The red curve is the linear fit from which the parameter γ is extracted. The distribution on the left (right) is at the beginning (end) of the alignment procedure. At $x_B \approx 4$ the edge of the sensor is reached and larger values of $\langle dx3 \rangle$ are observed. Such events do not affect the alignment procedure.

to events with cluster size 1 in all three sensor planes which are reconstructed at the centres of the pixels independent of the particle position. If the angular divergence of the beam and the distance between the sensors are small, a fraction of events will have the same reconstructed position in all three sensors and peaks appear in the Δx distribution. If the reconstructed position in one of the outer planes differs by pitch_x and the position in the other outer plane remains the same, the prediction in the central plane changes by $\text{pitch}_x/2$, which explains the observed distance between the peaks. The actual position of the two peaks in the Δx distribution and their frequency depends on the relative alignment of the pixels in the three sensor planes on the micrometer scale. Therefore, the determination of the resolution has a significant systematic uncertainty if the fraction of cluster size 1 events is large, which is the case for small angles of the particles to the normal of the sensors. It is assumed that for larger angles, in particular for the optimal angle, the effect of this correlation on the reconstructed positions is negligible, as the number of 1 pixel clusters is reduced. In Fig. 13.11, residual distributions at larger angles are presented, showing a single peak structure.

13.2.5 Spatial resolution definition

Clusters fulfilling the quality requirements described in Sec. 13.2.6 are used to reconstruct the tracks, defined as all possible combinations of straight lines linking a cluster in the A plane with a cluster in the C plane.

Each track is then interpolated to the central plane, in order to measure the residual difference Δx between the measured impact point on the DUT sensor and the predicted impact point from the track reconstruction, defined in Eq. 13.1.

Fig. 13.12a shows example residual distributions after the requirements in Sec. 13.2.6 are applied.

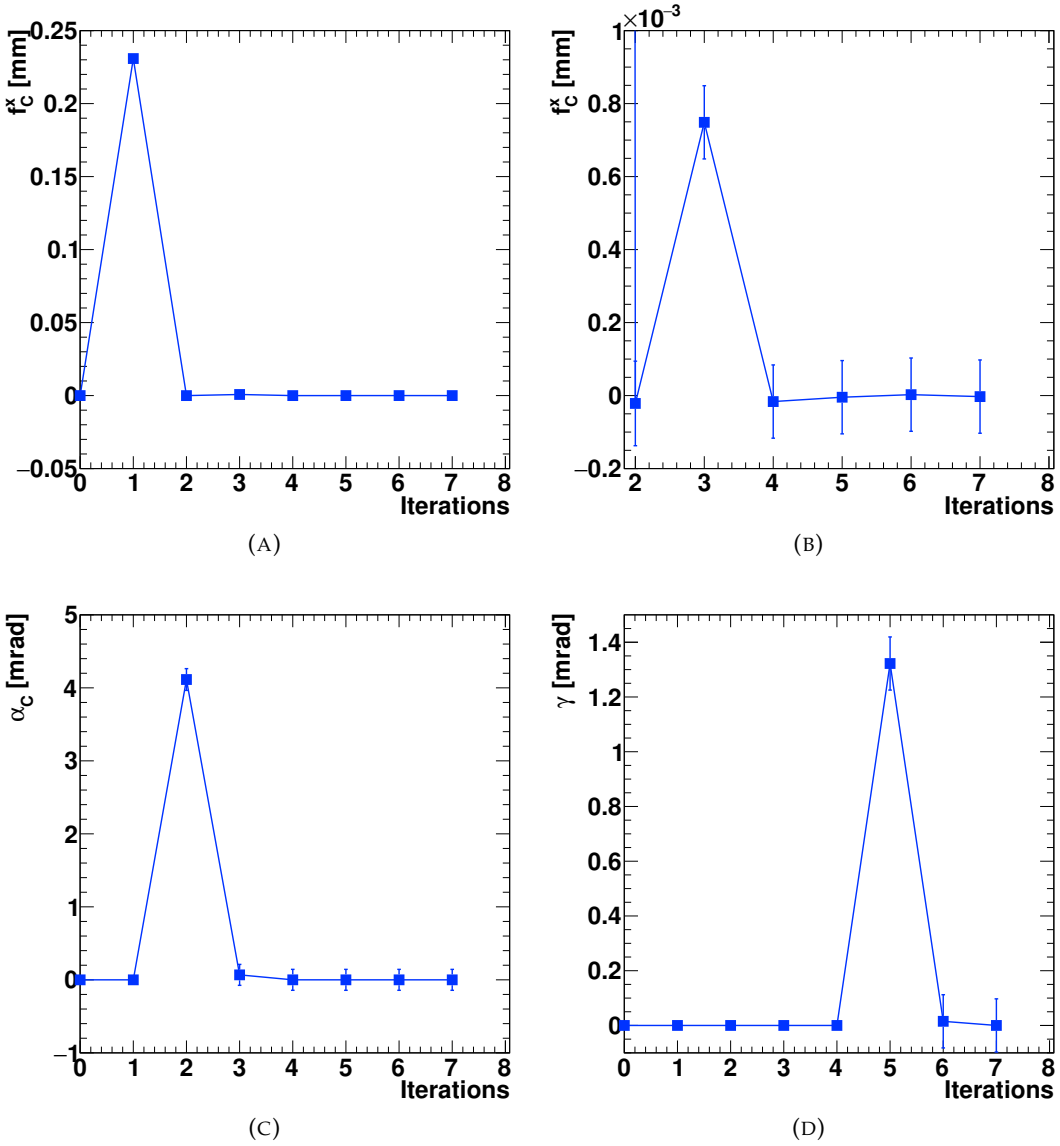


FIGURE 13.9: Values assumed by a selection of the alignment parameters at each step of the alignment procedure. Top: f_C^x, on the left for all iterations, on the right the zoom shows the small adjustments of f_i^x due to the corrections introduced by the angular parameters. Bottom left: α_C. Bottom right: γ.

Different methods have been used in literature (see Table 13.1) to estimate the width of the residual distribution. The approaches can be divided in two main subcategories:

- The width is the σ of a (generalized) Gaussian function
- The width is the RMS (in a certain range) of the residual distribution.

In this thesis, a *reduced RMS* as described below is used as an estimator of the spatial resolution due to non-Gaussian tails, for example because of delta electrons or the clustering algorithm. The usual RMS calculation can result in overestimating the width of the distribution due to the impact of the tails. Thus, the reduced RMS $\delta_{\Delta x}$ has been introduced to reduce the influence of the tails in the evaluation of the distribution width. The reduced RMS is obtained recursively excluding the tracks with

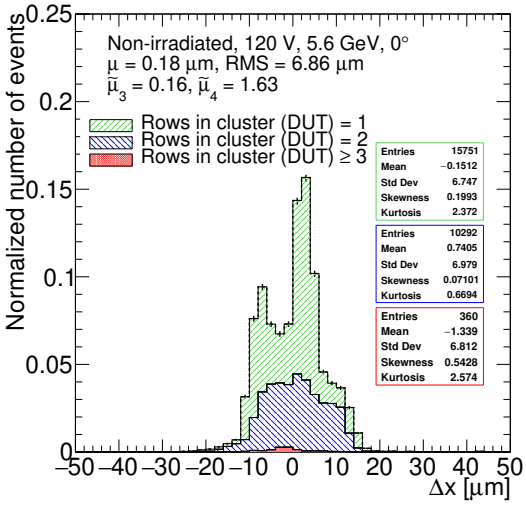
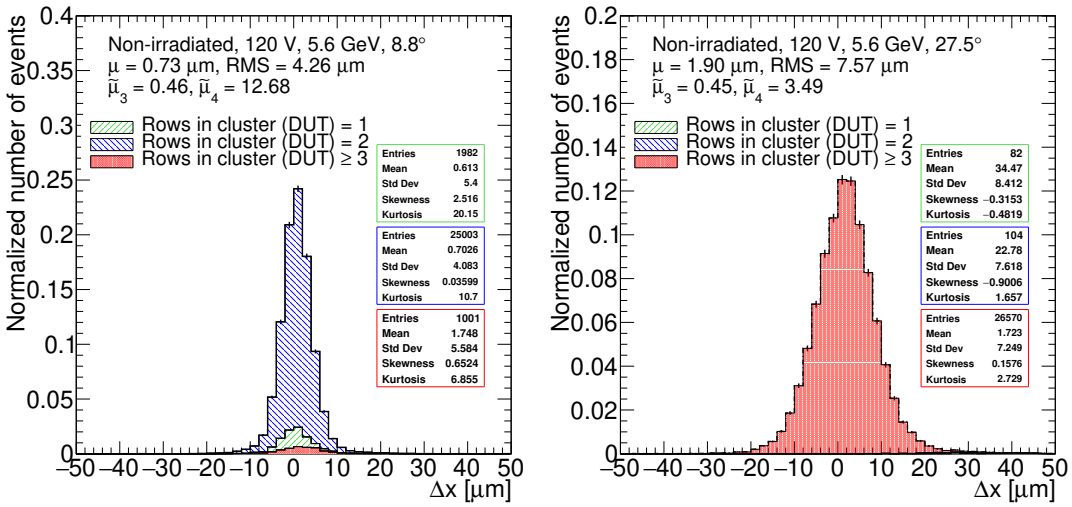


FIGURE 13.10: Residuals, Δx, for a non-irradiated sensor at vertical beam incidence angle. Filled areas of different colors represent the stacked contribution of different size clusters. Mean (μ), RMS, skewness ($\tilde{\mu}_3$) and kurtosis ($\tilde{\mu}_4$) combining all cluster sizes (legend) and for the different cluster sizes (boxes) are also reported. Published in Ref. [249].



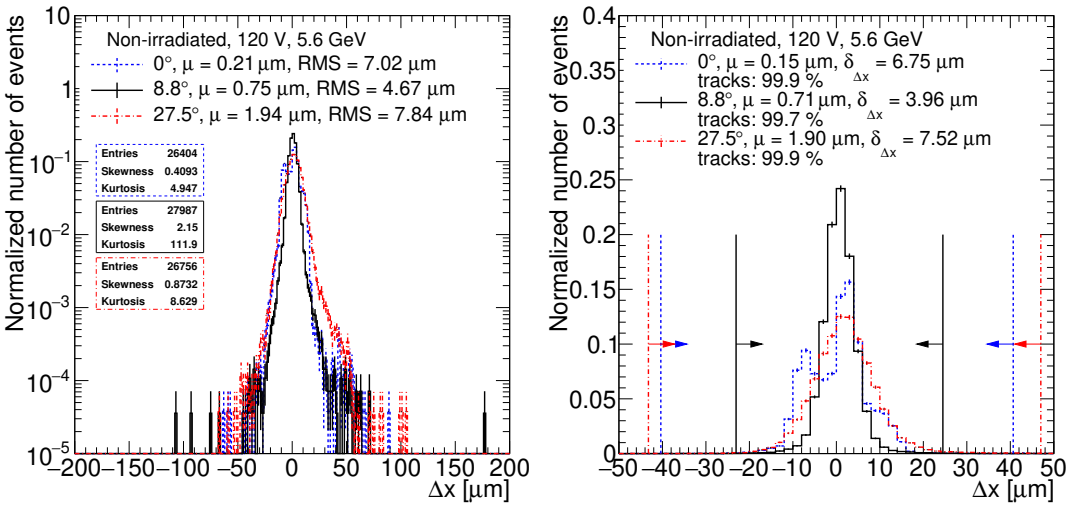
(A) (B)

FIGURE 13.11: Residuals, Δx, for a non-irradiated sensor for the optimal (left) and a shallow beam incidence angle (right). Filled areas of different colors represent the stacked contribution of different size clusters. Mean (μ), RMS, skewness ($\tilde{\mu}_3$) and kurtosis ($\tilde{\mu}_4$) combining all cluster sizes (legend) and for the different cluster sizes (boxes) are also reported. Published in Ref. [249].

$|\Delta x| > N \cdot \text{RMS}$. The procedure is repeated until the RMS converges to a stable value $\delta_{\Delta x}$. For the measurements described in this thesis the range was set to $\pm N \cdot \text{RMS}$ with $N = 6$ and $\approx 99\%$ of the tracks satisfying the requirements in Sec. 13.2.6 have been used in the evaluation of the spatial resolution.

Figure 13.12b shows again the residual distribution, this time with vertical lines indicating the range where the RMS converges to the stable values reported in the legend. The double peak structure at $\theta = 0^\circ$ is due to the impact of single pixel clusters, as shown in Fig. 13.10.

In Table 13.2 different methods of estimating the residuals distribution width are



(A) (B)

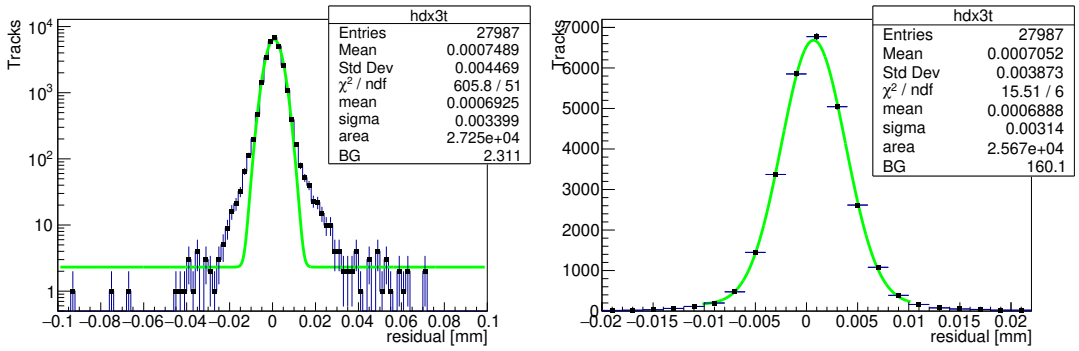
FIGURE 13.12: Residual Δx distribution. Different lines refer to different incidence angles θ. Left: A logarithmic scale is used to show the distribution tails. Right: Vertical lines indicate the range in which the RMS converges to the stable values reported in the legend and used to estimate the *Dreimaster* resolution. The legend also contains the mean μ of the residual distribution and the fraction of tracks in the range indicated by the vertical lines. Published in Ref. [249].

compared for a non-irradiated sample at the optimal angle. The same cluster quality requirements described in Sec. 13.2.6 are applied. For the fit functions, the same definitions as in Ref. [256] are employed. Two cases have been considered. In the first one, the fit has been performed over the full residuals distribution (Fig. 13.13a). Since the tails cannot be described by the fits, in the second case the fit has been restricted to $\pm\, 3\sigma$ of the width obtained in the first case (Fig. 13.13b).

Depending on the method, the width varies by up to 25% (more than 1 μm) with respect to the width definition adopted in this study, the reduced RMS $\delta_{\Delta x}$ ($\pm$ 6 RMS).

Measurement method	Width [μm]
Gaussian	3.40 ± 0.02
Gaussian ($\pm\, 3\sigma$)	3.14 ± 0.02
Generalized Error Function	2.98 ± 0.03
Generalized Error Function ($\pm\, 3\sigma$)	3.13 ± 0.05
Student's t	3.02 ± 0.02
Student's t ($\pm\, 3\sigma$)	3.12 ± 0.03
RMS	4.86 ± 0.02
RMS ($\pm$ pitch)	3.97 ± 0.02
RMS ($\pm$ pitch/2)	3.60 ± 0.02
$\delta_{\Delta x}$ ($\pm$ 3 RMS)	3.41 ± 0.02
$\delta_{\Delta x}$ ($\pm$ 6 RMS)	3.96 ± 0.02

TABLE 13.2: Spatial resolution for a non-irradiated sensor for a beam incidence angle of 8.8°. Different residual width definitions are compared. The range in which the width has been estimated is specified in brackets. If nothing is specified the full range has been used. Published in Ref. [249].



(A) (B)

FIGURE 13.13: Residual Δx distribution fitted with a Gaussian function. Left: A logarithmic scale is used to show the distribution tails. Right: The fit has been restricted to $\pm 3\sigma$ of the width obtained in the left case.

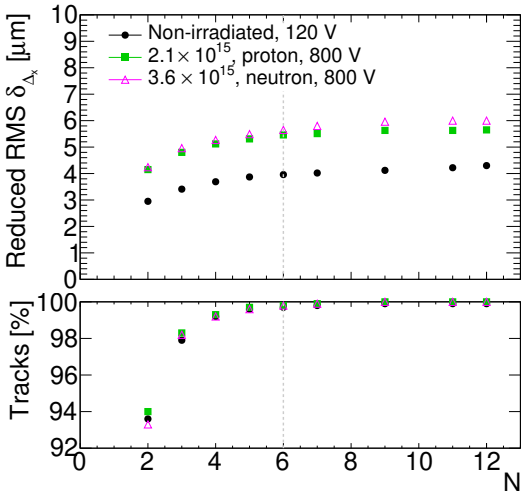


FIGURE 13.14: RMS and number of tracks entering its calculation in the reduced RMS method as a function of N. Different marker colors and styles refer to different fluences. The vertical dashed line marks the value used throughout this work. Published in Ref. [249].

In the reduced RMS method described in Sec 13.2.5, N=6 is chosen to maintain a stable fraction of tracks used in the resolution estimation for different fluences and beam incidence angles. Figure 13.14 shows the RMS and the number of tracks entering its calculation as a function of N. For small N values, a smaller range is considered, leading to better resolution but a lower number of tracks used. Increasing N both the estimated residual width and the number of tracks entering the estimation converge to stable values.

The reduced RMS $\delta_{\Delta x}$ of the residual distribution from measurements taken under the same conditions yields the *Dreimaster* resolution, with the contribution of all the three planes.

The *Dreimaster* method enables the measurement of the spatial resolution σ_x in the plane of the central inclined sensor.

Under the assumptions that the positions determined in the three planes of the

Dreimaster have the same resolution σ_x^0 and that they are uncorrelated, which however is not the case for small angles, σ_x^0 can be obtained from $\delta_{\Delta x}$:

$$\sigma_x^0 = \sqrt{\frac{2}{3}}\delta_{\Delta x}. \tag{13.2}$$

The assumption of same resolution in each plane is valid for three non-irradiated sensors. The assumption of uncorrelated resolutions is valid for tracks formed by hits reconstructed as center-of-gravity of a cluster having a size ≥ 2. This is, for small angles, not the majority of the cases, as shown in Fig. 13.10. For a proper treatment of clusters with size 1 a correction of the introduced bias should be applied first, as presented for instance in Ref. [286]. In addition, only Gauss functions have the property that the convolution of two Gauss functions is again a Gauss function. However, the resolution function of a single sensor can be very different from a Gauss function, as discussed above, and multiple scattering causes significant non-Gaussian tails. Equation 13.2 is therefore not always correct and its applicability should be carefully considered and verified against simulation.

When the central plane is irradiated a similar calculation can be used, considering the single hit uncertainty of the central plane as different from the one of the two reference planes. The spatial resolution σ_x^ϕ is obtained from the following formula, using for each setup configuration the resolution of the non-irradiated planes from Eq. 13.2,

$$\sigma_x^\phi = \sqrt{\delta_{\Delta x}^2 - \frac{1}{2}\sigma_x^{0^2}}. \tag{13.3}$$

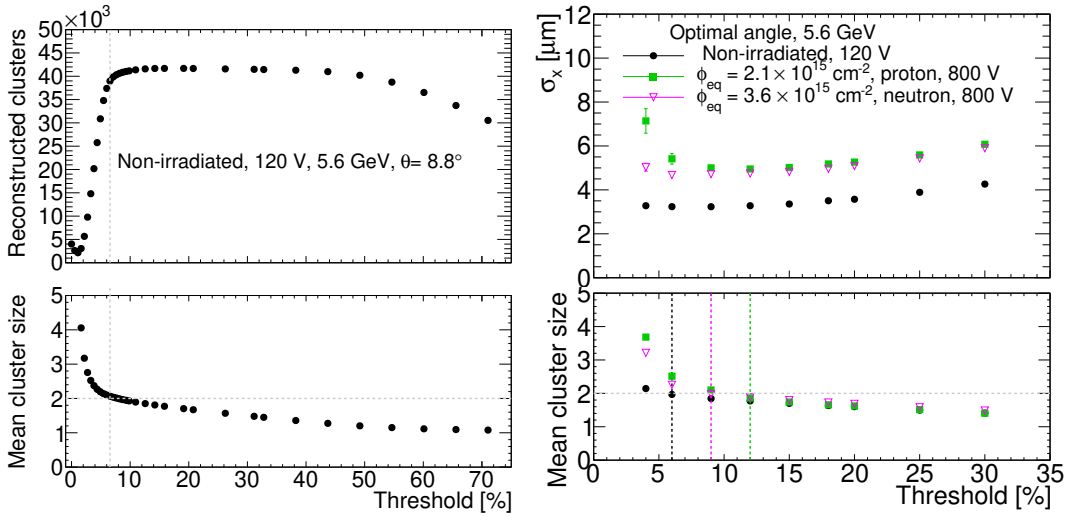
It has to be noted that Equations 13.2–13.3 do not take into account the significant multiple scattering contribution, present at the beam momenta at which the measurements are performed. MS contributions are addressed in the next chapter (e.g. Eq. 14.2).

13.2.6 Cluster quality requirements

Only clusters fulfilling specific requirements are used in the track reconstruction.

13.2.6.1 Offline threshold

First of all, to reduce spurious hits from noise, an offline threshold is applied to all pixels. A pixel is considered in the subsequent clustering step if its pulse-height after pedestal subtraction is higher than the offline threshold. Such threshold has been selected as the one providing the best performance in terms of resolution at the optimal angle for charge sharing. Figure 13.15a shows the number of reconstructed clusters and the average cluster size as a function of the offline threshold. A too low offline threshold results in including noisy pixels in the clusters. In the extreme case of no offline threshold applied, all pixels in the region-of-interest would form a single large cluster, potentially connecting more "true" clusters originated from different hits. Thus, at low thresholds very large clusters are observed and the total amount of clusters is lower than at higher thresholds. On the contrary, if the offline threshold is too high, the detection efficiency will be degraded. The thresholds are expressed as a percentage of the MPV of the cluster charge distribution. The MPV is obtained from the fit of a Landau function convolved with a Gaussian function to the cluster charge distribution at the optimal angle for charge sharing.



(A) (B)

FIGURE 13.15: Left: Number of reconstructed clusters and average cluster size as a function of the offline threshold. The horizontal dashed line marks the expected cluster size from charge sharing at the beam incidence angle at which the measurement was performed. Right: Resolution and cluster size as a function of the offline threshold. Different markers refer to different irradiations. For an easier comparison of the performance of sensors with different irradiations, threshold percentages have been rounded. In both figures, the vertical dashed lines show the offline thresholds used throughout this study. Published in Ref. [249].

In the following, the word threshold will be used to refer to the offline threshold for brevity.

To improve the precision of the position measurements, different thresholds were applied depending on the type and level of irradiation to take into account the different amount of noise in the sensors: 6.6% for non-irradiated sensors, 12.7% for proton-irradiated sensors at $\phi_{eq} = 2.1 \times 10^{15}\,cm^{-2}$ and 9.5% for neutron-irradiated sensors at $\phi_{eq} = 3.6 \times 10^{15}\,cm^{-2}$. In Fig. 13.15b the impact of the applied threshold on the resolution and the cluster size on sensors with different irradiations is shown. The non-irradiated sensor module is the least affected by different thresholds, but at higher thresholds a shift of the optimal angle towards higher values has been observed. The proton-irradiated sensor is the one showing a larger degradation of the resolution and increase in cluster size at low thresholds. The effect of applying the same threshold on all sensors despite the different irradiation level has been checked. If a 6.6% threshold, preferred for the non-irradiated sensor, was applied to the irradiated sensors the measurement would be affected by high noise level: at such threshold the average cluster size of the proton-irradiated sensor is 25% higher than expected from charge sharing. Raising the threshold to 12.7%, preferred for the proton-irradiated sensor, while measuring the resolution of the non-irradiated sensor would not affect the result at small beam incidence angles but would deteriorate the resolution at shallow angles by $\approx 5\%$.

The effect on the alignment of applying different thresholds has been checked for the proton-irradiated sensor module, whose resolution is the most affected by the threshold cut, at the optimal angle. Figure 13.16 shows the values assumed by a selection of the alignment parameters at each iteration. While the parameters assume different values during the alignment procedure, depending on the applied

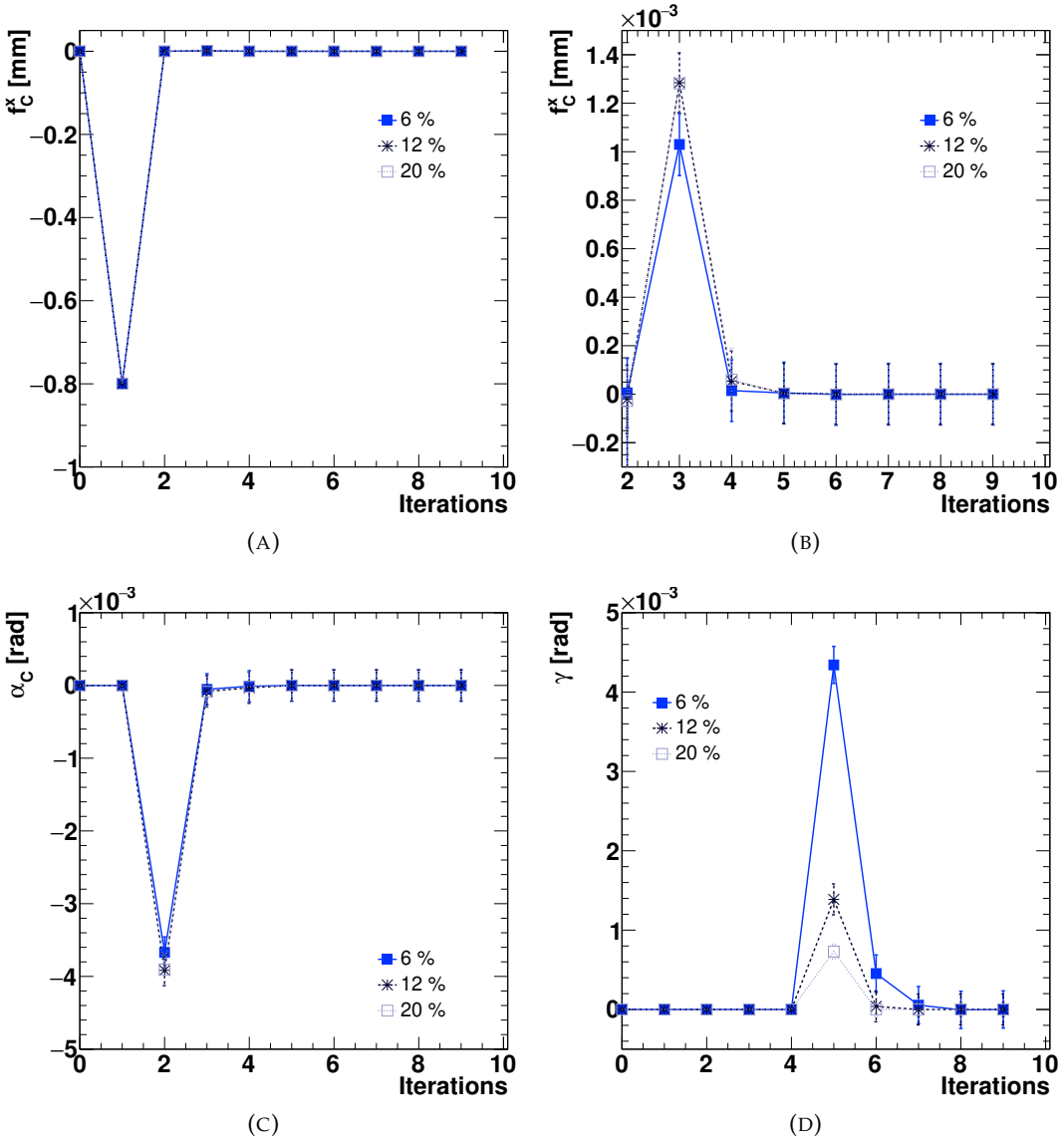


FIGURE 13.16: Values assumed by a selection of the alignment parameters at each step of the alignment procedure, different markers and lines refer to different thresholds. Top: f_C^x, on the left for all iterations, on the right the zoom shows the small adjustments of f_i^x due to the corrections introduced by the angular parameters. Bottom left: α_C. Bottom right: γ.

threshold, they all converge to zero once the alignment is completed. The final values assumed by all the parameters for the three different thresholds are reported in Table 13.3.

Online and offline thresholds are summarized in Table 13.4.

Figure 13.17 shows the cluster size distributions for a non-irradiated and a proton-irradiated sensor at different beam incidence angles θ with the 6.6% and 12.7% threshold applied, respectively. While in both cases the cluster size increases for increasing angles as expected from charge sharing, the proton-irradiated sensor has a lower average cluster size, due to the higher threshold.

Parameter	Offline Threshold [%]		
	6	12	20
f_A^x [µm]	0.005± 0.1	0.006± 0.1	-0.001± 0.1
f_A^y [µm]	7e-05± 0.5	0.0005± 0.5	-0.0003± 0.5
α_A [mrad]	-2e-06± 0.2	4e-06± 0.2	5e-06± 0.2
f_C^x [µm]	-0.0001± 0.1	0.0003± 0.1	0.0003± 0.1
f_C^y [µm]	-0.0004± 0.5	-0.0002± 0.5	0.0002± 0.5
α_C [mrad]	-9e-07± 0.2	2e-06± 0.2	-3e-06± 0.2
γ [mrad]	-2.7e-05± 0.2	-0.00022± 0.2	2e-09± 0.09

TABLE 13.3: The table lists the final values assumed by all the alignment parameters for three different thresholds for the alignment of the proton-irradiated sensor at the optimal angle. The final variation is consistent with zero with submicron precision for all alignment parameters.

Irradiation $[\phi_{eq}/10^{15}\,\mathrm{cm}^{-2}]$	Average seed pixel thr. [%]	offline Thr. [%]	MPV [ADC counts]
0	7.5	6.6%	183
2.1, proton	20.2	12.7%	118
3.6, neutron	11.6	9.5%	159

TABLE 13.4: The table lists the online and the offline thresholds used in this analysis as a percentage of the cluster charge MPV, for the fluences under study. The non-irradiated sensor module was operated at a bias voltage of 120 V and room temperature. The irradiated sensor modules were measured at a bias voltage of 800 V and temperature of $\approx -24\,^{\circ}$C. Published in Ref. [249].

13.2.6.2 Track requirements

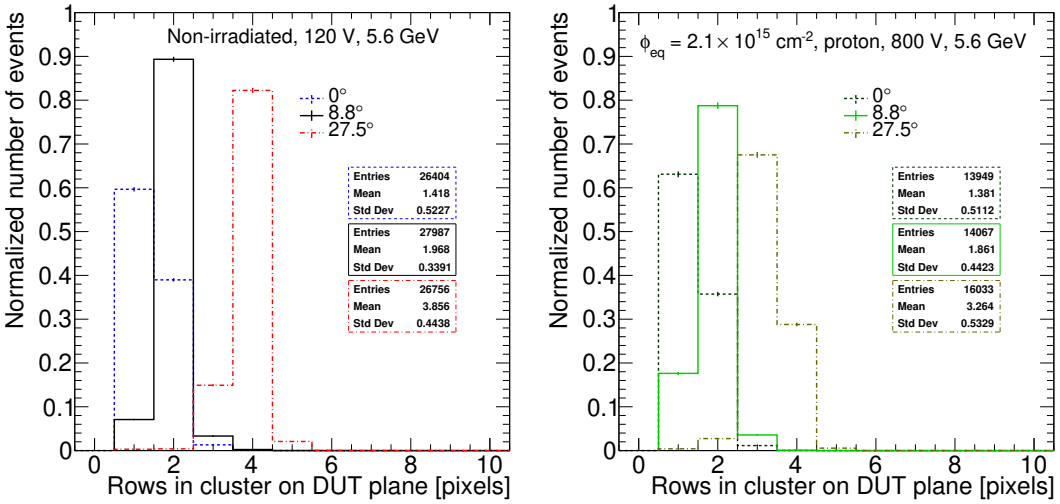
For a non-irradiated sensor, typically one cluster per event is recorded in each plane. In ≈5% of the events with at least one cluster, also a second (noise) hit is recorded. The number of events with more than two hits is $<<$ 5%.

If the distance between any of the pixels in two different clusters is lower than 600 µm, only the cluster with the higher charge will be used in the track reconstruction to eliminate remaining noise contributions. Furthermore, since more clusters can be present outside the 600 µm radius, only the cluster with the smallest $d_{xy} = \sqrt{(x_B - (x_A + x_C)/2)^2 + (y_B - (y_A + y_C)/2)^2}$ is considered, as it is assumed to originate from a real particle track.

Given the assumption in the alignment step of incoming particles parallel to the z-axis, the following requirements are also applied:

$$|x_A - x_C|,\ |y_A - y_C| < d_{AC} \cdot 3\, d_\alpha \cdot\ 5\,\mathrm{GeV}/p_{beam} \tag{13.4}$$

where d_{AC} is the 40 mm distance between the A and C planes, the factor 3 is to allow a 3σ tolerance on the beam divergence d_α introduced above, 5 GeV $/p_{beam}$ takes into account the dependence of the beam divergence on the beam momentum, normalized to 5 GeV. This restricts the analysis to straight tracks and reduces the combinatorics in track seeding.



(A) (B)

FIGURE 13.17: Cluster size. Different line colors and styles refer to different beam incidence angles θ. Left: non-irradiated sensor with a threshold at 6.6% of the cluster charge MPV. Published in Ref. [249]. Right: proton-irradiated to $\phi_{\text{eq}} = 2.1 \times 10^{15}$ cm^{-2} sensor with a 12.7% threshold.

The following additional requirement

$$|y_{\text{B}} - (y_{\text{A}} + y_{\text{C}})/2| < 2 \cdot \text{pitch}_y/\sqrt{12} \tag{13.5}$$

is applied for consistency with [242], which presents resolution studies on 50×50 µm^2 sensors of the same submission. It used to select tracks falling in a fiducial residual region in the direction not considered in the resolution measurement.

13.2.6.3 Cluster charge requirements

To reduce the contribution of delta electrons on the resolution whilst maintaining high statistics, hits above the 90% quantile of the cluster charge distribution are rejected on each of the three *Dreimaster* planes. This is illustrated in Fig. 13.18a where cluster charge distributions for a non-irradiated sensor at different beam incidence angles θ are shown. The vertical lines and arrows indicate the parts of the distributions that are kept. Similarly, Fig. 13.18b presents the cluster charge distributions for sensors with different fluences at the optimal angle for resolution.

In Fig. 13.18c, the solid line shows the pixel charge of all clusters above thresholds on the central *Dreimaster* plane. The dashed line indicates the pixel charge after hits above the 90% quantile of the cluster charge distribution are removed from all three *Dreimaster* planes.

Table 13.5 compares the spatial resolution excluding ("90% cut") and not excluding ("No cut") the clusters above the 90% quantile of the cluster charge distribution on the central plane of the *Dreimaster* in the reduced RMS calculation. When the additional requirement is not applied, the resolution is degraded by 10-20%, especially at shallow incidence angle. The resolution at 0° for the non-irradiated sample, with the "90% cut" applied, is the one quoted in Fig. 13.1.

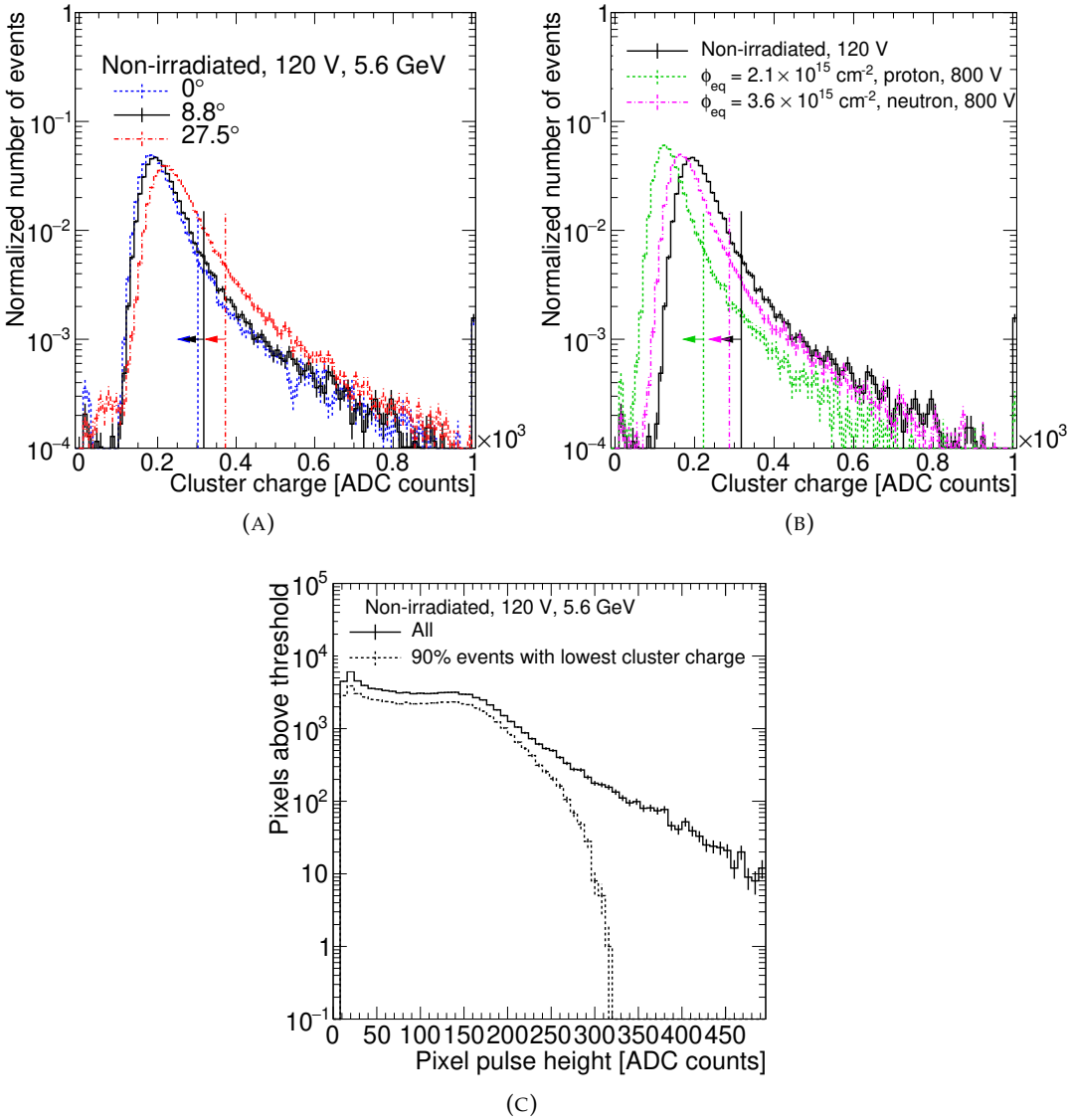


FIGURE 13.18: Top: Cluster charge of the DUT sensor. The vertical lines and arrows indicate the parts of the distributions that are kept. Underflow and overflow bins are also shown. Left: Non-irradiated sensor, different line colors and styles refer to different beam incidence angles θ. Right: Sensors with different fluences, indicated by different line styles and colors. The non-irradiated and proton-irradiated sensor are measured at $\theta = 8.8^\circ$ and the neutron-irradiated one at $\theta = 12^\circ$. Bottom: Pixel charge of all clusters above threshold on the central *Dreimaster* plane. A threshold at 6.6% of the cluster charge MPV is applied. The solid line shows all clusters above threshold. The dashed line indicates the pixel charge after hits above the 90% quantile of the cluster charge distribution are removed from all three planes. Published in Ref. [249].

The spatial resolution has also been measured in different ranges of the central *Dreimaster* plane's cluster charge distribution. The impact on the resolution once again depends on the beam incidence angle. At shallow angles, the resolution is more sensitive to requirements on the cluster charge as the incoming electrons are traversing more material, therefore increasing both the MPV and the width of the cluster charge distribution. The observed shift of the MPV is consistent with the

Irradiation $[\phi_{eq}/10^{15}\,\mathrm{cm}^{-2}]$	Spatial resolution [µm]					
	90% cut			No cut		
	0°	8.8°	27.5°	0°	8.8°	27.5°
0	5.51	3.23	6.14	6.3	4.05	7.6
2.1, proton	6.5	4.95	8.1	7.2	5.76	9.42
	8°	12°	22°	8°	12°	22°
3.6, neutron	5.6	5.0	6.6	6.5	5.9	7.2

TABLE 13.5: patial resolution for a non-irradiated sensor, a proton-irradiated ($\phi_{\mathrm{eq}} = 2.1 \times 10^{15}\,\mathrm{cm}^{-2}$) sensor, and a neutron-irradiated ($\phi_{\mathrm{eq}} = 3.6 \times 10^{15}\,\mathrm{cm}^{-2}$) sensor excluding ("90% cut") and not excluding ("No cut") the clusters above the 90% quantile of the cluster charge distribution on the central plane of the *Dreimaster*. The resolution at 0° for the non-irradiated sample, with the "90% cut" applied, is the one quoted in Fig. 13.1. Published in Ref. [249].

expected energy deposition in the increased material traversed by the incoming particles at shallow incidence angles. In Fig. 13.19a the spatial resolution as a function of each considered charge interval is shown for a shallow beam incidence angle. The cluster charge distribution of the central *Dreimaster* plane is also shown as reference (Fig. 13.19b). For both the non-irradiated and the proton irradiated sensors, the resolution has a minimum in the interval [MPV-σ, MPV+σ], where σ is the width of the cluster charge distribution obtained from the fit of a Landau function convolved with a Gaussian function. The cluster charge distribution for irradiated sensors is narrower than for non-irradiated ones, reducing the effect on the resolution of requiring a specific cluster charge interval.

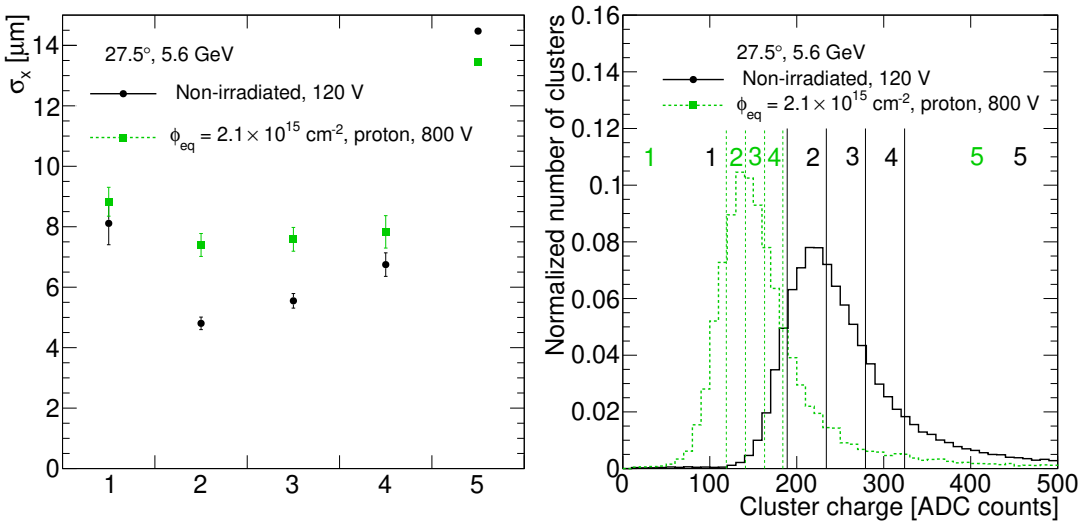
13.2.7 Summary of applied cuts and resolution definition

In this section, the quality requirements as well as the resolution definition discussed along the chapter are summarized. They are applied to the measurements presented in Chapter 14.

- An offline threshold of 6.6%, 12.7% and 9.5% of the cluster charge MPV is applied to pixels of non-irradiated, proton-irradiated to $\phi_{\mathrm{eq}} = 2.1 \times 10^{15}\,\mathrm{cm}^{-2}$, and neutron-irradiated to $\phi_{\mathrm{eq}} = 3.6 \times 10^{15}\,\mathrm{cm}^{-2}$ sensors, respectively
- If the distance among any of the pixels in two different clusters is lower than 600 µm, only the cluster with the higher charge is used in the track reconstruction
- The cluster with the smallest $d_{xy} = \sqrt{(x_{\mathrm{B}} - (x_{\mathrm{A}} + x_{\mathrm{C}})/2)^2 + (y_{\mathrm{B}} - (y_{\mathrm{A}} + y_{\mathrm{C}})/2)^2}$ is used in the track reconstruction
- $|x_{\mathrm{A}} - x_{\mathrm{C}}|,\ |y_{\mathrm{A}} - y_{\mathrm{C}}| < \mathrm{d_{AC}} \cdot 3\,\mathrm{d}_{\alpha} \cdot\ 5\,\mathrm{GeV}/\mathrm{p_{beam}}$
- $|y_{\mathrm{B}} - (y_{\mathrm{A}} + y_{\mathrm{C}})/2| < 2 \cdot \mathrm{pitch}_y/\sqrt{12}$
- Hits above the 90% quantile of the cluster charge distribution are rejected on each of the three *Dreimaster* planes.

The width of the residual distribution

$$\Delta x = x_{\mathrm{B}} - \frac{x_{\mathrm{A}} + x_{\mathrm{C}}}{2} \tag{13.6}$$



(A) (B)

FIGURE 13.19: Left: spatial resolution for the central *Dreimaster* sensor in different cluster charge intervals. Right: cluster charge distribution shown for reference, with vertical lines highlighting the considered charge intervals. Black markers and lines refer to a non-irradiated sensor and green markers and lines to a proton-irradiated sensor at $\phi_{\mathrm{eq}} = 2.1 \times 10^{15}\,\mathrm{cm}^{-2}$. The lower cluster charge threshold is fixed as discussed in Sec. 13.2.6.1. The intervals are defined as: 1 = [0, MPV-σ], 2 = [MPV-σ, MPV+σ], 3 = [MPV+σ, MPV+3σ], 4 = [MPV+3σ, MPV+5σ], 5 = ADC counts >MPV+5σ. Published in Ref. [249].

is estimated with the reduced RMS $\delta_{\Delta x}$, obtained from recursively evaluating the RMS, after setting the range of the residual distribution to a multiple of the RMS evaluated at the preceding step. The procedure is repeated until the RMS converges to a stable value $\delta_{\Delta x}$. For the measurements described in this thesis the range was set to $\pm N \cdot$RMS with $N = 6$.

Figure 13.20 shows the residual distributions for sensors at the three fluences under study after all the quality requirements listed above are applied. Vertical lines indicates the range in which the RMS converges to the stable value $\delta_{\Delta x}$.

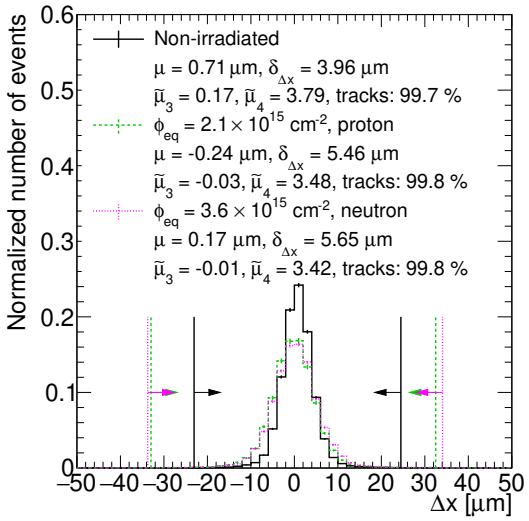


FIGURE 13.20: Residual Δx distribution. Different line colors and styles refer to different fluences. The non-irradiated and proton-irradiated sensors are measured at $\theta = 8.8^{\circ}$ and the neutron-irradiated one at $\theta = 12^{\circ}$. Vertical lines indicates the range where the RMS converges to the stable value reported in the legend and used to estimate the *Dreimaster* resolution. Published in Ref. [249].

14 Results and conclusions

The content of sections 14.1 and 14.2 has been collected and adapted by the author of this thesis to result in the public note in Ref. [249] and has been submitted to the Nuclear Instruments and Methods in Physics Research (Section A).

The author of this work took part in the data taking, performed the presented data analysis and the reported literature review. The work was performed under the supervision of Prof. Dr. Erika Garutti, in collaboration with Dr. Andreas Hinzmann and Dr. Daniel Pitzl. Finn Feindt developed the online acquisition software and some of the data corrections applied in this work. Dr. Jörn Schwandt designed the sensors. Dr. Georg Steinbrück organized the test beam periods. Dr. Aliakbar Ebrahimi, Finn Feindt, Prof. Dr. Erika Garutti, Dr. Paolo Gunnellini, Caroline Niemeyer, Dr. Daniel Pitzl, Dr. Jörn Schwandt, Dr. Georg Steinbrück and the author of this thesis contributed to the data taking. Prof Emer. Dr. Robert Klanner contributed to the discussion about validity and limitations of the employed method.

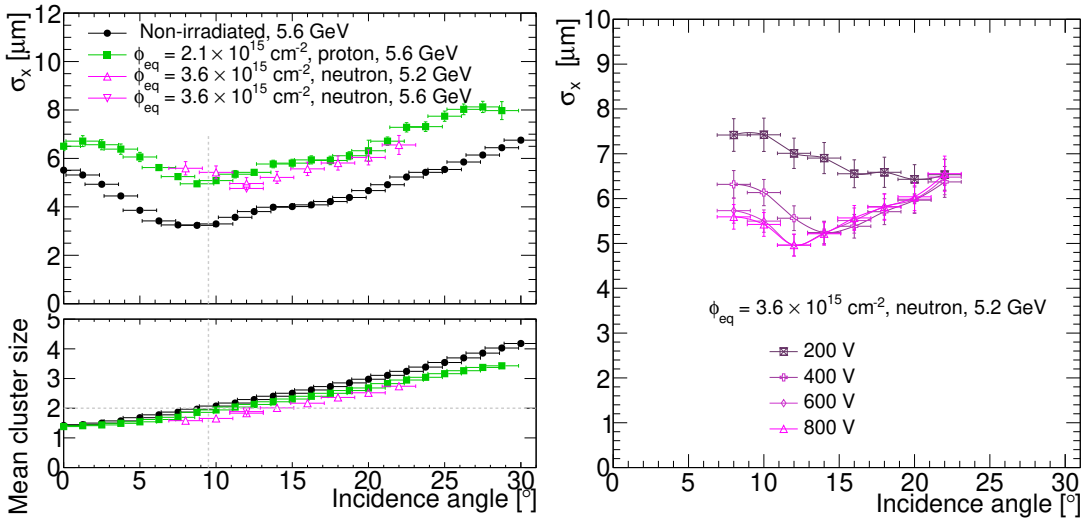
14.1 Results

In the previous chapter, the track quality requirements have been studied, analysing their effect on both the resolution and the analyzed sample size. The quantity used to measure the resolution, the reduced RMS, has been introduced. In the following, the requirements and definitions summarized in Sec. 13.2.7 are applied.

The non-irradiated sensors were operated at a bias voltage of 120 V. Measurements with different designs, listed in Table 12.1, are in agreement within uncertainties and thus only results with the design shown in Fig. 12.11b as DUT are shown in the following. All the irradiated sensors have the design shown in Fig. 12.11a and were operated at a bias voltage of 800 V, if not stated otherwise.

14.1.1 Scan of different beam incidence angles

First, the spatial resolution as a function of the beam incidence angle is presented. In realistic data-taking conditions in the CMS detector, not only particles originating from pp-collisions cross the layers of the detector at different incidence angles but the charge carriers produced in the sensors are subjected to the Lorentz force. This force, generated by the 3.8 T magnet, deflects the charges drifting to the electrodes. When tilting the sensors with respect to the beam axis, the generated charges spread over multiple pixels, mimicking the effect of the Lorentz drift and of the different inclination of the particles from pp-collisions. However, tilting the sensor is not equivalent to the effect of a magnetic field. In the absence of a magnetic field, the signal is independent of the depth where the charges are generated. On the contrary, the Lorentz force caused by a magnetic field deviates the charges from the trajectory they would follow if no magnetic field is present. Furthermore, the displacement is wider for charges created far away from the collecting electrodes.



(A) (B)

FIGURE 14.1: Left: Spatial resolution and average cluster size as a function of the beam incidence angle with respect to the normal to the sensors. Different markers refer to different irradiations. The electron beam momentum at which the measurements were taken is stated in the legend. The vertical (horizontal) dashed line marks the expected optimal angle 9.5° given by arctan (pitch_x/t) (the cluster size of two where the optimal resolution is expected). Right: Spatial resolution as a function of the beam incidence angle with respect to the normal to the sensors for the neutron-irradiated sensor. Different markers refer to different bias voltages applied to the sensor. The curves have the only purpose to guide the eyes. Published in Ref. [249].

In Fig. 14.1a the spatial resolution along the 25 µm pitch direction and the average cluster size are shown as a function of the beam incidence angle with respect to the normal to the sensors (θ in Fig. 13.4) for sensors at the three different fluences under study. The beam momentum was 5.6 GeV for the non-irradiated and the proton-irradiated sensors. For the neutron-irradiated sensor module, measurements for beam momenta of 5.2 GeV and 5.6 GeV are reported. Vertical error bars represent the statistical Poisson uncertainties, horizontal error bars show the uncertainty on the angle determination: $\pm\,1^\circ$ due to the mechanical precision of the rotating device and its alignment in the mounting process, and a statistical uncertainty of $\pm\,0.5^\circ$. Poisson uncertainties have been estimated as in Ref. [256].

The incidence angle at which a minimum in the spatial resolution (*best resolution*) is achieved will be indicated as the optimal angle in the following. At such angle the average cluster size is about 2. Within uncertainty, the optimal angle is consistent with 9.5°, given by arctan (pitch_x/t), for both the non-irradiated and proton-irradiated samples, with t the sensor thickness. The best resolution for the neutron-irradiated sample corresponds to an angle of 12°, with an average cluster size of about 2. Differences between neutron and proton irradiated samples can be due to different shapes of the electric fields in the sensors, providing a less uniform charge collection in neutron-irradiated sensors [287, 288], and more severe ionising energy loss to the readout chip for proton-irradiated modules. Nevertheless, the measured spatial resolution at the respective optimal angle is the same. To further investigate the dependence of the resolution on the beam angle for the neutron-irradiated sensor, the measurement was repeated at lower bias voltages, as presented in Fig. 14.1b. The curves have the only purpose to guide the eyes. The angle for optimal resolution

increases and the spatial resolution degrades with decreasing bias voltage. Above a bias voltage of 600 V both the resolution and the angle at which the minimum is situated converge to stable values.

14.1.2 Scan of different beam momenta

At beam momenta of 5.2 GeV and 5.6 GeV, at which the measurements in Fig. 14.1 were performed, the multiple scattering contribution has a significant impact on the resolution.

Charged particles traversing a medium are deflected by many small angle scatters, mostly due to Coulomb scattering from the material nuclei [15]. The net distribution of the scattering angle Θ_0 and the displacement caused by several small angle scatters are Gaussian owing to the central limit theorem and the following approximation [289] can be used:

$$\Theta_0 = \frac{13.6\text{MeV}}{\beta c p} z \sqrt{x/X_0} \left[1 + 0.088 \log \frac{x z^2}{X_0 \beta^2} \right] \tag{14.1}$$

where p, βc, and z are the momentum (in MeV/c), velocity, and charge number of the incident particle, and x/X_0 is the thickness of the scattering medium in radiation lengths.

As Eq. 14.1 shows, the net scattering will be larger at lower beam momentum, thus worsening the measurement of the spatial resolution.

To extrapolate the spatial resolution to infinite beam momentum, thus when the MS contribution can be neglected, measurements at different beam momenta between 1 GeV and 6 GeV were performed. The spatial resolution squared as a function of the beam momentum squared is shown in Fig. 14.2, for a non-irradiated sample and a proton-irradiated sample at $\phi_{\text{eq}} = 2.1 \times 10^{15}\,\text{cm}^{-2}$, operated at 120 V and 600 V respectively. The uncertainties on p_{beam} represent the absolute momentum spread σ_p introduced in Sec. 13.2.

The function

$$\sigma_x^2(p_{\text{beam}}) = \sigma_{\text{extr}}^2 + (\sigma_{\text{MS}}/p_{\text{beam}})^2 \tag{14.2}$$

has been fitted to the data (solid lines), where σ_{extr} is the extrapolated resolution at infinite beam momentum and σ_{MS} the contribution to the resolution due to multiple scattering. $\sigma_x^2(p_{\text{beam}})$ corresponds to σ_x^0 (σ_x^ϕ) defined in Eq. 13.2 (Eq. 13.3) in the case of a non-irradiated (irradiated) central sensor. Equation 13.3 is valid if σ_x^0 and σ_x^ϕ are obtained under the same conditions, thus to extract $\sigma_x^\phi(p_{\text{beam}})$, σ_x^0 should be measured at the same p_{beam}. Since σ_x^0 has been sampled in larger p_{beam} intervals than σ_x^ϕ, a different procedure has been used to extract σ_x^ϕ: σ_x^0 is not directly obtained from measurements but from the fit function in Eq. 14.2 at a given p_{beam}.

A spatial resolution of 2.4 ± 0.1 µm was found for the non-irradiated sensor at the optimal angle when extrapolating to infinite beam momentum. For the proton-irradiated sample the extrapolated spatial resolution is 4.1 ± 0.2 µm.

14.2 Conclusions

In this study spatial resolution measurements for planar n^+p sensors with an active thickness of 150 µm and a pixel size of 100×25 µm^2 produced by Hamamatsu Photonics were presented for non-irradiated, proton-irradiated and neutron-irradiated

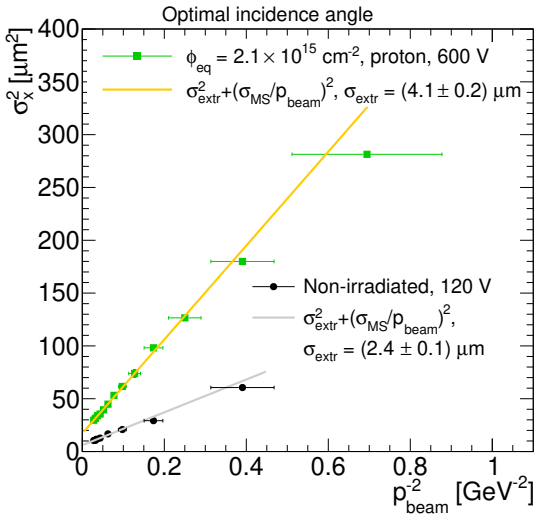


FIGURE 14.2: Spatial resolution squared as a function of the inverse of the beam momentum squared p_{beam}^{-2} for a non-irradiated and a proton-irradiated sensor. The function $\sigma_x^2 = \sigma_{\text{extr}}^2 + (\sigma_{\text{MS}}/p_{\text{beam}})^2$ has been fitted to the data (solid lines). The uncertainties on p_{beam} represent the absolute momentum spread σ_p introduced in Sec. 13.2. Published in Ref. [249].

sensors bump bonded to ROC4Sens analog readout-chips.

The measurements were performed in the DESY test beam with three parallel layers of sensors, not relying on any external tracking detector. The material budget in front and between each sensor is minimized to reduce the effects of multiple Coulomb scattering of the low energy electron beam, allowing to perform spatial resolution measurements with a pixel pitch of 25 µm at different beam incidence angles and while cooling irradiated devices.

Track selections have been studied and a set of requirements to compare sensors with different fluences has been identified. The effects on the spatial resolution of different thresholds and requirements on the cluster charge have also been presented. A new observable to define the spatial resolution has been defined and compared to methods used in literature.

Four different sensor designs are compared using non-irradiated samples and no significant difference in the spatial resolution is found.

Beam incidence angles between 0° and 30° were investigated. The Center-of-Gravity method with threshold cuts is used consistently for all the analysis to determine the position of the beam particle, which causes uncertainties of the determination of the resolution at small angles. While the absolute results may still be improved with optimized algorithms, here emphasis is given to the relative variation of resolution after irradiation and as a function of pixel threshold and beam incidence angle.

The spatial resolution determined varies between 3 and 7 µm for the non-irradiated sensor and between 5 and 8 µm for the irradiated sensors. The angle expected for the optimal resolution of a non-irradiated detector is $9.5^\circ = \arctan(\text{pitch}_x/t)$, which agrees with the values observed for the non-irradiated sensor and the proton-irradiated sensor at $\phi_{\text{eq}} = 2.1 \times 10^{15}\,\text{cm}^{-2}$. For the neutron-irradiated sensor with $\phi_{\text{eq}} = 3.6 \times 10^{15}\,\text{cm}^{-2}$ the optimal resolution is observed at 12°. Proton and neutron irradiated sensors achieve similar spatial resolutions.

Measurements for the neutron-irradiated sensor at $\phi_{\text{eq}} = 3.6 \times 10^{15}\,\text{cm}^{-2}$ were

repeated at lower bias voltages, showing a deterioration of the position resolution and a shift of the optimal angle to higher values. Above a bias voltage of 600 V both the resolution and the optimal angle converge to stable values.

A spatial resolution of $2.4 \pm 0.1\,\mu m$ was found for a non-irradiated sensor at the optimal angle when extrapolating to infinite beam momentum. For the proton-irradiated sample the extrapolated spatial resolution is $4.1 \pm 0.1\,\mu m$. Excellent resolution is therefore maintained up to the tested fluences of $\phi_{eq} = 2.1 \times 10^{15}\,cm^{-2}$ and $\phi_{eq} = 3.6 \times 10^{15}\,cm^{-2}$, which correspond to the lifetime fluence of layer 3 and more than 70% of the lifetime fluence of layer 2 of the CMS Phase-2 Inner Tracker, respectively.

The results presented in this thesis show that the tested sensors are suitable candidates for high precision measurements at the HL-LHC. Results in Refs. [241, 242] show that the sensors also maintain a high hit efficiency after irradiation. The spatial resolution for sensors with 50 µm pitch is presented in Refs. [241, 242]. The sensors employed in this work and in Refs. [241, 242] are bump-bonded to an analog readout chip with 12 bit resolution to allow for detailed studies of the sensor properties. The chip that will be used in the Phase-2 IT adopts a Time Over Threshold (ToT) approach and has a 4 bit resolution. Sensors similar to the ones investigated in this thesis are also being tested with such readout [243] to demonstrate their suitability for the Phase-2 upgrade of the CMS experiment at the HL-LHC.

14.3 Outlook

The analysis presented above aims at a detailed study of the sensor properties, understanding the effect that parameters like the applied threshold have on the spatial resolution. While the ROC4Sens analog readout chip with 12 bit resolution is suited for such in-depth sensor studies, it does not reflect the read out type that will be used in the upgraded detector. Thus, in the context of Phase-2, similar studies are being repeated with a prototype of the final chip to evaluate the effect that the ToT read out has on the resolution. Such measurements have been performed for non irradiated sensors [243, 272] using standard telescopes for the track reconstruction and measurements on irradiated modules are being carried out, with the challenges described in Sec. 13.2.1, where the motivation for this study was described.

The experimental setup used for this work, which features a low material budget, allows the measurement of the resolution of irradiated sensors, proving the validity of the method, but only fluences up to $\phi_{eq} = 3.6 \times 10^{15}\,cm^{-2}$, which is below the expected lifetime fluence of the Phase-2 IT, could be tested. The measurement of sensors irradiated to higher fluences was prevented by the limited cooling possibility of the *Dreimaster*. Improving the thermal isolation of the devices, while keeping a reasonably low material budget, would allow reaching lower temperatures and thus testing sensors with higher fluences.

As it was introduced in Sec. 13.1, the spatial resolution of segmented detectors strongly depends on the pitch size. For this work the smallest pitch among those proposed for the detector upgrade was tested but more and eventually even smaller pitch sizes could be probed with this setup, for instance for the Phase-3 upgrade of the CMS IT which will feature even finer segmentation.

Concerning the method, the Center-of-Gravity algorithm with threshold cuts has been adopted consistently for all the analysis to determine the position of the beam particle, as emphasis is given to the relative variation of resolution after irradiation and as a function of pixel threshold and beam incidence angle. The absolute

measured resolution may still be improved with optimized algorithms, such as the η-algorithm [283, 284] for small incidence angles and the head-tail algorithm [285] at large incidence angles.

The measurement of non-irradiated sensors presented in this work are compared to the simulation developed in Ref. [286]: The average cluster size agrees to better than 0.1 and the spatial resolution agrees within 0.2 µm up to 25°. While a limitation of the CoG method in the position determination at small incidence angles was described in Sec. 13.2.4, Ref. [286] presents a method to correct the position determination and a considerable improvement to the measured resolution is expected when applying such corrections. The results in Ref. [286] validate the quality of the data collected and suggest a method to increase the accuracy of the position measurement with segmented detectors.

Part III

Summary

15 | Summary

In this thesis, different aspects of the search for new physics have been presented, from analyzing the data collected at the LHC to the validation of new techniques and detector development for future applications.

Several theoretical models address the shortcomings of the standard model predicting the existence of new particles with masses of the order of the TeV that can be produced at colliders. The analysis framework described in this thesis allows one single search to probe many proposed extensions of the standard model and can be further adapted to test even more exotic scenarios. Moreover, the method's flexibility permitted to perform the first combined search for heavy resonances decaying in hadronic WW, WZ, ZZ, WH and ZH final states, presented in this thesis. Dedicated categories were introduced to distinguish resonances produced through gluon-gluon fusion (ggF) and quark-antiquark annihilation (DY) from vector boson fusion (VBF), characterized by two extra jets in the final state. The bosons produced in the decay of these massive resonances have a high momentum. Thus, they decay into two collimated quarks, and the sprays of particles produced in the quarks hadronization are merged into a single large radius jet. Therefore, jet substructure techniques are employed to identify jets originating from vector bosons or containing b-quark pairs from jets stemming from a single quark or gluon. Algorithms based on machine learning (ML) techniques are used to achieve the highest discriminating power. This study's background estimation and signal extraction procedures rely on a multidimensional maximum likelihood fit to the masses of the jets originating from the bosons and their combined dijet system. The signal is resonant in the considered dimensions and a three-dimensional search for a bump on smoothly falling backgrounds is performed. A decorrelation method has been validated to prevent ML-based algorithms from introducing sculpting in the shape of the jet mass used as observable in the fit.

The search presented in this thesis analyzed 138 fb^{-1} of pp-collisions data collected by the CMS experiment in the 2016–2018 data taking periods at a centre-of-mass energy of 13 TeV. While local excesses on the SM expectation have been observed, their global significance in the broad analysis phase space does not indicate a deviation from the SM. Exclusion limits are set on 16 different signal hypotheses taking into account the resonances' various production and decay modes. Moreover, since resonances can decay into multiple combinations of boson pairs, 10 such combinations are considered to achieve the best sensitivity. The analysis establishes a up to an order of magnitude improvement on limits on signal production cross section with respect to previous searches. Spin-0 resonances produced through ggF are excluded up to masses of 2.7 TeV. This search can exclude ggF produced spin-2 resonances decaying hadronically for masses below 1.4 TeV, to which a previous search based on 77 fb^{-1} was not sensitive to. Concerning the VBF production mode, upper limits on the production cross section are set from 3 fb for a resonance mass of

1.3 TeV to 0.1 fb at 6 TeV for the Radion$\rightarrow$ VV. Furthermore, limits on the production cross section of VBF produced $G^{\tilde{\kappa}=0.5}_{\text{bulk}} \rightarrow$ VV combining all hadronic WW and ZZ final states are set for the first time, from 4 fb for a resonance mass of 1.3 TeV to 0.2 fb at 6 TeV. DY produced spin-1 resonances are excluded up to 4.8 TeV combining VV and VH channels, the best to date mass exclusion limits. In addition, for resonances produced through VBF, limits on the production cross section are set from 7–10 fb at 1.3 TeV to 0.3–0.4 fb at 6 TeV, for the first time in the WH channel, for the first time in the all-jets final states and for the first time combining VV and VH decay modes.

Searches for heavy resonances as the ones described above will significantly benefit from the future data taking periods at the LHC, where datasets more than 10 times larger than those analyzed in this thesis will be collected at the increased centre-of-mass energy of 14 TeV. Furthermore, upgrades of the CMS sub-detectors are planned for the high-luminosity phase of the LHC. In this thesis, a simulation study of the performance of the future tracker is described, conducted in the context of the upgrade technical design report. It shows that the finer granularity of the upgraded tracker improves the reconstruction of high momentum jets and leads to better identification of hadronically decaying high momentum bosons. Furthermore, for 4 TeV signals like the ones described in the first part of this thesis, two times better signal efficiency and cross section limits are predicted from such improvements, for typical analysis conditions.

A finer segmentation of the detectors improves the spatial resolution, the precision of the measurement of a particle position. However, during operation at the HL-LHC, the number of pp interactions in the same bunch crossing could reach an average of 200 and the detectors will receive a high radiation dose, which causes a degradation of their performance. The pixel detector, the closest system to the interaction point, will face the most challenging conditions.

However, radiation effects were relevant also for the pixel detectors used in the 2016–2018 data taking period and should be taken into account to precisely compare the simulated event to the collected data, a crucial step in almost all performed analyses. This thesis features the validation of a method introduced to include the effect of radiation damage in the CMS simulation framework without significantly increasing the computing time. A weight, obtained from an external setup, taking into account the radiation effects is applied to the charges simulated in absence of radiation damage. The implementation has been tested, and the obtained simulation was compared to the recorded data, showing an improved agreement with respect to the non-weighted simulation.

To operate in the HL-LHC conditions, a new generation of radiation tolerant, fine pitch pixel sensors is being developed. In the second part of this thesis, a detailed study of the spatial resolution of prototype planar silicon pixel sensors for operation at the HL-LHC has been performed, comparing the performance before and after irradiation. The prototypes are characterized by a $100\times25\,\mu m^2$ pitch, six times smaller than the one currently in use in the CMS experiment, to ensure efficient tracking in a high pileup environment and improve the two-track separation, even inside highly energetic jets. To perform precise measurements in the shorter pitch direction, a dedicated setup is used, composed of three parallel planes of sensors. The measurements were performed in the DESY II test beam facility with a 5 GeV electron beam. The setup minimizes the material between sensors to mitigate the contribution due to multiple scattering while allowing the cooling of irradiated detectors. The tested irradiated sensors are a sensor irradiated with neutrons to $\phi_{\text{eq}} = 3.6 \times 10^{15}\,\text{cm}^{-2}$ and

of two sensors irradiated with protons to $\phi_{eq} = 2.1 \times 10^{15}\,cm^{-2}$, corresponding to more than 70% of the full lifetime fluence of the second barrel layer and to the full lifetime fluence of the third layer, respectively.

The thesis presents a review of the different quantities adopted in literature to define the spatial resolution and introduces a new variable. The effect on the resolution of various track selection criteria has been assessed. The measurements were repeated for different beam incidence angles to determine the angle providing the best resolution. The 9.5° angle expected for the optimal resolution of a non-irradiated detector agrees with the values observed for non-irradiated and proton-irradiated sensors. For the neutron-irradiated sensor the optimal resolution is observed at 12°. Proton and neutron irradiated sensors achieve similar spatial resolutions. Furthermore, the measurements at the optimal angle were repeated for different beam momenta to extrapolate the resolution at infinite beam momentum, where the multiple scattering contribution can be neglected: a spatial resolution of $2.4 \pm 0.1\,\mu m$ ($4.1 \pm 0.1\,\mu m$) was found for a non-irradiated (proton-irradiated) sensor. The results presented in this thesis show that the tested sensors are suitable candidates for high precision measurements at the HL-LHC.

A | Simulated and data samples

The first step of the signal production was curated by the author of this thesis, with the help of Dr. Anna Benecke for part of the production. The simulated samples, produced centrally by the CMS Collaboration, need several level of post-processing before they can be used in the multidimensional fit. The first processing part was done by the author of this thesis together with Dr. Anna Benecke and Dr. Clemens Lange. While the author of the thesis and Dr. Anna Benecke contributed to the software needed for the handling of the second processing part, this was done by Dr. Jennifer Ngadiuba, that took care also of the NLO correction factors for the V+jets simulation. Dr. Gerrit van Onsem took care of the rescaling of the 2017 and 2018 signal sample weights.

A.1 Signal samples

The complete list of signal samples used in this thesis can be found in Table A.1. For each signal benchmark masses of 1.2, 1.4, 1.6, 1.8, 2.0, 2.5, 3.0, 3.5, 4, 4.5, 5.0, 5.5, 6.0 TeV have been generated with a number of events between 30k and 50k. The campaigns and global tags used to produced the samples are reported in Table A.2 for each data-taking year.

The BSM models predicting the signals studied in this thesis are described in Sec. 2.3. For the Bulk Graviton samples, a WED benchmark with $\tilde{k} = 0.5$ is used, while for the Radion the chosen benchmark model parameters are $\Lambda_R = 3$ TeV and $kl = 35$. For the W' and Z', the heavy vector triplet (HVT) framework is used to assume different scenarios. In this framework the new heavy vector bosons, W' and Z', couple to the Higgs and SM gauge bosons with the parameters c_H and g_V, and to the fermions via the combination $(g^2/g_V)c_F$, where c_F is the fermion coupling and g is the SM $SU(2)_L$ gauge coupling. The couplings are expected to be of order unity in most models. Three benchmark models, denoted as models A, B, and C are considered. In model A (g_V = 1, c_H = -0.556, c_F= -1.316), the branching fractions of decays to fermions and gauge bosons are comparable and the heavy resonances decay predominantly to fermions, as is the case in some extensions of the SM gauge group. In model B (g_V = 3, c_H = -0.976, c_F = 1.024), the fermionic couplings are suppressed, as in composite Higgs models. In model C (g_V = 1, c_H = 1, c_F = 0), the fermionic couplings are set to zero, so the resonances are produced only through vector boson fusion (VBF) and decay exclusively to a pair of SM bosons.

Simulated events of these signals are generated at leading-order (LO) in quantum chromodynamics (QCD) with MADGRAPH5_aMC@NLO version 2.4.2 and 2.6.0 and the MLM matching scheme [290], while the hadronization showering is simulated with PYTHIA version 8.226 for 2016 and 8.230 for 2017 and 2018. No extra jets are generated in any of the samples except for the VBF production mode where the two forward jets are fully simulated. Table A.1 lists also the PDF sets used for the 2016 sample production. For 2017 and 2018 the NNPDF3.1 NNLO

(NNPDF31_nnlo_hessian_pdfas) set was used. For high resonance masses, the use of this not positive definite pdf set may yield to a large fraction of negative (unphysical) event weights. To avoid this issue, the 2017 and 2018 samples have been rescaled to NNPDF30_lo_as_0130. This correction is achieved by multiplying the generator event weight by the ratio of the weight corresponding to the LO PDF set to the weight corresponding to the NNLO PDF set and is applied on an event-by-event basis.

The CUETP8M1 [123] (CP5 [124]) underlying event tunes in Pythia8 for 2016 (2017 and 2018) conditions.

The NLO theoretical cross sections for the bulk graviton and radion produced through gluon fusion are taken from [44] and [46], respectively. In the VBF production mode the NLO theoretical cross sections are taken from [45] and [47], for the bulk graviton and radion, respectively. These are all calculated with MADGRAPH5_aMC@NLO interfaced to LHAPDF6; the PDF4LHC15_nlo_mc_pdfas set and the four flavor scheme for the proton are used. The product of the predicted production cross section and the branching ratio for the decay into two SM bosons for these benchmarks are shown in Fig. 2.5 as a function of the resonance mass hypothesis.

The LO cross sections for the HVT model have been derived using the online tools provided by the authors of the framework [51]. For completeness these are shown for model B in Fig. 2.7 for the DY production mode and together with the branching ratios for the decays of the resonance to two SM bosons. Cross sections for the model C describing the VBF production mode are also shown in the same figure. Actual numbers can be found in Ref. [53] for model A and B, and in Ref. [54] for model C.

Sample name	pdf set 2016
BulkGravToWW_narrow_M-**X**_13TeV-madgraph	NNPDF30_lo_as_0130
VBF_BulkGravToWWinclusive_narrow_M-**X**_13TeV-madgraph-pythia8	NNPDF30_lo_as_0130_nf_4
BulkGravToZZToZhadZhad_narrow_M-**X**_13TeV-madgraph	NNPDF30_lo_as_0130
VBF_BulkGravToZZinclusive_narrow_M-**X**_13TeV-madgraph-pythia8	NNPDF30_lo_as_0130_nf_4
WprimeToWZToWhadZhad_narrow_M-**X**_13TeV-madgraph	NNPDF30_lo_as_0130
VBF_WprimeToWZinclusive_narrow_M-**X**_13TeV-madgraph-pythia8	NNPDF30_lo_as_0130_nf_4
WprimeToWhToWhadhinc_narrow_M-**X**_13TeV-madgraph-pythia8	NNPDF30_lo_as_0130_nf_4
VBF_WprimeToWhToWhadhinc_narrow_M**X**_13TeV-madgraph-pythia	NNPDF30_lo_as_0130_nf_4
ZprimeToWW_narrow_M-**X**_13TeV-madgraph	NNPDF30_lo_as_0130
VBF_ZprimeToWWinclusive_narrow_M-**X**_13TeV-madgraph-pythia8	NNPDF30_lo_as_0130_nf_4
ZprimeToZhToZhadhinc_narrow_M-**X**_13TeV-madgraph-pythia8	NNPDF30_lo_as_0130_nf_4
VBF_ZprimeToZhToZhadhinc_narrow_M**X**_13TeV-madgraph-pythia	NNPDF30_lo_as_0130
RadionToWW_narrow_M-**X**_13TeV-madgraph	NNPDF30_lo_as_0130
VBF_RadionToWW_narrow_M-**X**_13TeV-madgraph	NNPDF30_lo_as_0130_nf_4
RadionToZZ_narrow_M-**X**_13TeV-madgraph	NNPDF30_lo_as_0130
VBF_RadionToZZ_narrow_M-**X**_13TeV-madgraph	NNPDF30_lo_as_0130_nf_4

TABLE A.1: Signal samples and PDF sets used for this analysis.

Year	**Campaign**
2016	RunIISummer16MiniAODv3
2017	RunIIFall17MiniAODv2
2018	RunIIAutumn18MiniAOD

TABLE A.2: MC campaigns and corresponding Global Tag used to simulate signal events for the three data-taking years.

A.2 Background samples

Simulated samples of the SM background processes are used to optimize the analysis and provide flexible background templates. QCD multijet production is simulated with three generator configurations:

1. PYTHIA8 standalone;
2. the LO mode of MADGRAPH5_aMC@NLO matched with PYTHIA8;
3. HERWIG++ 2.7.1 with tune CUETHS1 [123] for 2016, and HERWIG7 with tune CH3 for 2017 and 2018.

Top quark pair production is modelled at NLO with POWHEG and showered with PYTHIA. In order to account for differences between the shapes of the measured and simulated top p_T spectra, due to the absence of NNLO order in the simulation, a reweight is applied to the $t\bar{t}$ MC samples using p_T-dependent scale factors. For a given event, the weight $SF(p_T) = e^{(\alpha-\beta p_T)}$ derived in Ref. [188, 189] are applied, with the nominal parameters $\alpha = 0.0615$ and $\beta = 0.0005$.

W+jets and Z+jets production are simulated with the LO mode of MADGRAPH5_aMC@NLO matched with PYTHIA. Following the same procedure as [291] and [292] these LO simulation samples are reweighted using higher-order corrections separately corresponding to NLO QCD and NLO EW terms. Scale factors corresponding to NLO EW corrections are obtained from Ref. [293] and applied as a function of the generator-level boson p_T. As these scale factors are provided only up to a p_T of 1.2 TeV, a fit to the available data with an exponential function is performed to obtain a smooth extrapolation up to higher p_T.

Scale factors corresponding to NLO QCD corrections for W and Z production are obtained comparing the central CMS samples produced in the "Fall17" campaign, in which large LO and NLO samples are available. The scale factors are derived by normalizing the distributions to their respective cross sections, and then dividing them as SF = NLO/LO. The EWK and QCD NLO scale factors are shown in Fig. A.1.

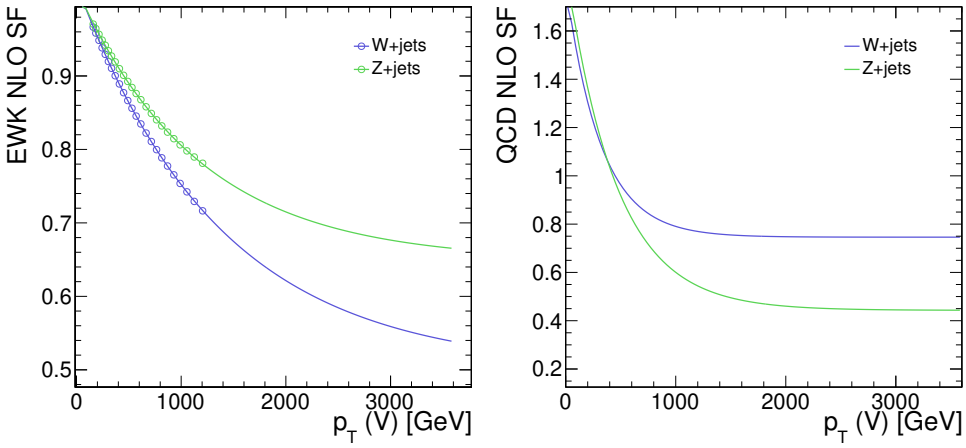


FIGURE A.1: EW (left) and QCD (right) NLO correction factors for W and Z production as a function of the generator-level boson p_T.

All samples are processed through a GEANT4-based simulation of the CMS detector. To simulate the effect of additional proton-proton collisions within the same or adjacent bunch crossings (pileup), additional inelastic events are generated using

PYTHIA and superimposed on the hard-scattering events. The MC simulated events are weighted to reproduce the distribution of the number of pileup interactions observed in data. The campaigns are the same as for the signal samples reported in Table A.2 for each data-taking year. The complete list of background samples used in the analysis can be found in Tables A.3–A.5 for the three data-taking years together with the corresponding cross sections. Cross sections for the H_T and p_T binned QCD samples have been computed using the tools provided in Ref. [294]; the W/Z+jets from Ref. [295]; the SM VV diboson from Ref. [296–298]; the SM VH from Ref. [299]; the single top from Ref. [300, 301]. The $t\bar{t}$ cross sections are obtained multiplying the inclusive NNLO cross section of 831.76 pb from Ref. [188, 189] by the filter efficiency (0.0921 for `TT_Mtt-700to1000` and 0.02474 for `TT_Mtt-1000toInf`).

A.2.1 Pileup Reweighting

As is shown in Fig. A.2, data has a slightly different pileup profile than what was used to generate MC and therefore a weight is applied to the MC simulation to match the number of primary vertices observed in data, separately for each period.The minimum bias cross section of 69.2 mb is used for all datasets. As recommended, an uncertainty of 4.6% on the minimum bias cross section is included as a systematic.

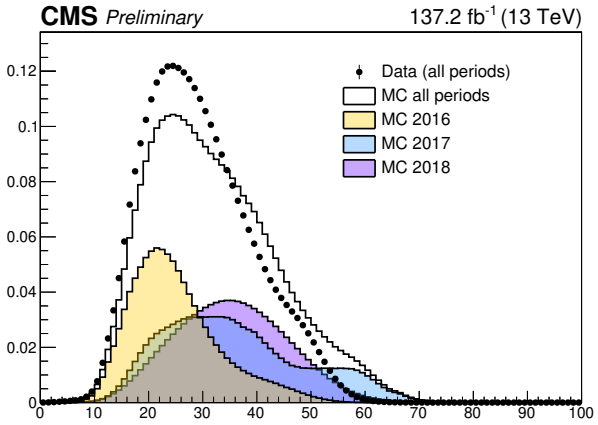


FIGURE A.2: Pileup profile of data versus MC for the sum of the three data taking periods. Due to the difference in these profiles, a reweighting scheme is used in order to match MC to the number of primary vertices observed in data.

A.3 Dataset

This thesis is based on the data collected in 2016, 2017 and 2018, corresponding to an integrated luminosity of 138 fb^{-1}. The analysis has been performed with the `CMSSW 10_2_X` release. The data sample names are listed in Table A.6, where the corresponding integrated luminosity are reported as calculated with all the runs certified as "good" for all subsystems.

Sample name	N_{events}	Cross section [pb]
/QCD_Pt-15to7000_TuneCUETHS1_Flat_13TeV_herwigpp/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM	9936496	–
/QCD_Pt_170to300_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	14796774	117276
/QCD_Pt_300to470_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	22470404	7823
/QCD_Pt_470to600_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	19566912	648.2
/QCD_Pt_600to800_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	32569319	186.9
/QCD_Pt_800to1000_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	19697092	32.293
/QCD_Pt_1000to1400_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	9846615	9.4183
/QCD_Pt_1400to1800_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	2873427	0.84265
/QCD_Pt_1800to2400_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	1982038	0.114943
/QCD_Pt_2400to3200_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	996130	0.00682981
/QCD_Pt_3200toInf_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM	391735	0.000165445
/QCD_HT100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM	82293477	27990000
/QCD_HT200to300_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	57580393	1712000
/QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	54552852	347700
/QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	62622029	32100
/QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	37233786	6831
/QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	15210939	1207
/QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	11839357	119.9
/QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	6019541	25.24
/TT_Mtt-700to1000_TuneCUETP8M2T4_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM	38591757	76.605
/TT_Mtt-1000toInf_TuneCUETP8M2T4_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM	24577608	20.578
/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM	998276	35.85
/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM	992024	35.85
/WW_TuneCUETP8M1_13TeV-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	7982180	118.7
/WWTo4Q_13TeV-powheg/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM	1998400	53.94
/WZ_TuneCUETP8M1_13TeV-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	3997571	47.13
/ZZ_TuneCUETP8M1_13TeV-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM	1988098	16.5
/ZZTo4Q_13TeV_amcatnloFXFX_madspin_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM	30636849	8.06
/WplusH_HToBB_WToQQ_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM	499597	0.840*0.5824*0.6741
/WminusH_HToBB_WToQQ_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM	499994	0.533*0.5824*0.6741
/ZH_HToBB_ZToQQ_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM	499200	0.8839*0.5824*0.69911

TABLE A.3: List of 2016 background samples used in this analysis together with the corresponding cross section.

Sample name	N_{events}	Cross section [pb]
/QCD_Pt-15to7000_TuneCH3_Flat_13TeV_herwig7/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	19286000	–
/QCD_Pt_170to300_TuneCP5_13TeV_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14*/MINIAODSIM	56298920	117276
/QCD_Pt_300to470_TuneCP5_13TeV_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14*/MINIAODSIM	111229780	7823
/QCD_Pt_470to600_TuneCP5_13TeV_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14*/MINIAODSIM	55503028	648.2
/QCD_Pt_600to800_TuneCP5_13TeV_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14*/MINIAODSIM	128548964	186.9
/QCD_Pt_800to1000_TuneCP5_13TeV_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14*/MINIAODSIM	78116008	32.293
/QCD_Pt_1000to1400_TuneCP5_13TeV_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14*/MINIAODSIM	35819814	9.4183
/QCD_Pt_1400to1800_TuneCP5_13TeV_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14*/MINIAODSIM	11353270	0.84265
/QCD_Pt_1800to2400_TuneCP5_13TeV_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	2923941	0.114943
/QCD_Pt_2400to3200_TuneCP5_13TeV_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	1910526	0.00682981
/QCD_Pt_3200toInf_TuneCP5_13TeV_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14*/MINIAODSIM	956837	0.000165445
/QCD_HT100to200_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14*/MINIAODSIM	171876686	27990000
/QCD_HT200to300_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v2/MINIAODSIM	59427619	1712000
/QCD_HT300to500_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v1/MINIAODSIM	59569132	347700
/QCD_HT500to700_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM	56207744	32100
/QCD_HT700to1000_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v2/MINIAODSIM	46840955	6831
/QCD_HT1000to1500_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v1/MINIAODSIM	16882838	1207
/QCD_HT1500to2000_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM	11634434	119.9
/QCD_HT2000toInf_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM	5941306	25.24
/WJetsToQQ_HT400to600_qc19_3j_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	9738307	315.6
/WJetsToQQ_HT600to800_qc19_3j_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	8798398	68.57
/WJetsToQQ_HT-800toInf_qc19_3j_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	8081153	34.9
/ZJetsToQQ_HT400to600_qc19_4j_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	10316727	145.4
/ZJetsToQQ_HT600to800_qc19_4j_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	8882592	34.0
/ZJetsToQQ_HT-800toInf_qc19_4j_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	7818660	18.67
/TT_Mtt-700to1000_TuneCP5_PSweights_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14_ext1-v4/MINIAODSIM	35495198	76.605
/TT_Mtt-1000toInf_TuneCP5_PSweights_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14_ext1-v1/MINIAODSIM	19153722	20.578
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM	7977430	35.85
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM	7794186	35.85
/WW_TuneCP5_13TeV-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM	7765828	118.7
/WWTo4Q_NNPDF31_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v1/MINIAODSIM	2000000	53.94
/WZ_TuneCP5_13TeV-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	3928630	47.13
/ZZ_TuneCP5_13TeV-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v2/MINIAODSIM	1925931	16.5
/WplusH_HToBB_WToQQ_M125_13TeV_powheg_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	504997	0.840*0.5824*0.6741
/WminusH_HToBB_WToQQ_M125_13TeV_powheg_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	484662	0.533*0.5824*0.6741
/ZH_HToBB_ZToQQ_M125_13TeV_powheg_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	488450	0.8839*0.5824*0.69911

TABLE A.4: List of 2017 background samples used in this analysis together with the corresponding cross section.

Sample name	N_{events}	Cross section [pb]
/QCD_Pt-15to7000_TuneCH3_Flat_13TeV_herwig7/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	19481000	–
/QCD_Pt_170to300_TuneCP5_13TeV_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	29710000	117276
/QCD_Pt_300to470_TuneCP5_13TeV_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	41744000	7823
/QCD_Pt_470to600_TuneCP5_13TeV_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15*/MINIAODSIM	42459973	648.2
/QCD_Pt_600to800_TuneCP5_13TeV_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	64061000	186.9
/QCD_Pt_800to1000_TuneCP5_13TeV_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15_ext1-v2/MINIAODSIM	37598000	32.293
/QCD_Pt_1000to1400_TuneCP5_13TeV_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	18485000	9.4183
/QCD_Pt_1400to1800_TuneCP5_13TeV_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15*/MINIAODSIM	6928000	0.84265
/QCD_Pt_1800to2400_TuneCP5_13TeV_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15*/MINIAODSIM	4017800	0.114943
/QCD_Pt_2400to3200_TuneCP5_13TeV_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15*/MINIAODSIM	2394000	0.00682981
/QCD_Pt_3200toInf_TuneCP5_13TeV_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15*/MINIAODSIM	925000	0.000165445
/QCD_HT100to200_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	93972378	27990000
/QCD_HT200to300_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	54289442	1712000
/QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	54661579	347700
/QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	55152960	32100
/QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	48158738	6831
/QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	15466225	1207
/QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	10955087	119.9
/QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	5475677	25.24
/WJetsToQQ_HT400to600_qc19_3j_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	10071273	315.6
WJetsToQQ_HT600to800_qc19_3j_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	15298056	68.57
/WJetsToQQ_HT-800toInf_qc19_3j_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	14627242	34.9
/ZJetsToQQ_HT400to600_qc19_4j_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	16704355	145.4
ZJetsToQQ_HT600to800_qc19_4j_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	14642701	34.0
/ZJetsToQQ_HT-800toInf_qc19_4j_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	10561192	18.67
/TT_Mtt-700to1000_TuneCP5_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	34823327	76.605
/TT_Mtt-1000toInf_TuneCP5_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v2/MINIAODSIM	22971051	20.578
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15_ext1-v1/MINIAODSIM	7623000	35.85
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15_ext1-v1/MINIAODSIM	9598000	35.85
/WW_TuneCP5_13TeV-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v2/MINIAODSIM	7850000	118.7
/WWTo4Q_NNPDF31_TuneCP5_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	3808800	53.94
/WZ_TuneCP5_13TeV-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v3/MINIAODSIM	3885000	47.13
/ZZ_TuneCP5_13TeV-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v2/MINIAODSIM	1979000	16.5
/WplusH_HToBB_WToQQ_M125_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	746195	0.840*0.5824*0.6741
/WminusH_HToBB_WToQQ_M125_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	778619	0.533*0.5824*0.6741
/ZH_HToBB_ZToQQ_M125_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	761658	0.8839*0.5824*0.69911

TABLE A.5: List of 2018 background samples used in this analysis together with the corresponding cross section.

Sample name	**Integrated luminosity [fb^{-1}]**
/JetHT/Run2016B-17Jul2018_ver1-v1/MINIAOD /JetHT/Run2016B-17Jul2018_ver2-v2/MINIAOD	5.8
/JetHT/Run2016C-17Jul2018-v1/MINIAOD	2.6
/JetHT/Run2016D-17Jul2018-v1/MINIAOD	4.2
/JetHT/Run2016E-17Jul2018-v1/MINIAOD	4.0
/JetHT/Run2016F-17Jul2018-v1/MINIAOD	3.1
/JetHT/Run2016G-17Jul2018-v1/MINIAOD	7.6
/JetHT/Run2016H-17Jul2018-v1/MINIAOD	8.7
/JetHT/Run2017B-31Mar2018-v1/MINIAOD	4.8
/JetHT/Run2017C-31Mar2018-v1/MINIAOD	9.7
/JetHT/Run2017D-31Mar2018-v1/MINIAOD	4.3
/JetHT/Run2017E-31Mar2018-v1/MINIAOD	9.3
/JetHT/Run2017F-31Mar2018-v1/MINIAOD	13.5
/JetHT/Run2018A-17Sep2018-v1/MINIAOD	4.8
/JetHT/Run2018B-17Sep2018-v1/MINIAOD	9.7
/JetHT/Run2018C-17Sep2018-v1/MINIAOD	4.3
/JetHT/Run2018D-PromptReco-v*/MINIAOD	9.3
Total integrated luminosity	137.6

TABLE A.6: Data samples together with the corresponding integrated luminosity.

B Signal and background modelling: additional distributions

B.1 Signal modelling

Figure B.1 shows the 3D signal modelling for VBF produced $W' \rightarrow WH$, $Z' \rightarrow ZH$ and $W' \rightarrow WZ$ signals.

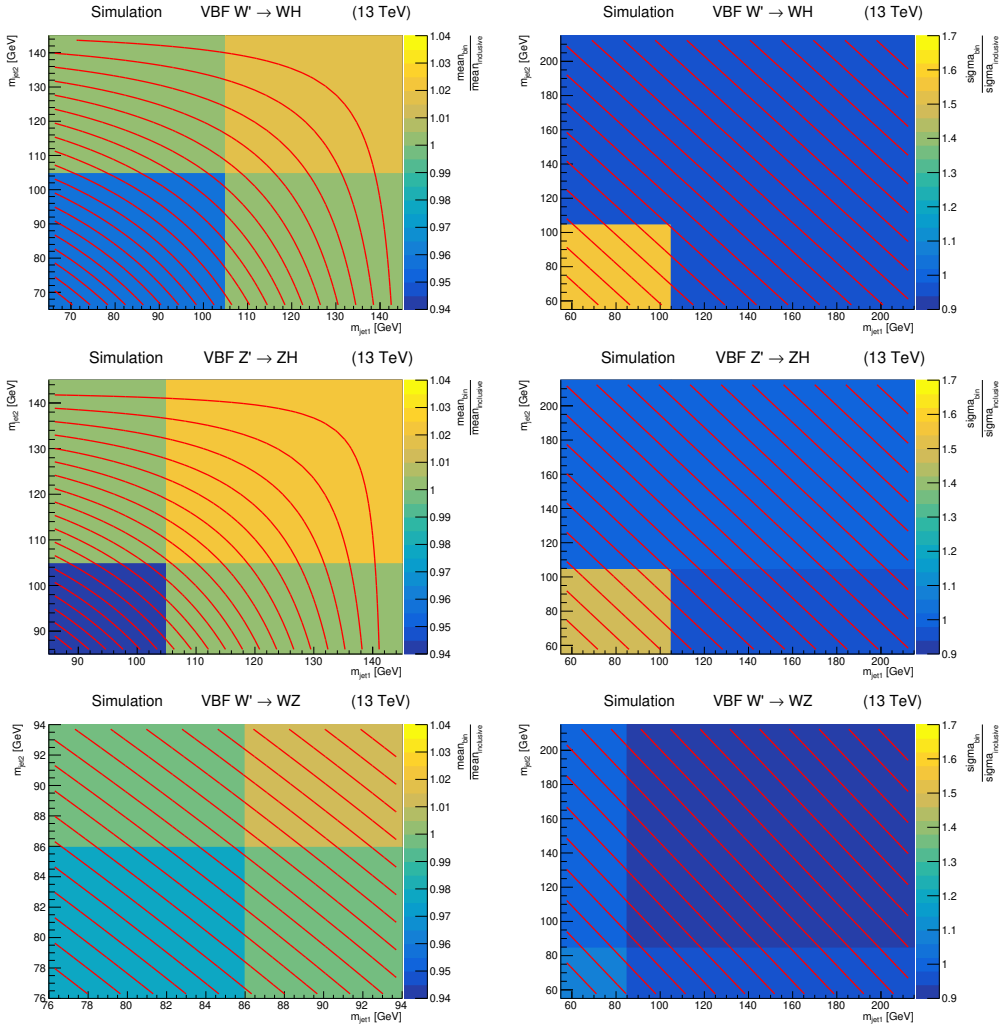


FIGURE B.1: Two-dimensional dependence of the mean (left) and width (right) of the dijet invariant mass distribution on the two jet masses for the VBF $W' \rightarrow WH$ (top), VBF $Z' \rightarrow ZH$ (middle), and VBF $W' \rightarrow WZ$ signal (bottom).

Figure B.2 (Figure B.3 for VBF signals) shows the fit with a spline, linear or parabolic function to the dCB parameters for P_{j1}, i.e. the jet mass as a function of M_X, after only the analysis preselections have been applied, i.e. without applying any H/W-tagging cuts. Figure B.4 (Figure B.5) shows the same for the fits of the dijet invariant mass shapes.

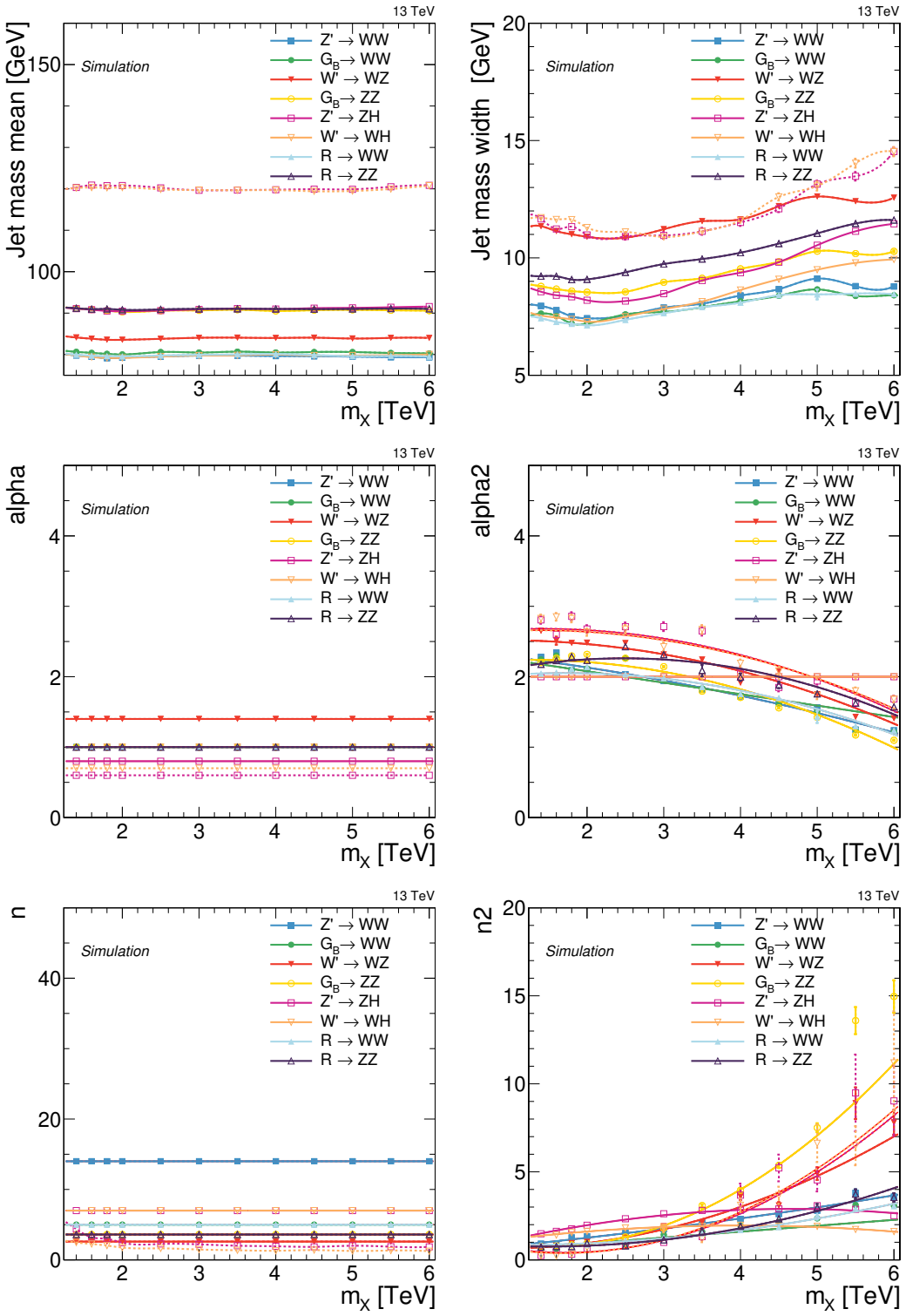


FIGURE B.2: Interpolation of the signal m_{jet1} shape parameters as a function of M_X. The results for m_{jet2} are identical because of the random jet labelling and therefore omitted.

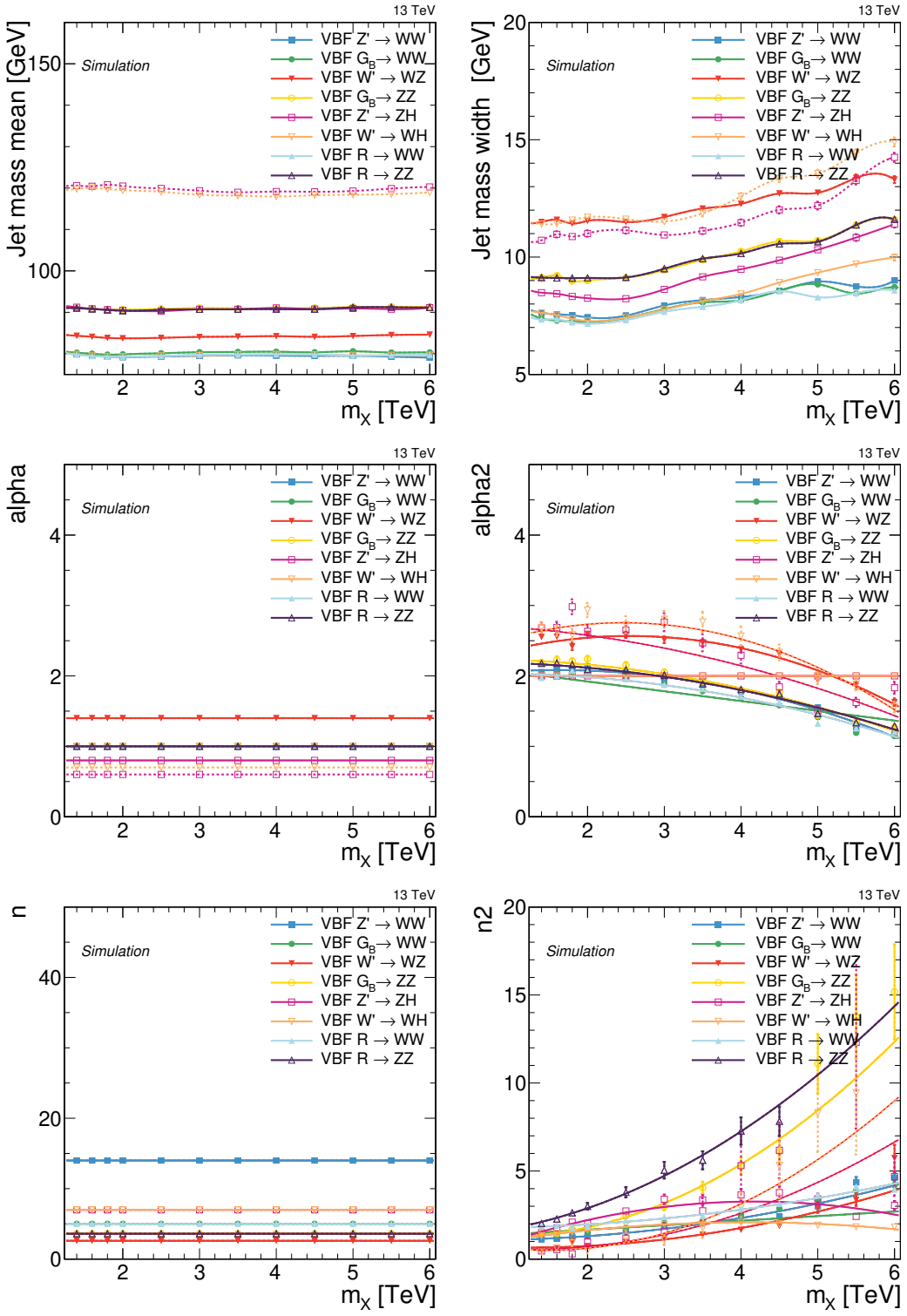


FIGURE B.3: Interpolation of the signal m_{jet1} shape parameters as a function of M_X. The results for m_{jet2} are identical because of the random jet labelling and therefore omitted.

For each mass point $M_X$and each jet tagging category the signal efficiency is also interpolated as a function of M_X(Fig. B.6 and Fig. B.7) in order to extract signal yields for arbitrary resonance masses. The interpolation is obtained with a polynomial function.

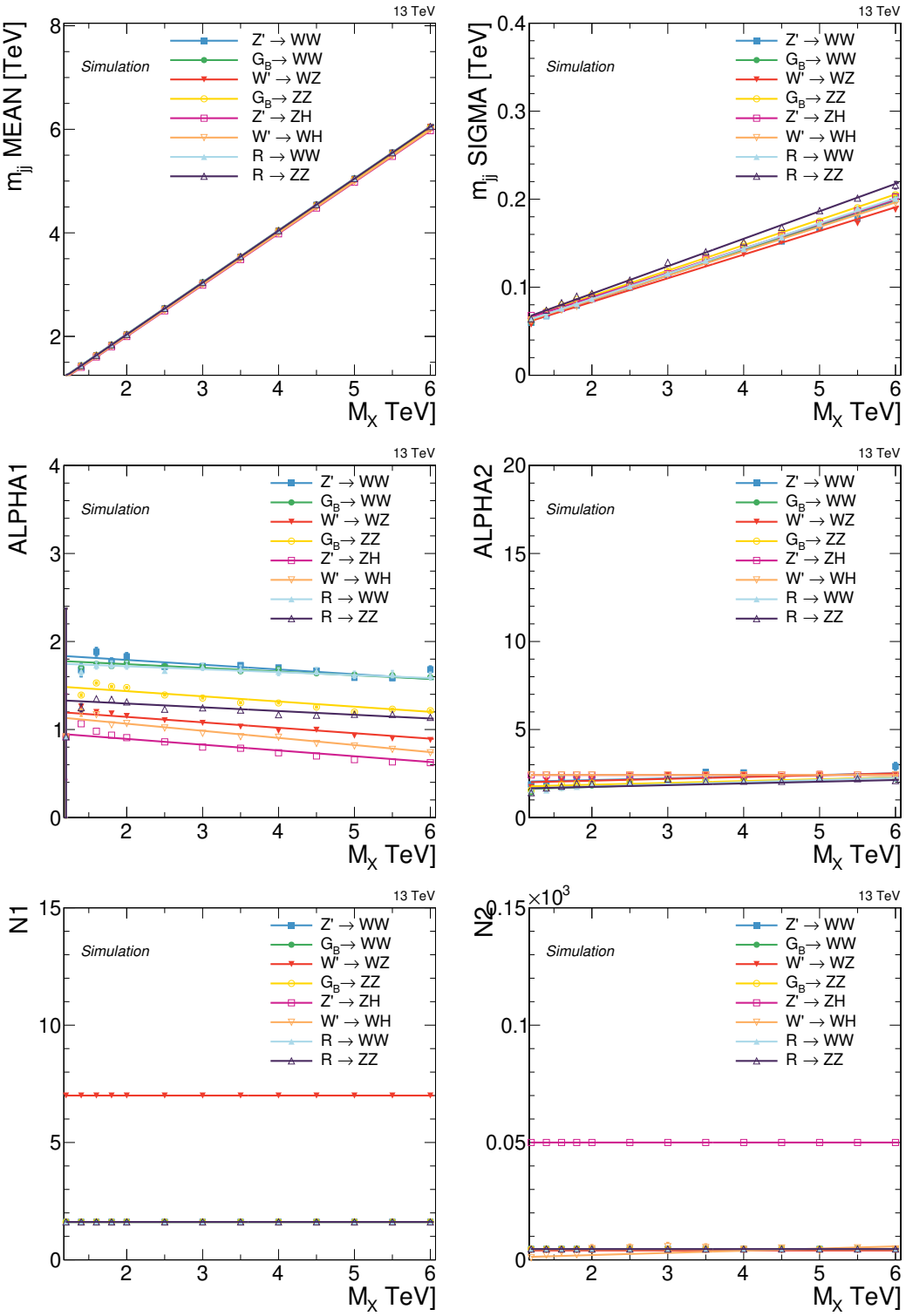


FIGURE B.4: Interpolation of the signal m_{jj} shape parameters as a function of M_X.

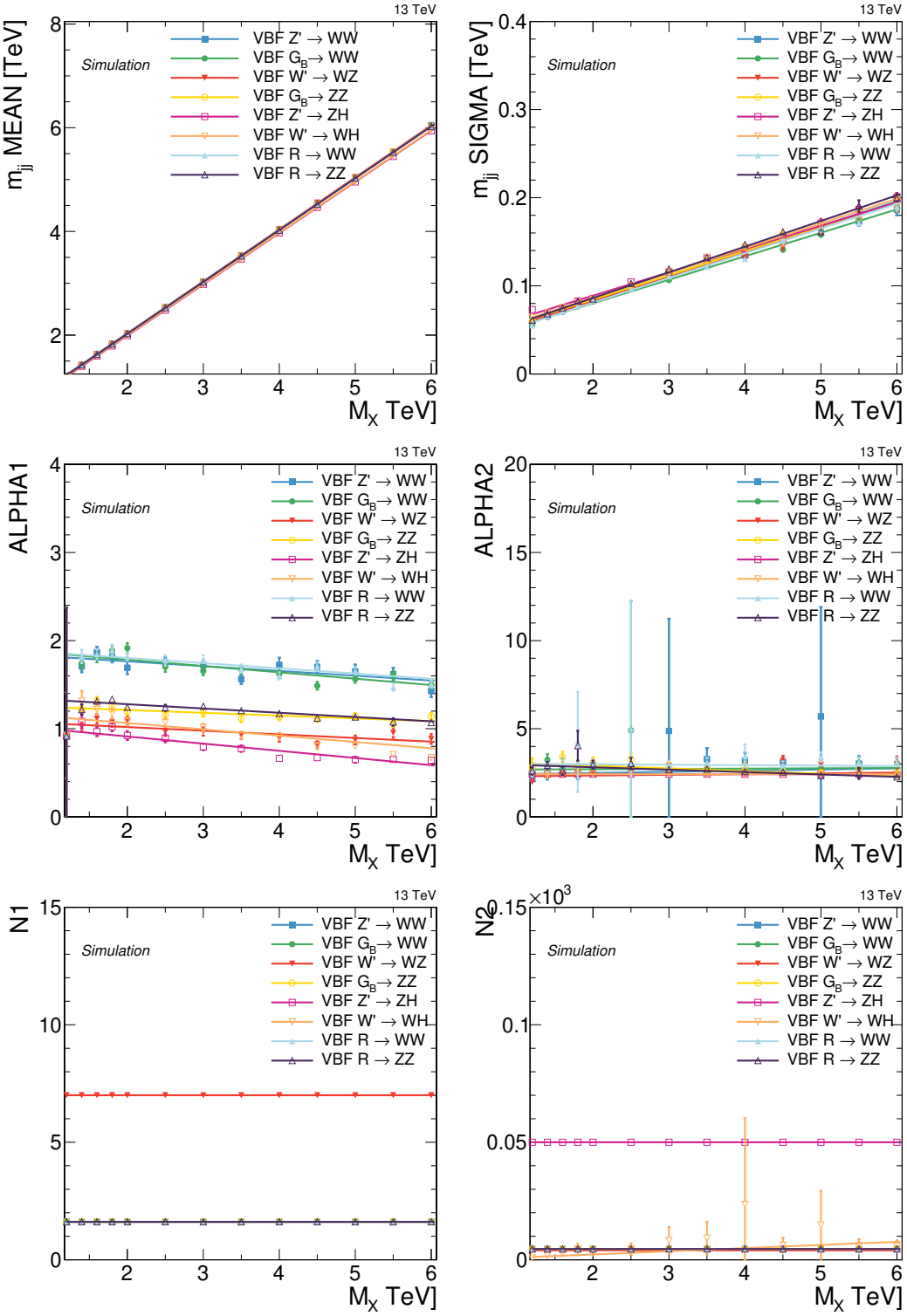


FIGURE B.5: Interpolation of the signal m_{jj} shape parameters as a function of M_X.

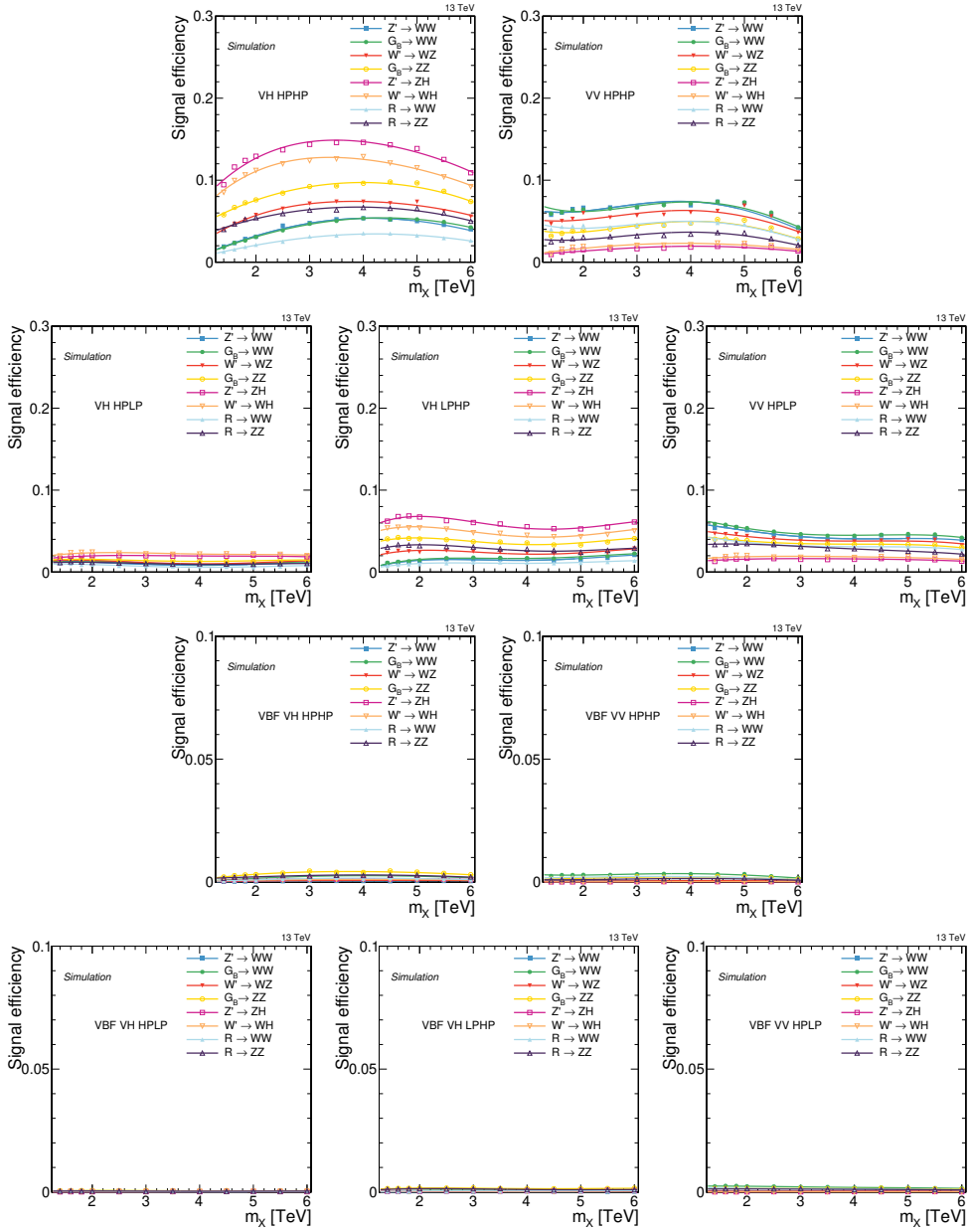


FIGURE B.6: Signal efficiency as a function of M_X for all categories. The polynomial fit is also shown a solid line. The various ggF or DY produced signal models are shown.

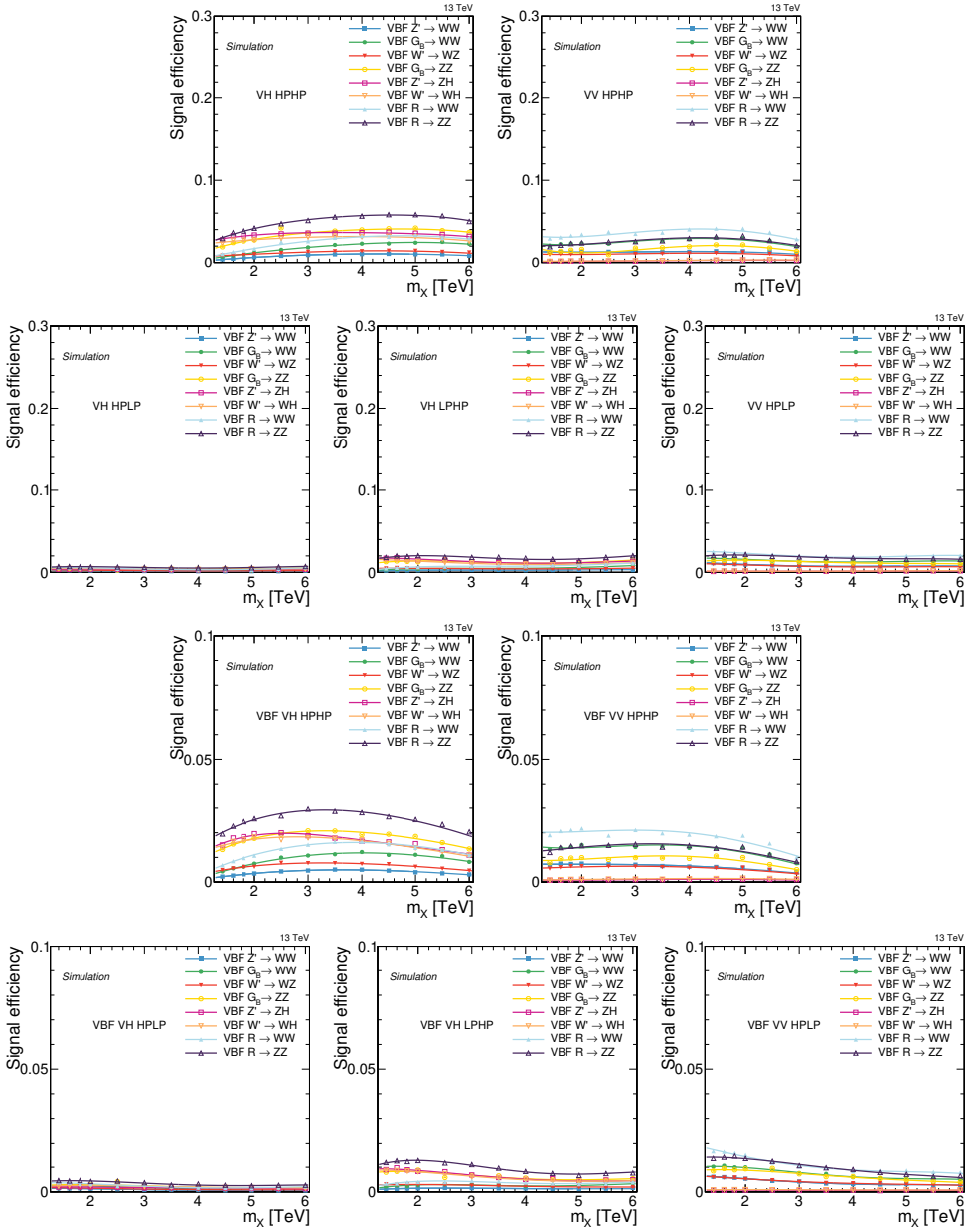


FIGURE B.7: Signal efficiency as a function of M_X for all categories. The polynomial fit is also shown a solid line. The various VBF produced signal models are shown.

B.2 QCD modelling: additional distributions

B.2.1 ggF/DY categories

Figure B.8 shows the comparison between the derived templates and MC simulation for each of the categories for the nominal QCD PYTHIA8 samples. The same is shown for the QCD MADGRAPH5_aMC@NLO (Fig. B.9) and HERWIG (Fig. B.10). For the HERWIG case, only inclusive projections are shown as statistics for this sample is rather limited with respect to the other samples. Good agreement is observed between simulation and templates.

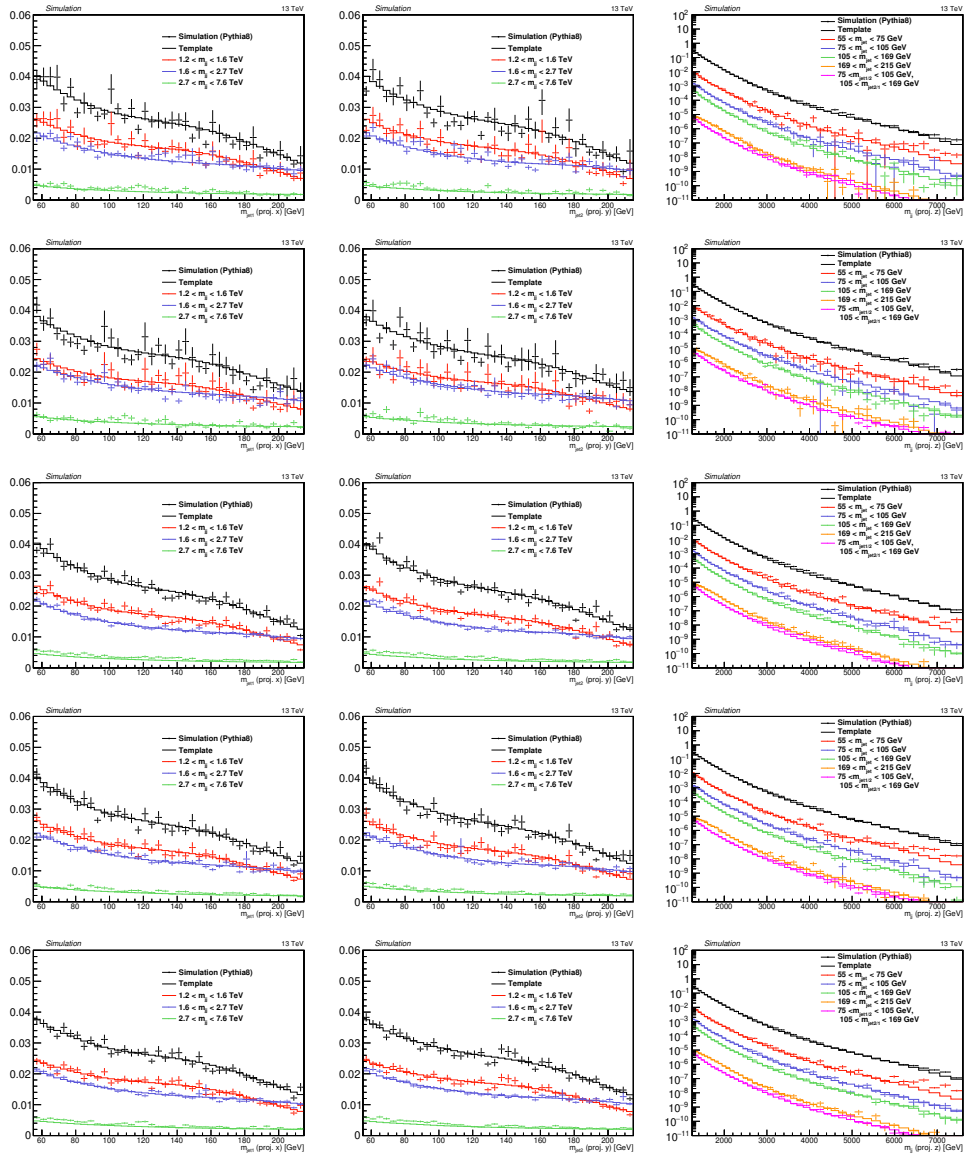


FIGURE B.8: Comparison between simulation (markers) and templates derived from generator level quantities (lines) for the nominal QCD PYTHIA8 sample. The templates are shown for m_{jet1} (left), m_{jet2} (middle) and m_{jj} (right) and, from top to bottom, for VH HPHP, VV HPHP, VH LPHP, VH HPLP and VV HPLP categories. In the m_{jj} distribution, the jet mass selections reported in the legend are applied to both jets.

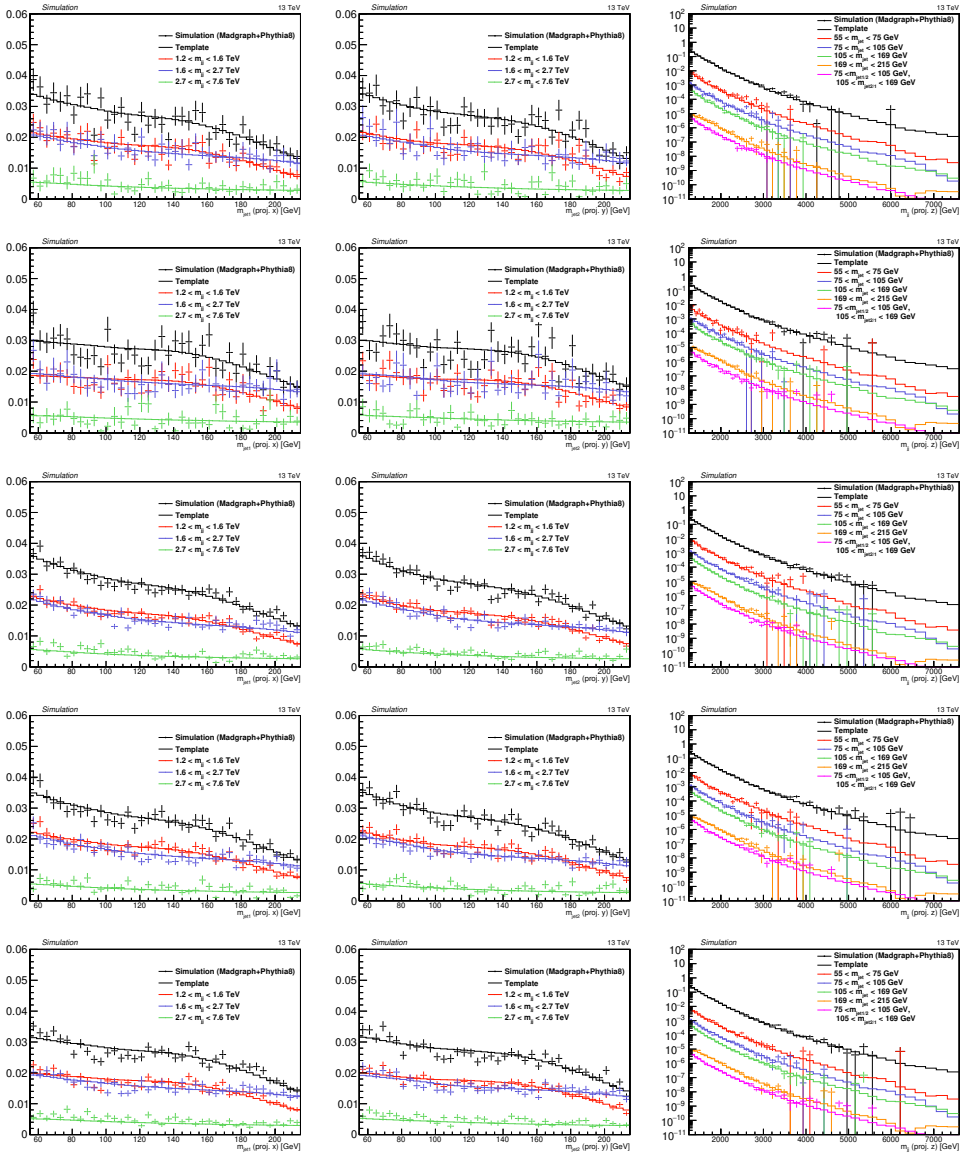


FIGURE B.9: Comparison between simulation (markers) and templates derived from generator level quantities (lines) for the nominal QCD MADGRAPH5_aMC@NLO. The templates are shown for m_{jet1} (left), m_{jet2} (middle) and m_{jj} (right) and, from top to bottom, for VH HPHP, VV HPHP, VH LPHP, VH HPLP and VV HPLP categories. In the m_{jj} distribution, the jet mass selections reported in the legend are applied to both jets.

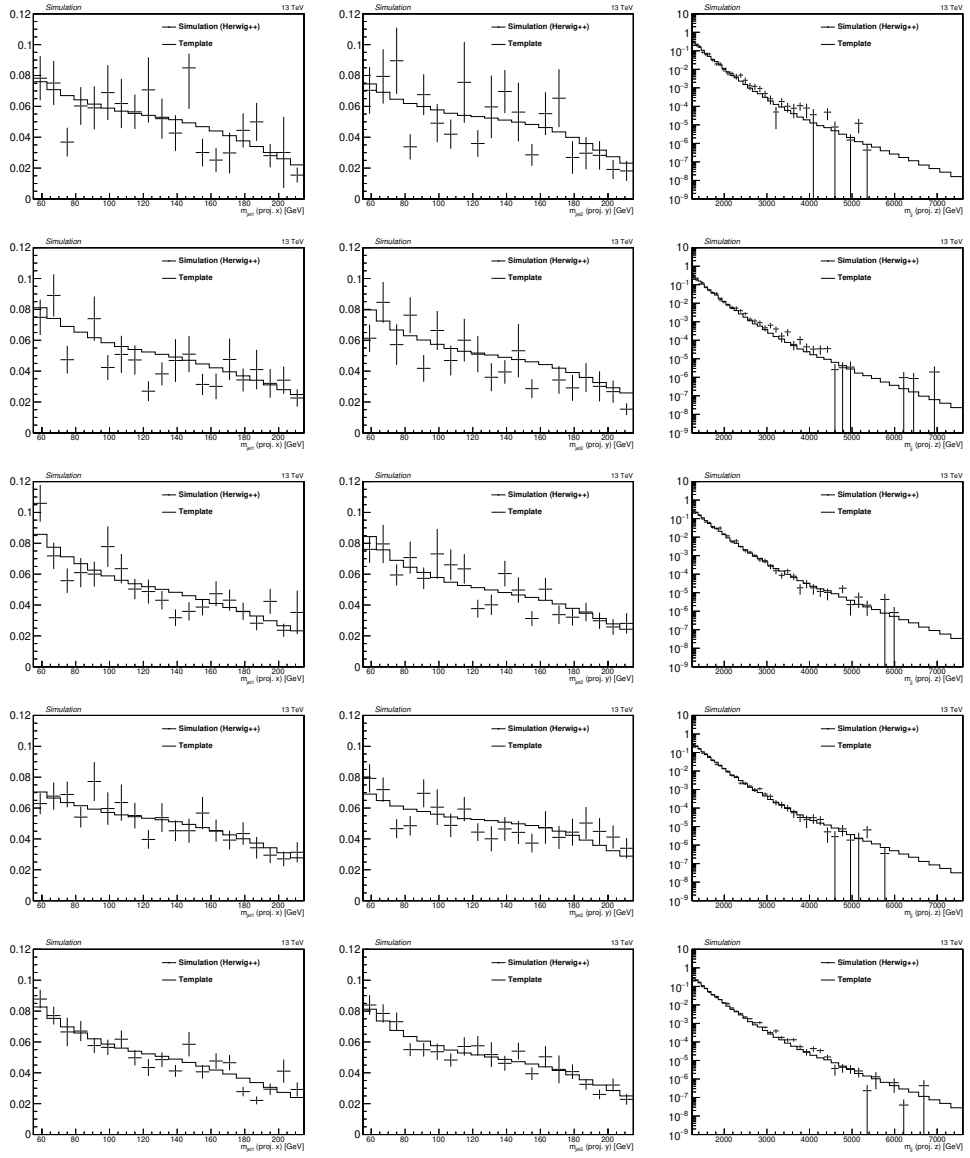


FIGURE B.10: Comparison between simulation (markers) and templates derived from generator level quantities (lines) for the nominal QCD HERWIG. The templates are shown for m_{jet1} (left), m_{jet2} (middle) and m_{jj} (right) and, from top to bottom, for VH HPHP, VV HPHP, VH LPHP, VH HPLP and VV HPLP categories.

B.2.2 VBF categories

Figures B.11 shows the comparison between the ggF/DY derived templates and MC simulation for each of the VBF categories for the nominal QCD PYTHIA8 samples. Only inclusive projections are shown as statistics for the VBF categories is rather limited with respect to the ggF/DY categories. Good agreement in within statistical uncertainties between MC and templates in all three dimensions and for all samples is observed.

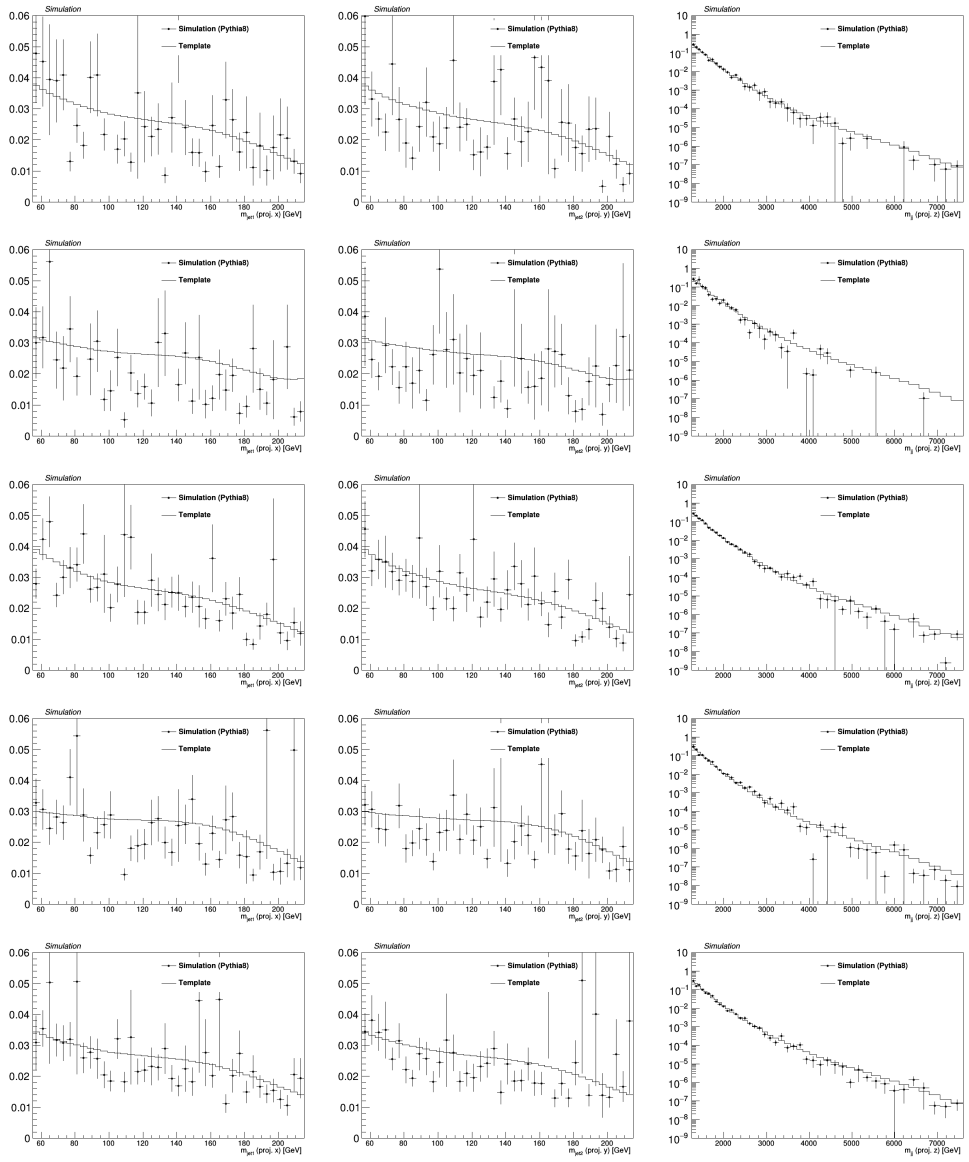


FIGURE B.11: Comparison between simulation (markers) and templates derived from generator level quantities (lines) for the nominal QCD PYTHIA8. The templates are shown for m_{jet1} (left), m_{jet2} (middle) and m_{jj} (right) and, from top to bottom, for VH HPHP, VV HPHP, VH LPHP, VH HPLP and VV HPLP categories.

B.3 $t\bar{t}$ modelling: additional distributions

The jet mass projection onto the whole dijet invariant mass range is shown in Fig. B.12, where the MC is fitted with a double Gaussian, one describing the top and one describing the W resonant components, and an Error Function Exponential (ErxExp) for the non-resonant part of the spectra.

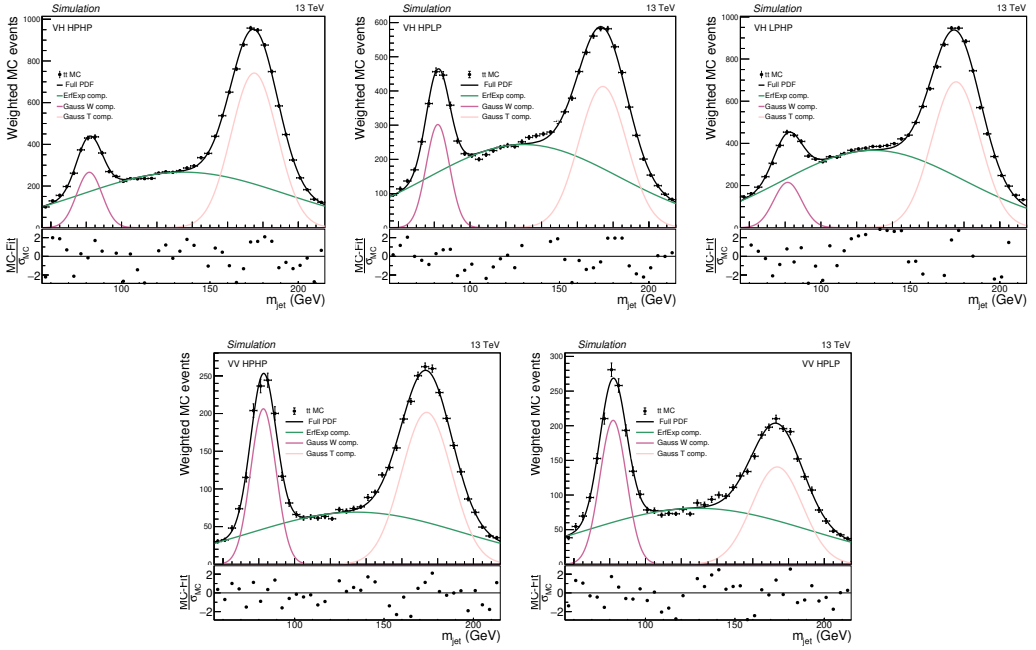


FIGURE B.12: Fit to the inclusive jet mass distributions (all bins of m_{jj}) for the $t\bar{t}$ background in the VH categories (top) and VV categories (bottom).

Then the jet mass is sampled in coarse bins of m_{jj}. The binning is chosen such as to have sufficient statistics in each slice, which, due to the SM $t\bar{t}$ background being concentrated at low-m_{jj}, is chosen to be fine at lower dijet invariant masses and coarser a higher m_{jj}. The binning is: 1246–1404 GeV, 1404–1563 GeV, 1563–1722 GeV and 1722–7600 GeV. The distribution in each bin is then fitted with the chosen PDF (2Gauss+ErfExp) using the starting parameters evaluated from the fit to the inclusive distribution. These fits are shown in Fig B.13. In the VV categories, the contribution on top and W are of equal size and in the VH categories, the top contribution dominates. In all categories, the fraction of merged tops to merged Ws increase.

Each fit parameter is then extraxted in each m_{jj}-bin, and the fit parameter distribution is parametrised as a function of dijet invariant mass. The m_{jj}-value for each bin is chosen to be the average m_{jj}-value in that bin. The last point, at $\approx$ 1700 has been duplicated three times at higher m_{jj} values to ensure that the parametrised value at higher-m_{jj} is constrained by the value in the highest fittable m_{jj} bin. The parameterisation of each fit parameter in each analysis category is shown in Figures B.14-B.18.

The fitted jet mass distributions for m_{jet1} and m_{jet2} are shown in Figure B.19-B.23.

The final 1D m_{jj} templates are shown in Fig. B.24 -B.28 compared to MC simulation, for the 6 $t\bar{t}$ contribution and ggF/DY categories.

Once the final pdf described in Eq. **??** is built, the full model is fit to the full $t\bar{t}$ simulation to obtain the prefit normalization values for each $t\bar{t}$ contribution that will

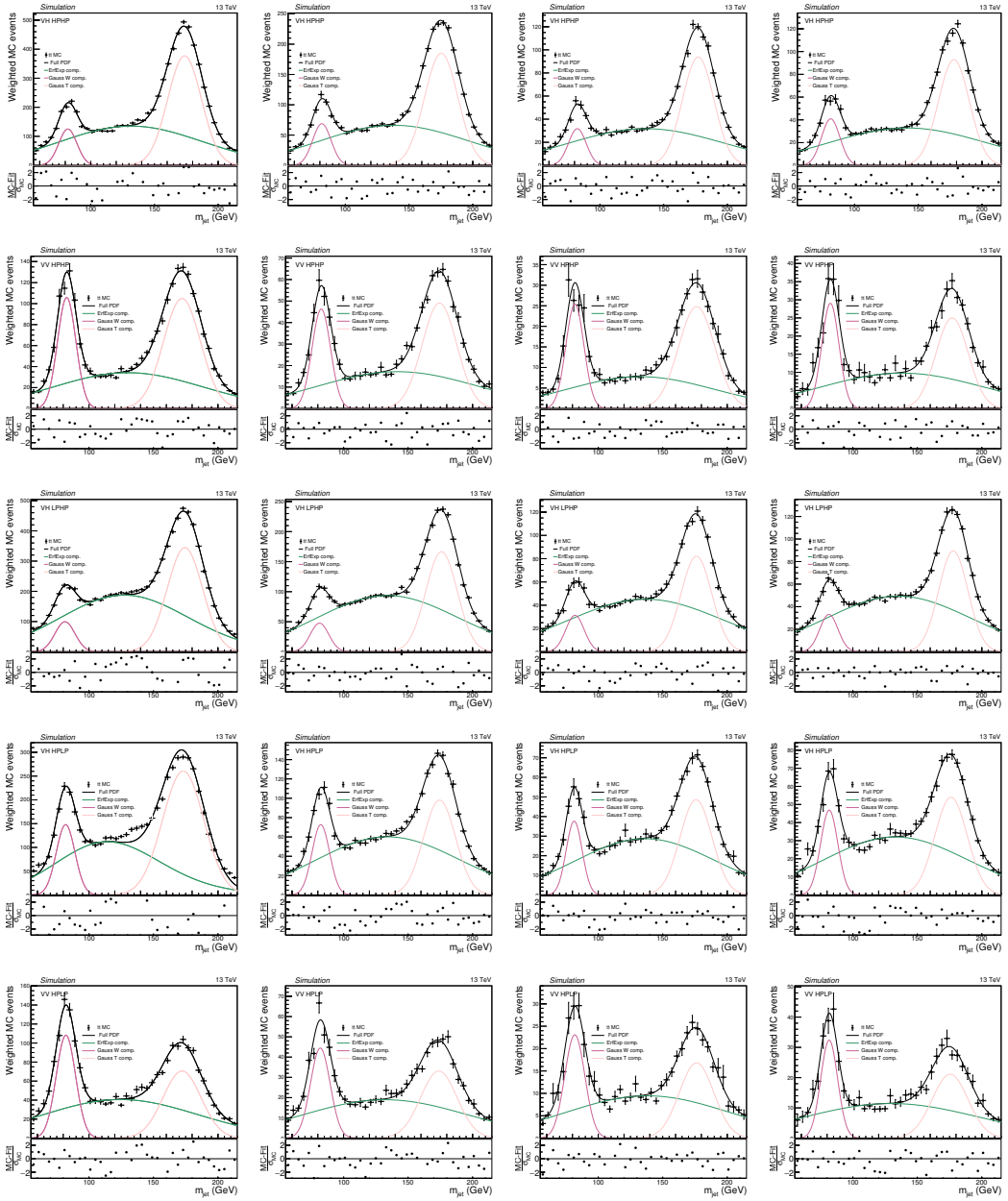


FIGURE B.13: The fitted jet mass distribution in increasing m_{jj} bins: 1246–1404 GeV (left), 1404–1563 GeV, 1563–1722 GeV and 1722–7600 GeV (right). From top to bottom: VH HPHP, VV HPHP, VH LPHP, VH HPLP and VV HPLP categories.

enter the final fit to data. This is done because the generator-level matching procedure above yields typically lower yields than what expected because of generator-level matching efficiency. The results of these fit are shown in Fig. B.29 and Fig. B.30 for all ggF/DY and VBF categories and are used as prefit to the final fit to data.

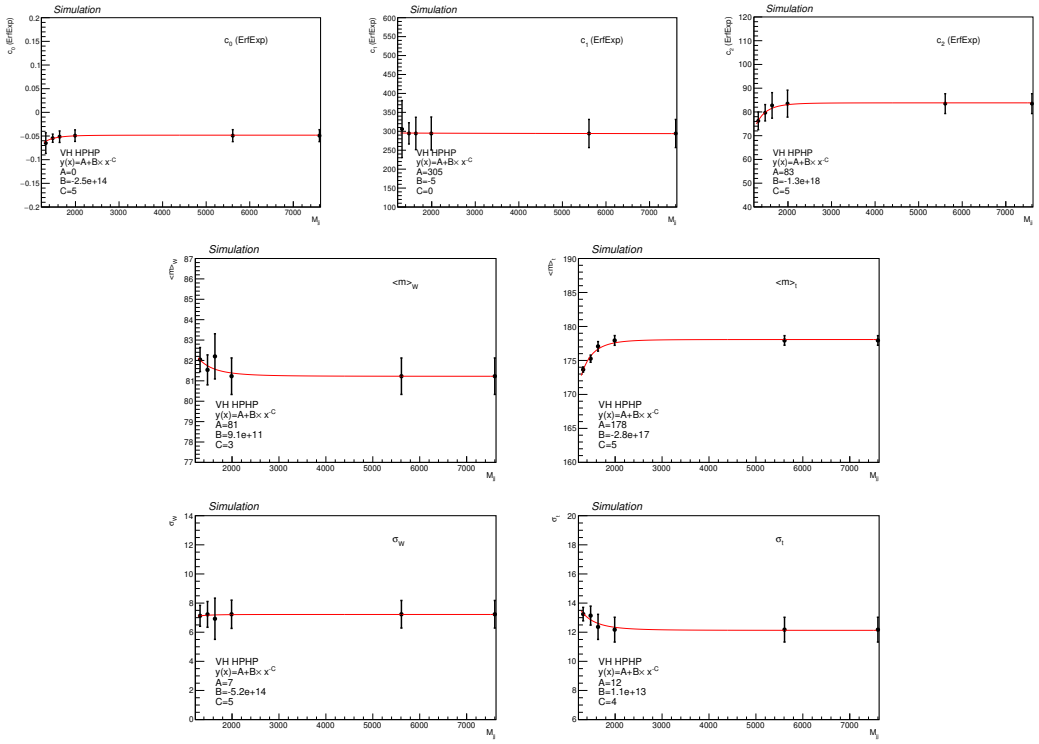


FIGURE B.14: The parametrisation of the fitted parameters for the $t\bar{t}$ background in the VH HPHP category. The first row shows the parameters of the non resonant component while the second and third row show the mean and resolution for the W boson (left) and top quark (right) peaks.

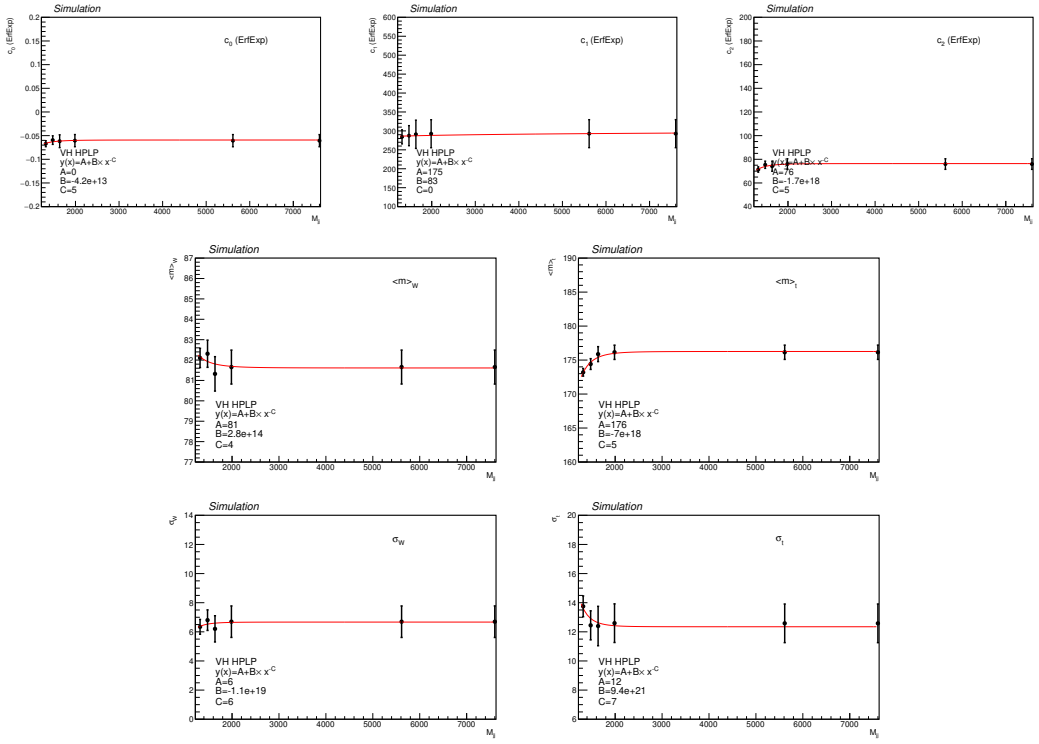


FIGURE B.15: The parametrisation of the fitted parameters for the $t\bar{t}$ background in the VH HPLP category. The first row shows the parameters of the non resonant component while the second and third row show the mean and resolution for the W boson (left) and top quark (right) peaks.

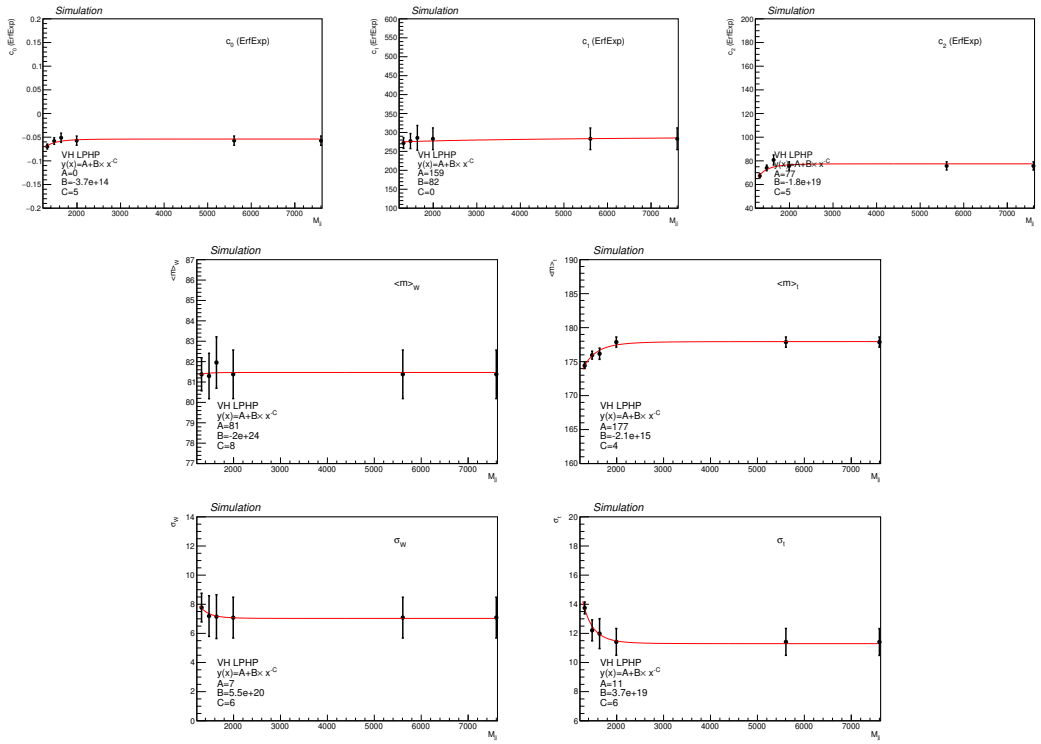


FIGURE B.16: The parametrisation of the fitted parameters for the $t\bar{t}$ background in the VH LPHP category. The first row shows the parameters of the non resonant component while the second and third row show the mean and resolution for the W boson (left) and top quark (right) peaks.

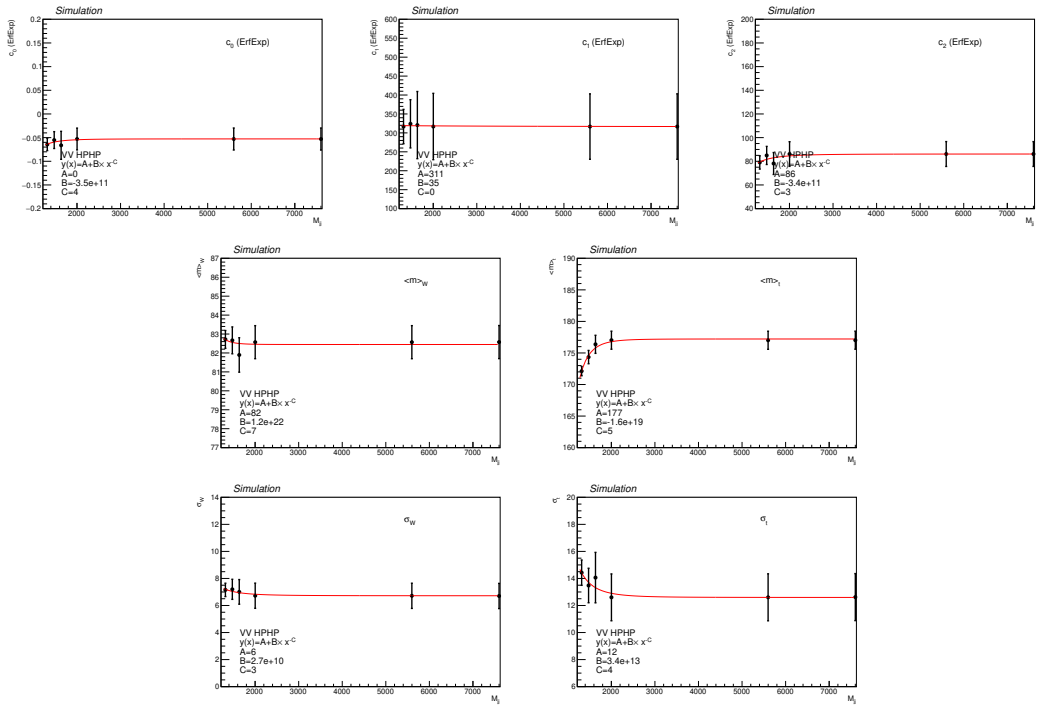


FIGURE B.17: The parametrisation of the fitted parameters for the $t\bar{t}$ background in the VV HPHP category. The first row shows the parameters of the non resonant component while the second and third row show the mean and resolution for the W boson (left) and top quark (right) peaks.

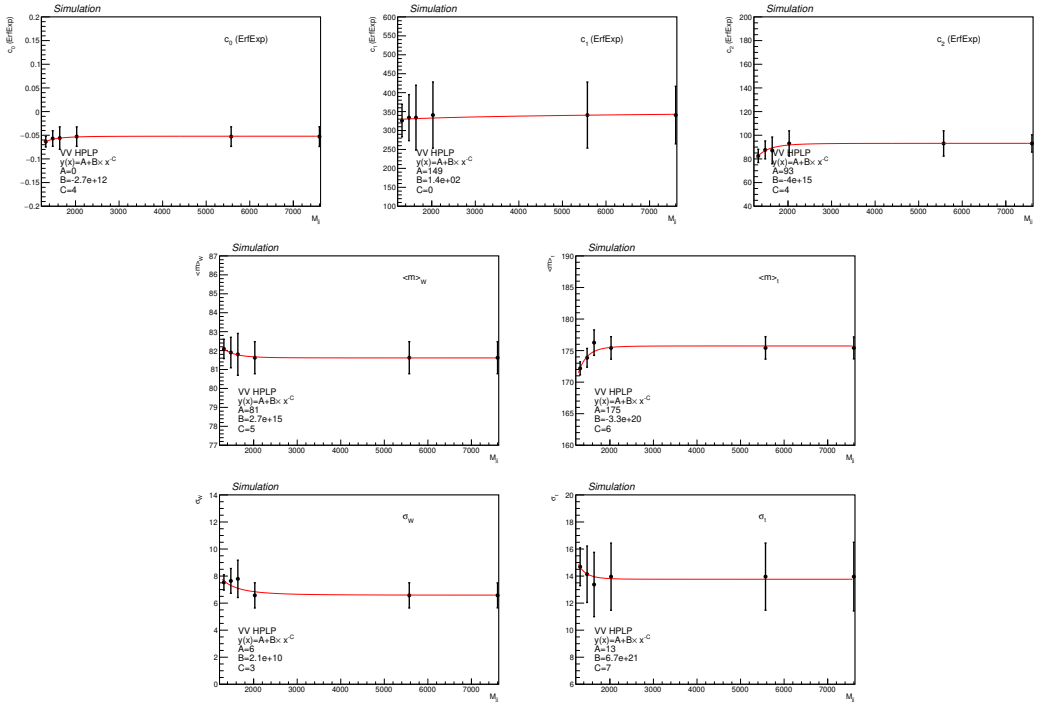


FIGURE B.18: The parametrisation of the fitted parameters for the $t\bar{t}$ background in the VV HPLP category. The first row shows the parameters of the non resonant component while the second and third row show the mean and resolution for the W boson (left) and top quark (right) peaks.

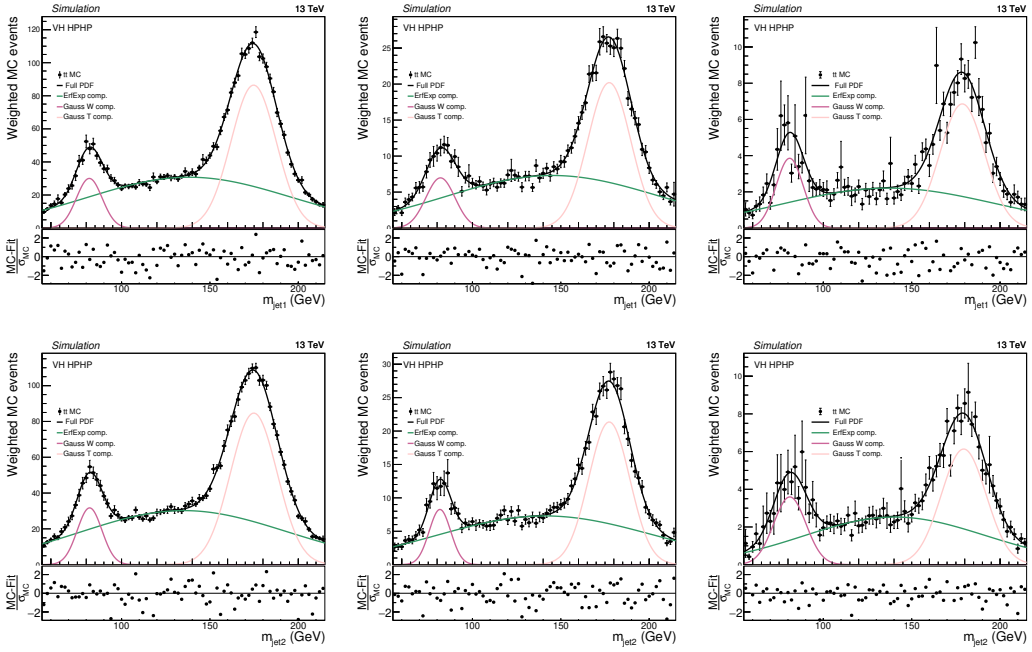


FIGURE B.19: Closure test for the $t\bar{t}$ background pdfs in the VH HPHP category. Here, m_{jet1} (top) and m_{jet2} (bottom) are fit separately starting from the parametrisations of the $t\bar{t}$ background as a function of m_{jj}. This is shown in increasing m_{jj} bins: 1325–1563 GeV (left), 1643–1881 GeV (middle), 2040–7600 GeV (right).

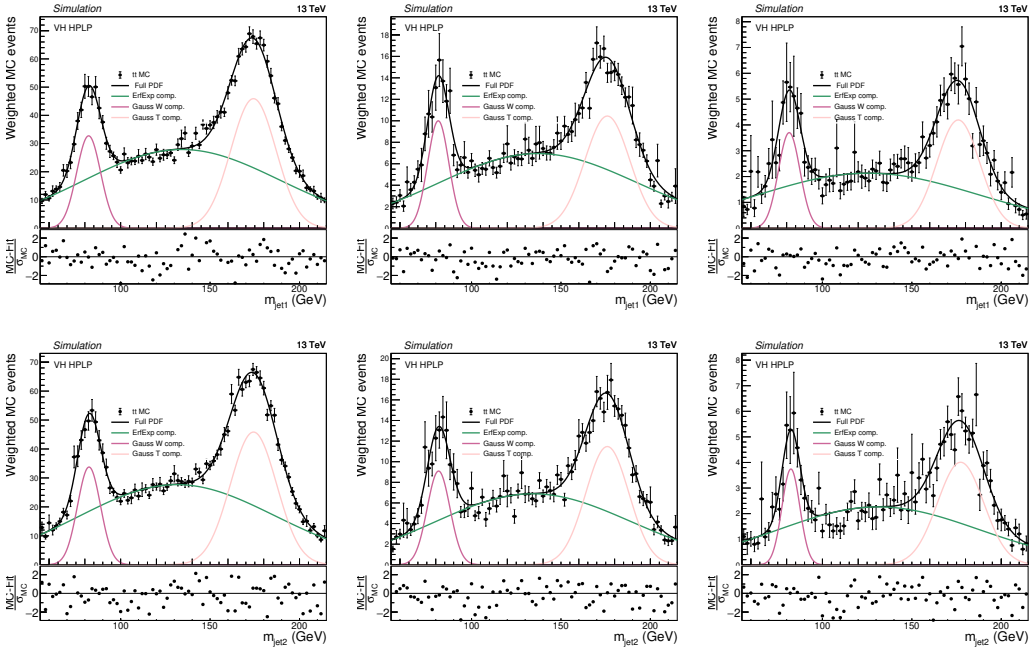


FIGURE B.20: Closure test for the $t\bar{t}$ background pdfs in the VH HPLP category. Here, m_{jet1} (top) and m_{jet2} (bottom) are fit separately starting from the parametrisations of the $t\bar{t}$ background as a function of m_{jj}. This is shown in increasing m_{jj} bins: 1325–1563 GeV (left), 1643–1881 GeV (middle), 2040–7600 GeV (right).

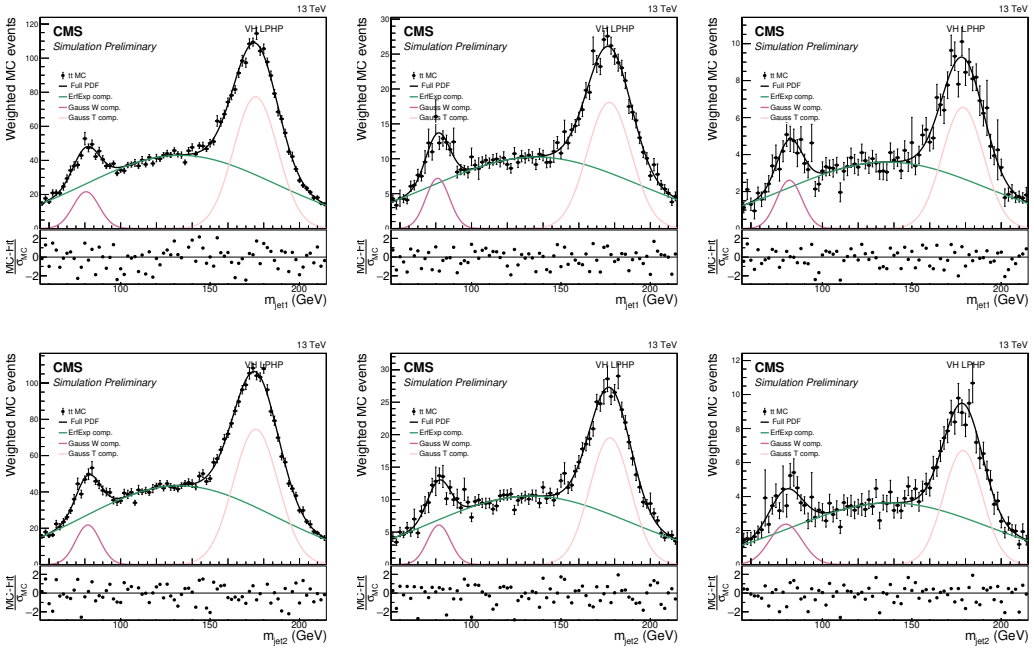


FIGURE B.21: Closure test for the $t\bar{t}$ background pdfs in the VH LPHP category. Here, m_{jet1} (top) and m_{jet2} (bottom) are fit separately starting from the parametrisations of the $t\bar{t}$ background as a function of m_{jj}. This is shown in increasing m_{jj} bins: 1325–1563 GeV (left), 1643–1881 GeV (middle), 2040–7600 GeV (right).

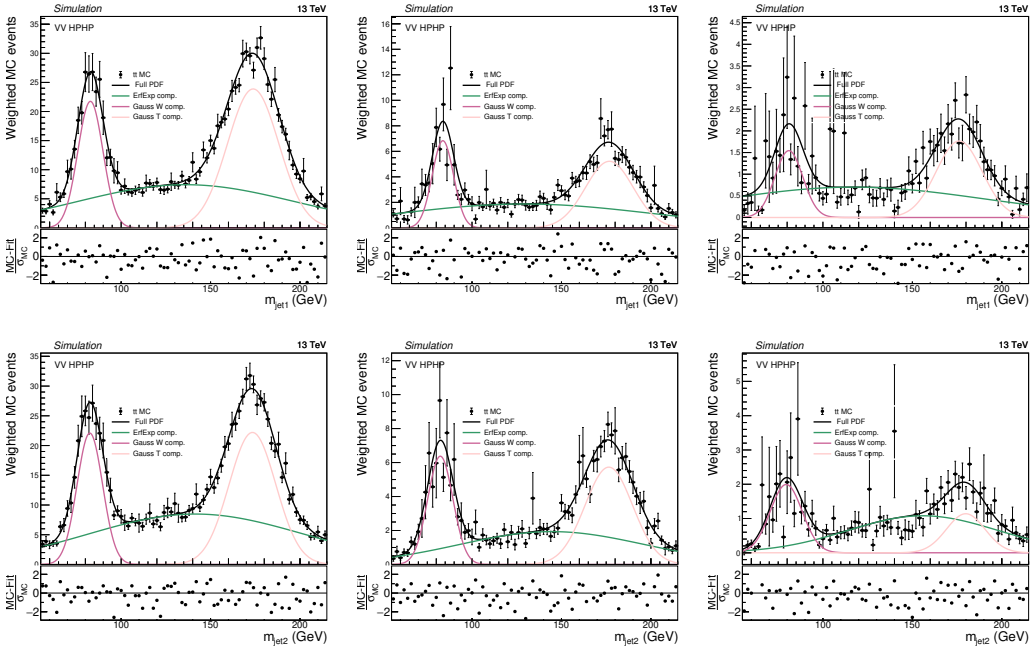


FIGURE B.22: Closure test for the $t\bar{t}$ background pdfs in the VV HPHP category. Here, m_{jet1} (top) and m_{jet2} (bottom) are fit separately starting from the parametrisations of the $t\bar{t}$ background as a function of m_{jj}. This is shown in increasing m_{jj} bins: 1325–1563 GeV (left), 1643–1881 GeV (middle), 2040–7600 GeV (right).

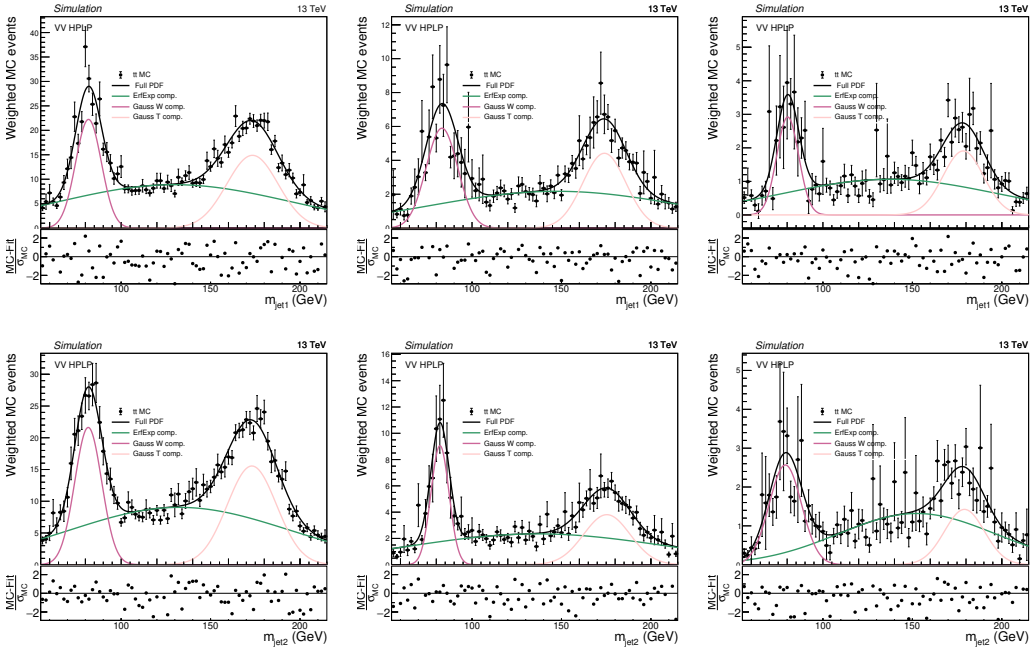


FIGURE B.23: Closure test for the $t\bar{t}$ background pdfs in the VV HPLP category. Here, m_{jet1} (top) and m_{jet2} (bottom) are fit separately starting from the parametrisations of the $t\bar{t}$ background as a function of m_{jj}. This is shown in increasing m_{jj} bins: 1325–1563 GeV (left), 1643–1881 GeV (middle), 2040–7600 GeV (right).

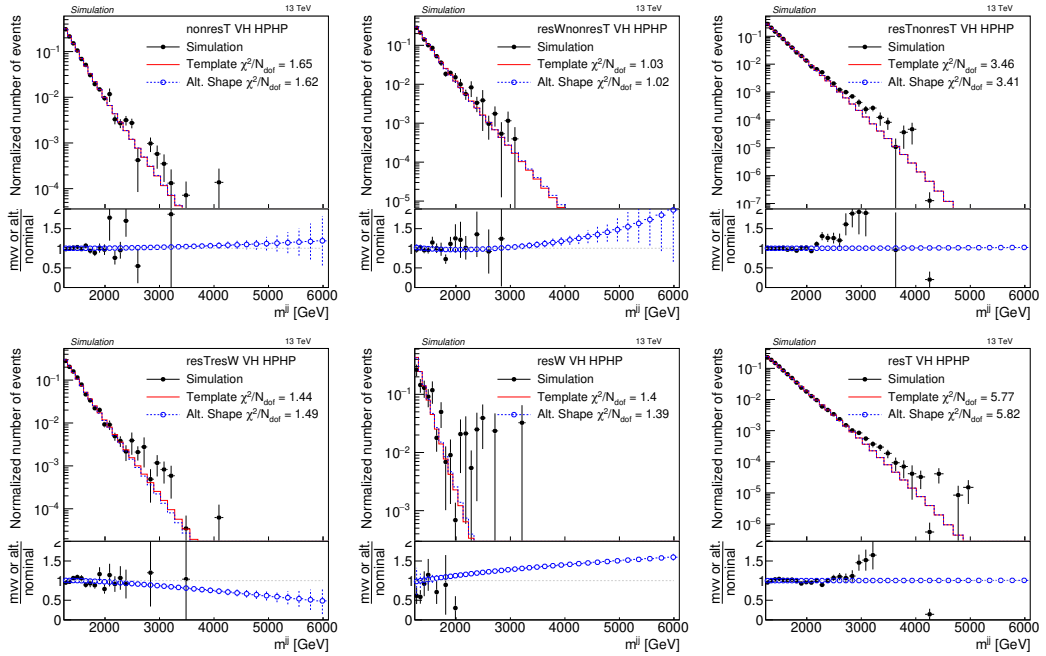


FIGURE B.24: Final one-dimensional m_{jj} templates in the VH HPHP category. The nominal shape derived from the fitting procedure can be seen as red solid line, the alternative shape derived from varying the slope of the m_{jj} spectrum is shown as the blue dashed line. Top, from left to right: nonRes, WnonRes and TnonRes; bottom from left to right: WT, WW, TT.

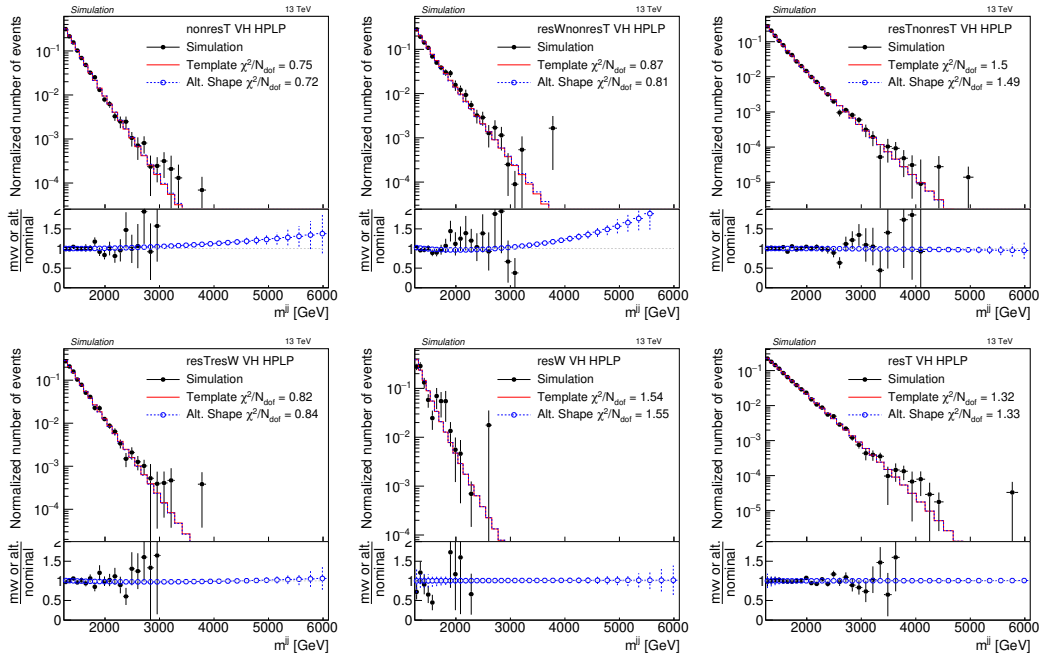


FIGURE B.25: Final one-dimensional m_{jj} templates in the VH HPLP category. The nominal shape derived from the fitting procedure can be seen as red solid line, the alternative shape derived from varying the slope of the m_{jj} spectrum is shown as the blue dashed line. Top, from left to right: nonRes, WnonRes and TnonRes; bottom from left to right: WT, WW, TT.

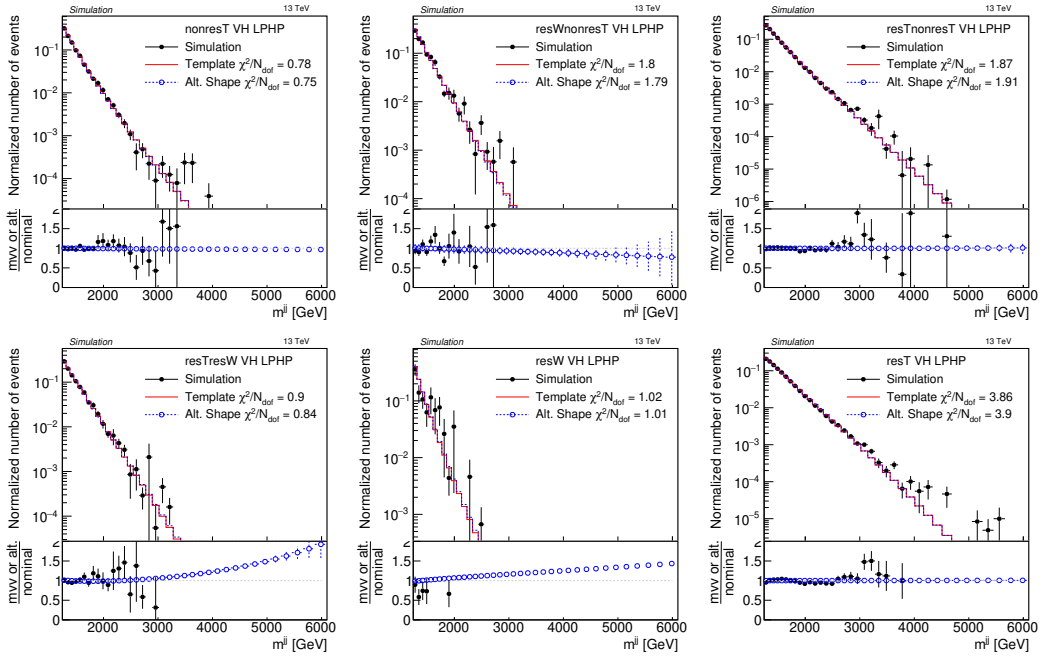


FIGURE B.26: Final one-dimensional m_{jj} templates in the VH LPHP category. The nominal shape derived from the fitting procedure can be seen as red solid line, the alternative shape derived from varying the slope of the m_{jj} spectrum is shown as the blue dashed line. Top, from left to right: nonRes, WnonRes and TnonRes; bottom from left to right: WT, WW, TT.

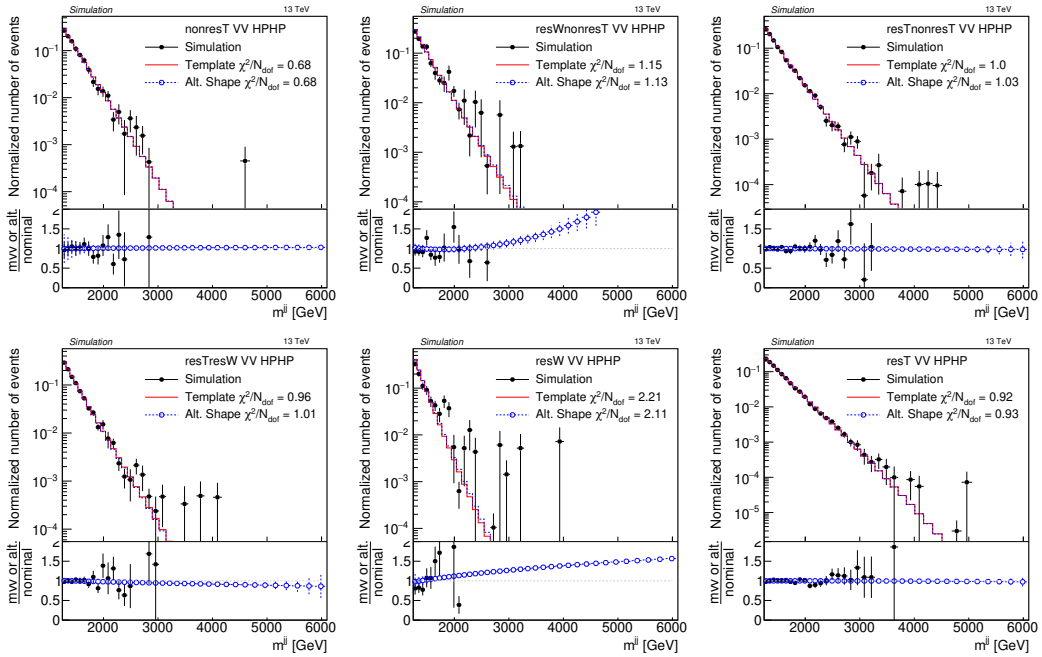


FIGURE B.27: Final one-dimensional m_{jj} templates in the VV HPHP category. The nominal shape derived from the fitting procedure can be seen as red solid line, the alternative shape derived from varying the slope of the m_{jj} spectrum is shown as the blue dashed line. Top, from left to right: nonRes, WnonRes and TnonRes; bottom from left to right: WT, WW, TT.

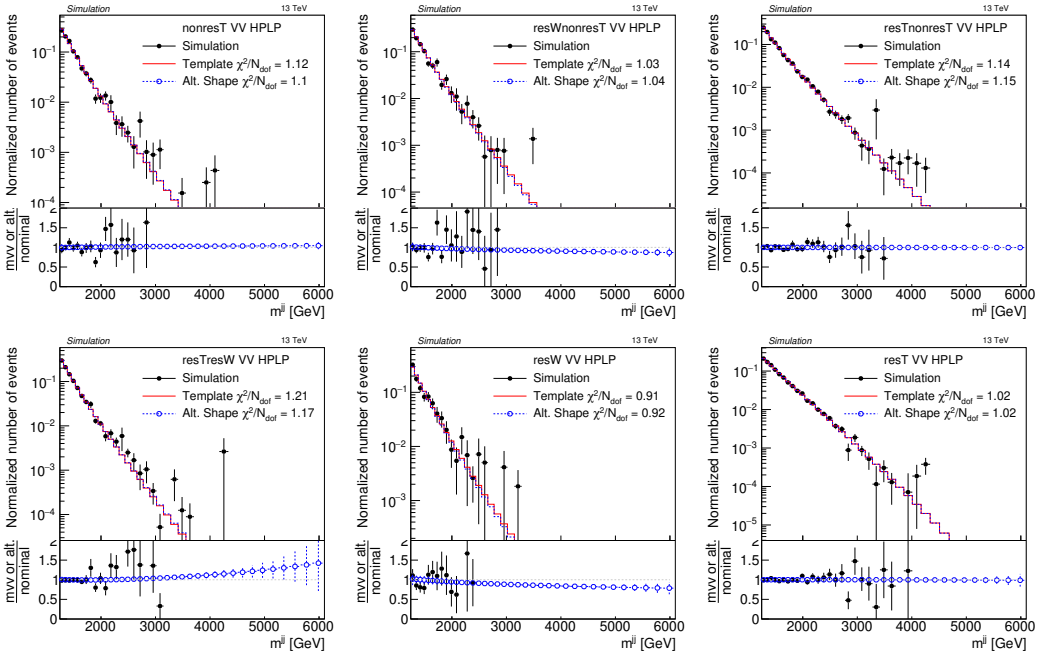


FIGURE B.28: Final one-dimensional m_{jj} templates in the VV HPLP category. The nominal shape derived from the fitting procedure can be seen as red solid line, the alternative shape derived from varying the slope of the m_{jj} spectrum is shown as the blue dashed line. Top, from left to right: nonRes, WnonRes and TnonRes; bottom from left to right: WT, WW, TT.

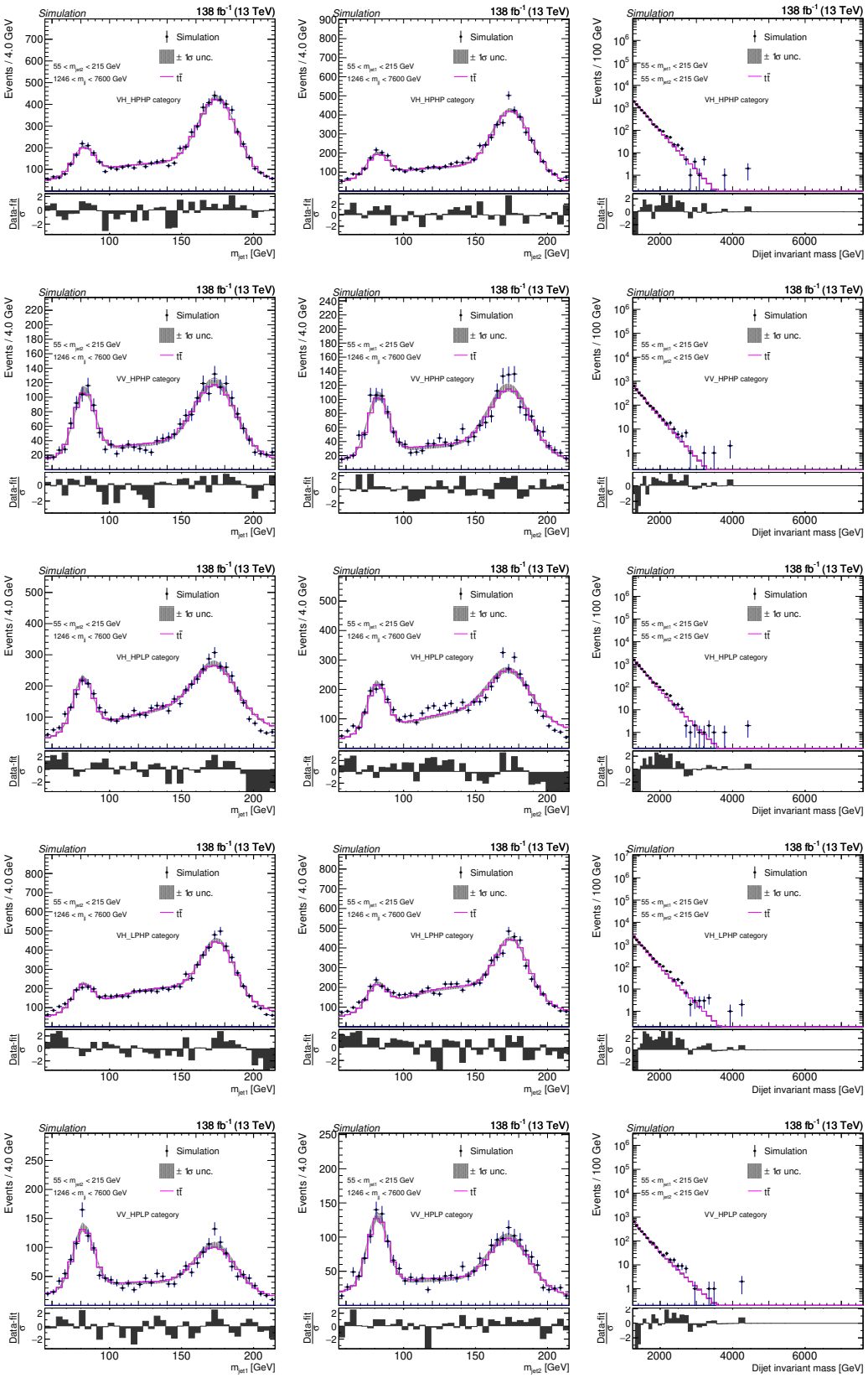


FIGURE B.29: Post fit to the $t\bar{t}$ MC in, from top to bottom, the VH HPHP, VV HPHP, VH HPLP, VH LPHP and VV HPLP categories.

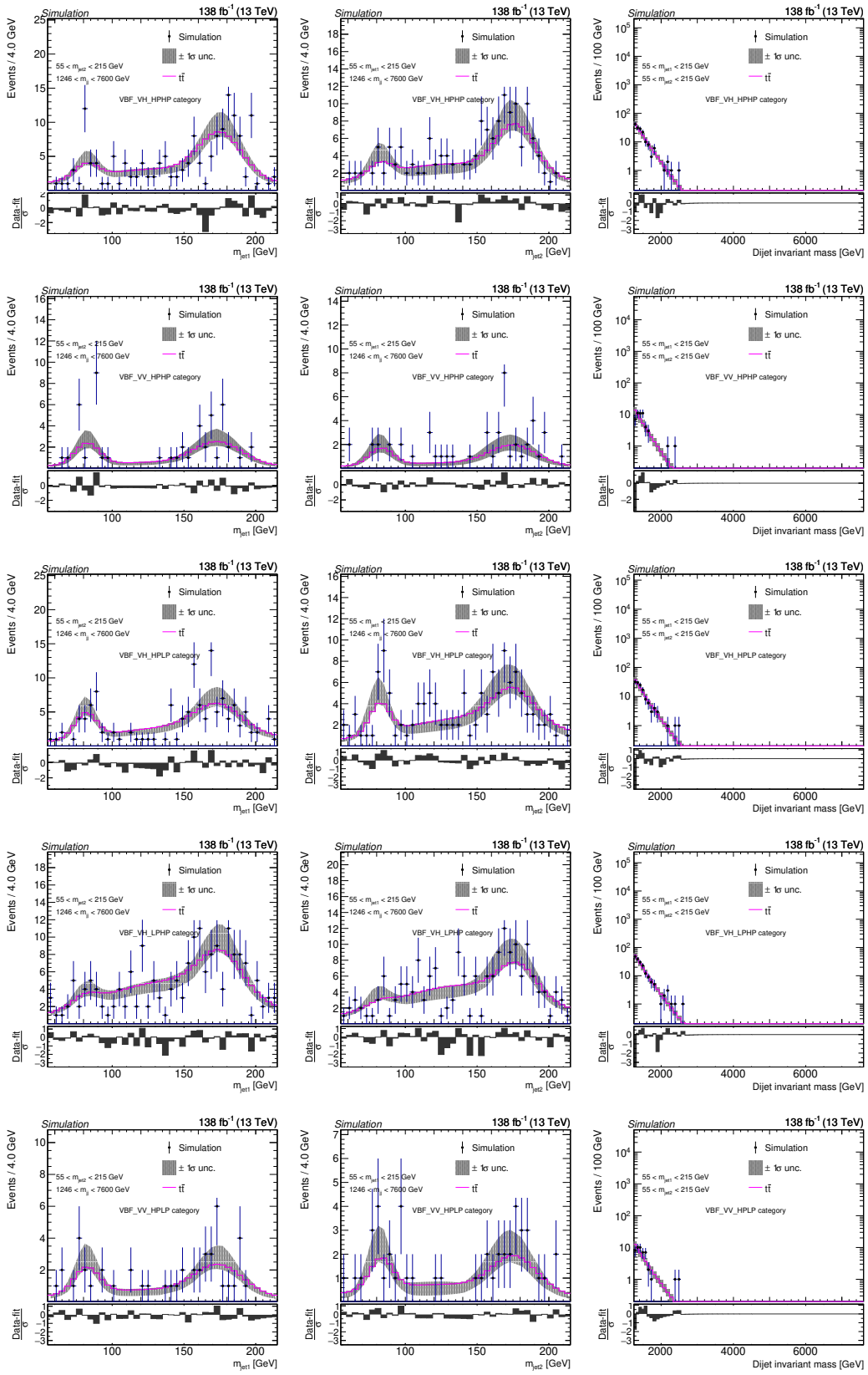


FIGURE B.30: Post fit to the $t\bar{t}$ MC in, from top to bottom, the VBF VH HPHP, VBF VV HPHP, VBF VH HPLP, VBF VH LPHP and VBF VV HPLP categories.

B.4 V+jets modelling: additional distributions

In Fig. B.31 the final fit of the dCB function to the resonant part of the W+jets and Z+jets spectrum is shown for all the categories.

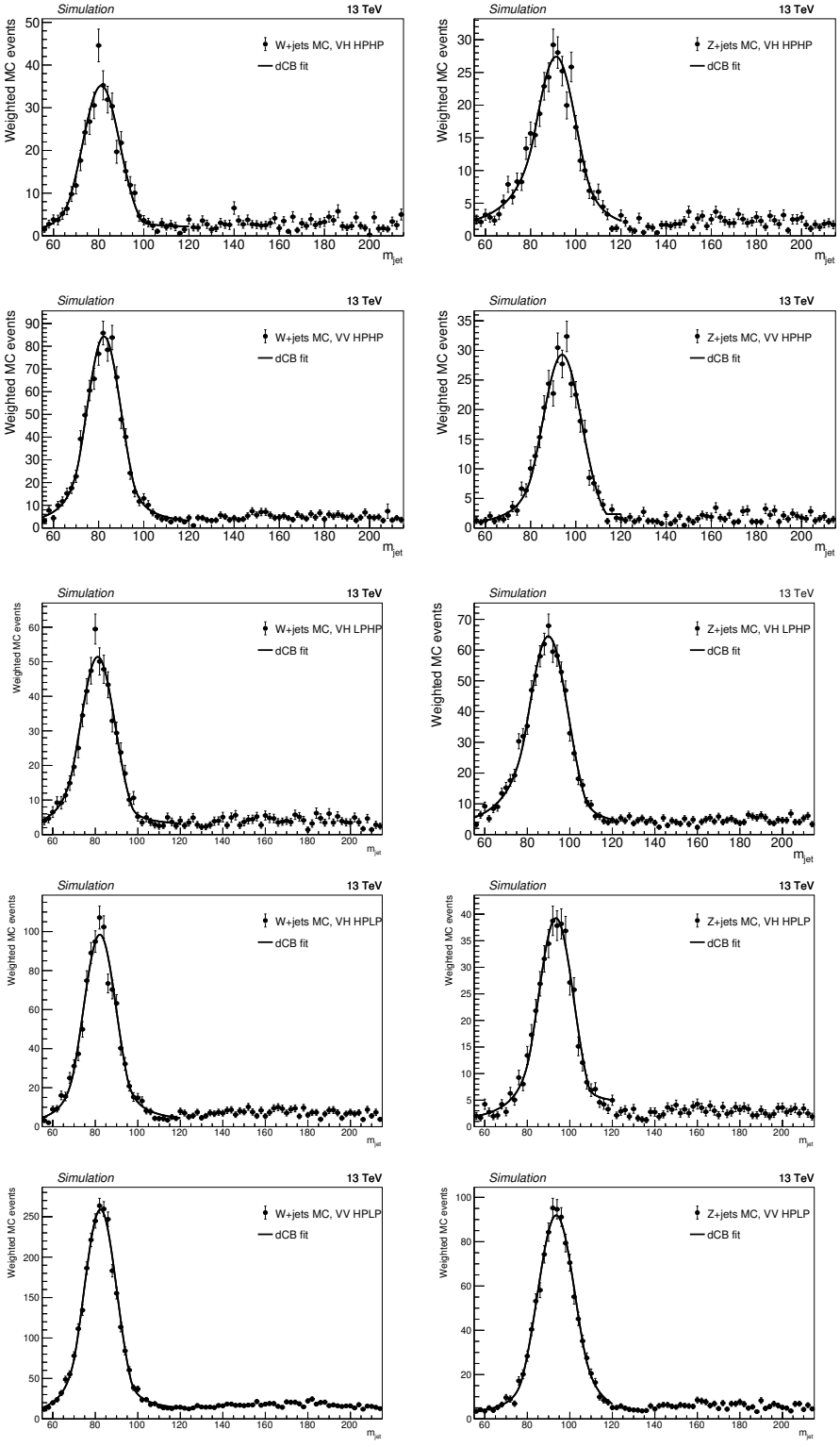


FIGURE B.31: Fit to the resonant part of the m_{SD} spectrum (black markers) with a dCB function (black line) in each of the category, for W+jets (left) and Z+jets (right).

The non-resonant part of the V+jets backgrounds is modelled using a simple fit with a Gaussian to the non-resonant part of the spectrum as shown in Fig. B.32 for W+jets and Z+jets.

The 1D m_{jj} templates are shown in Fig. B.33 compared to MC simulation.

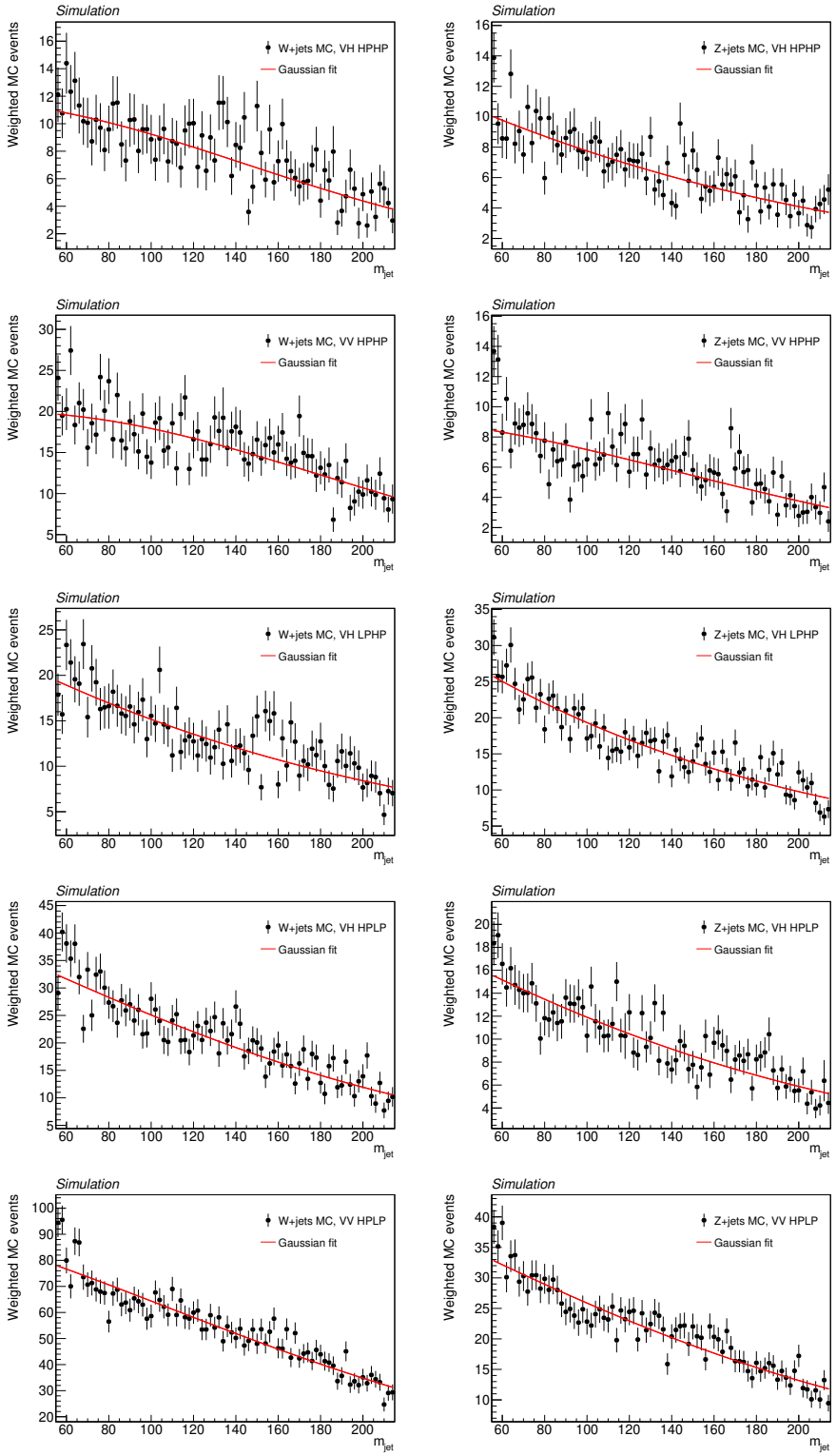


FIGURE B.32: Fit to the non-resonant part of the m_{SD} spectrum (black markers) with a dCB function (red line) in each of the category, for W+jets (left) and Z+jets (right).

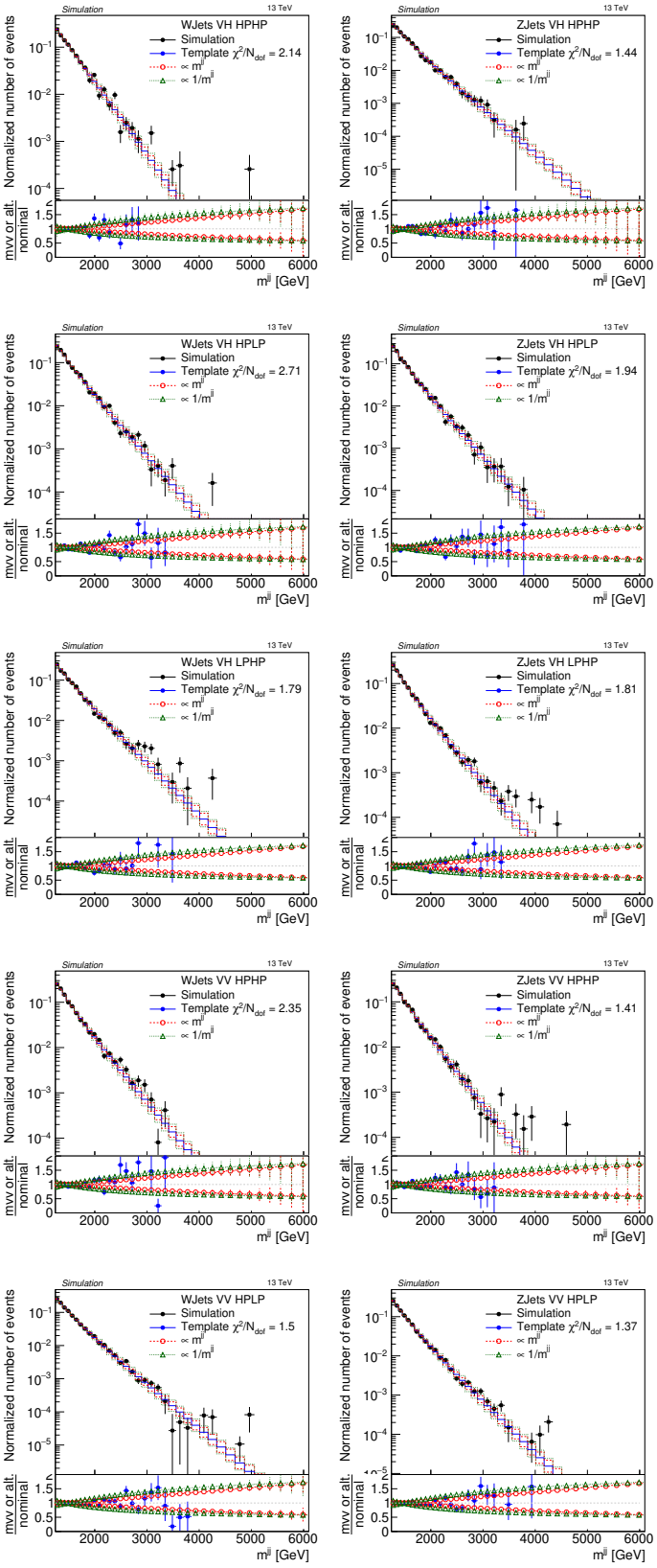


FIGURE B.33: One-dimensional m_{jj} templates for the resonant W+jets (left) and Z+jets (right) backgrounds compared to MC simulation for all categories. The nominal shape derived from the fitting procedure can be seen as blue line, alternative shapes derived from varying the slope of the m_{jj} spectrum are shown in green and red.

C | Systematic uncertainties and fit validation tests

C.1 Systematic uncertainties

C.1.1 Signal PDFs acceptance uncertainties

The relative variations for one signal benchmark are shown in Fig. C.1.

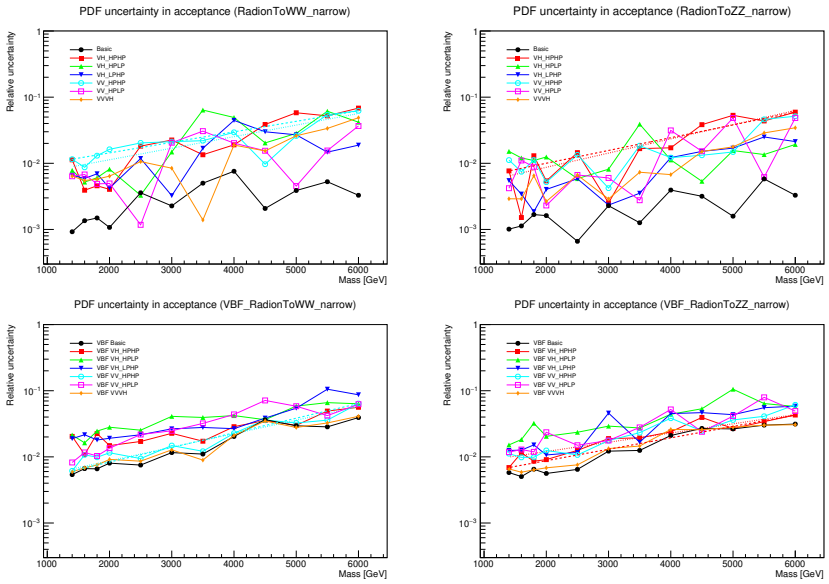


FIGURE C.1: Relative variations in signal efficiency due to PDFs as a function of the resonance mass and for all tagging categories for a Radion→ WW (left) and Radion→ ZZ (right) produced through gluon fusion (top) and vector boson fusion (bottom) mechanisms. In the top (bottom) plots only ggF/DY (VBF) selection categories are shown. The dashed line corresponds to the M_X-dependent uncertainty inserted in the final fit where the variations for the lowest and highest mass hypothesis in the most sensitive category are fitted versus M_X as explained in the text. As comparison, the dotted line corresponds to an alternative definition where a fit to the average of all variations per mass is performed. The first approach is chosen as found to be the most conservative.

C.1.2 QCD multijet alternative shapes

Those five shape uncertainties are assigned very large pre-fit values (allowed to float within 2 standard deviations), effectively allowing the simulation to take any value to fit the data. The alternate shapes described above are shown in Fig. C.2.

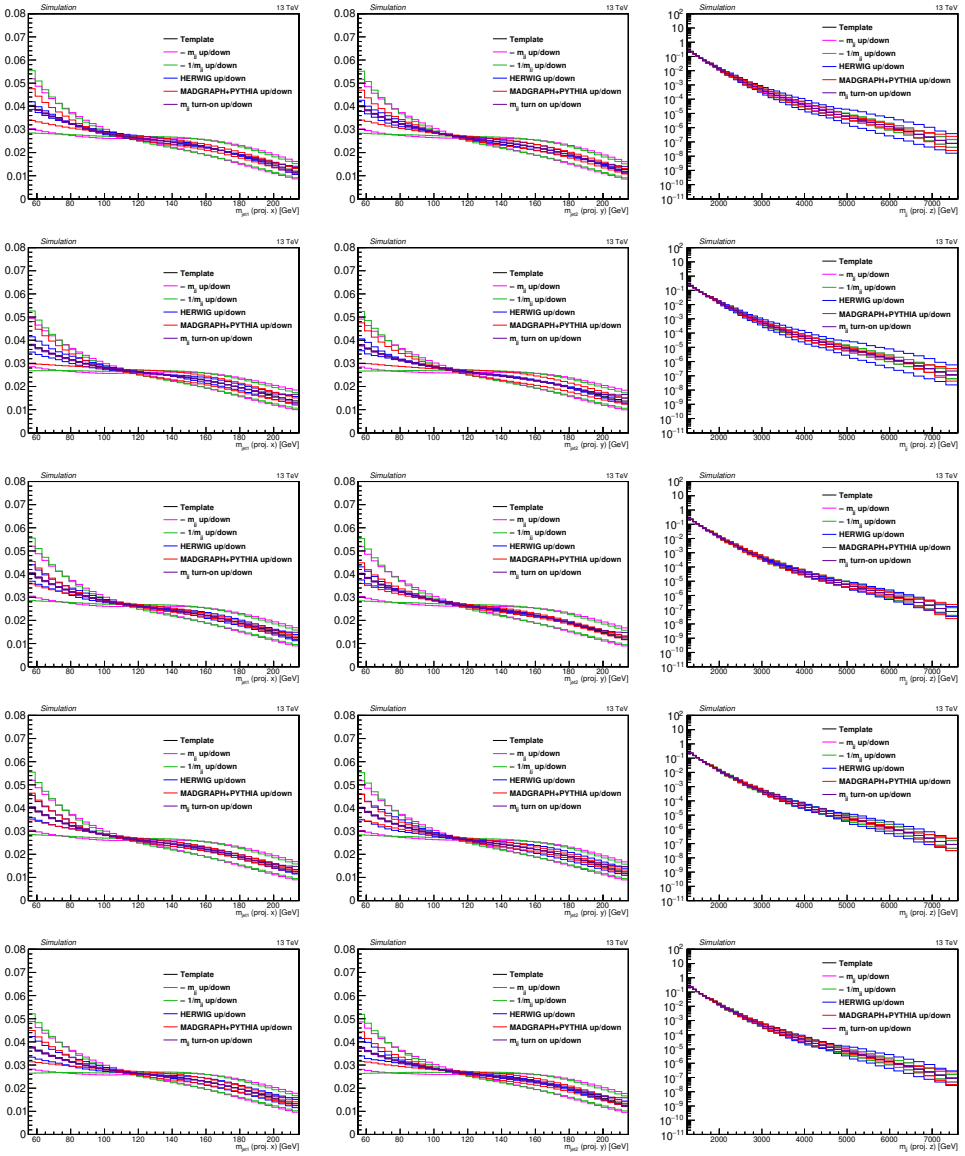


FIGURE C.2: The QCD PYTHIA8 MC data and nominal template is shown together with the five alternate shapes added to the fit as shape nuisance parameters for, from top to bottom, the VH HPHP, VV HPHP, VH LPHP, VH HPLP and VV HPLP categories.

C.2 Fit validation tests

Figure C.3 (C.4) shows the post-fit distributions in all ggF/DY (VBF) categories, projected onto the m_{jet1}, m_{jet2}, and m_{jj} dimensions. A good agreement is observed in all the categories.

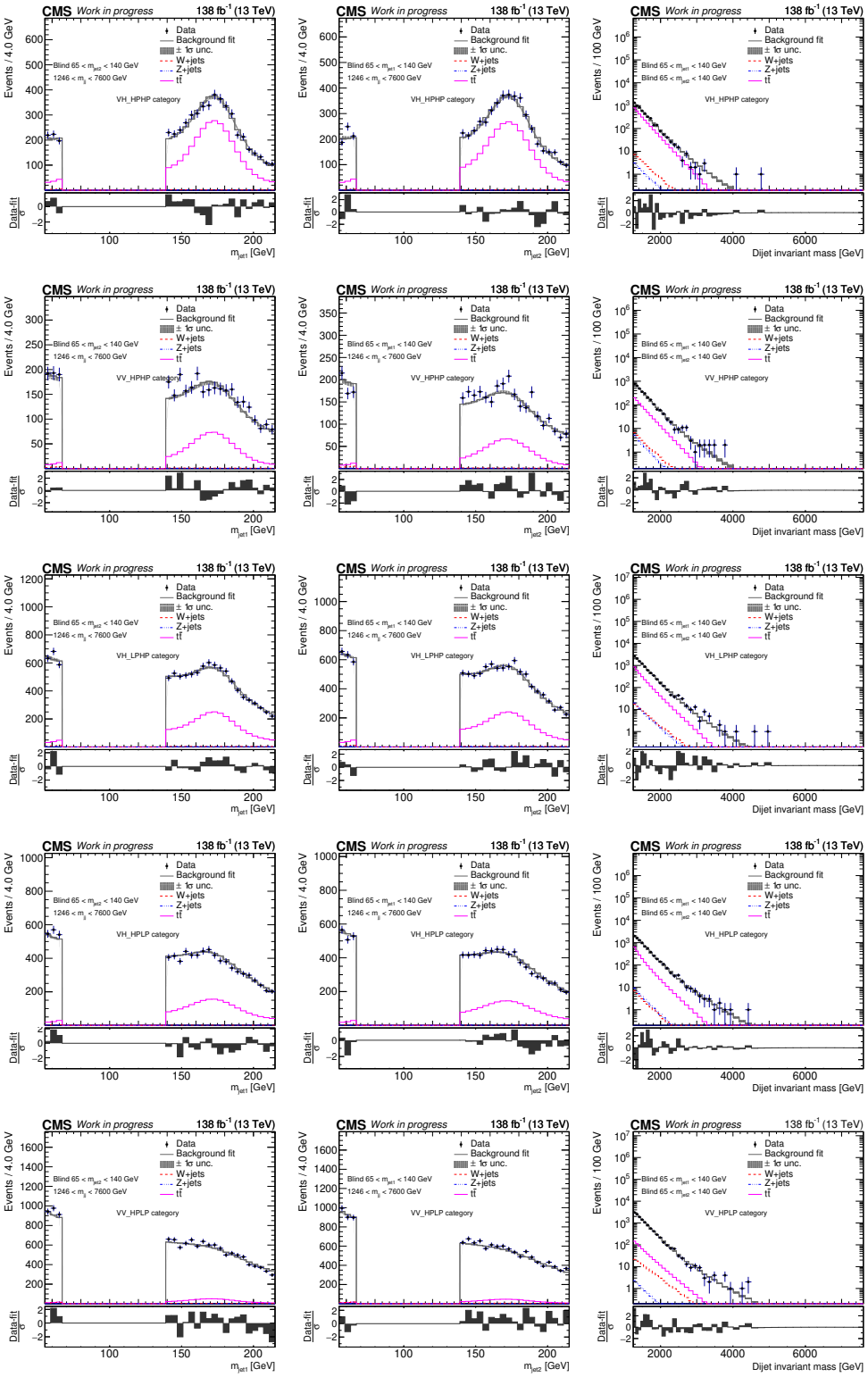


FIGURE C.3: Projections of data and post-fit distributions onto the m_{jet1} (left), m_{jet2} (middle), and m_{jj} (right) dimensions for, from top to bottom, the VH HPHP, VV HPHP, VH LPHP, VH HPLP and VV HPLP categories.

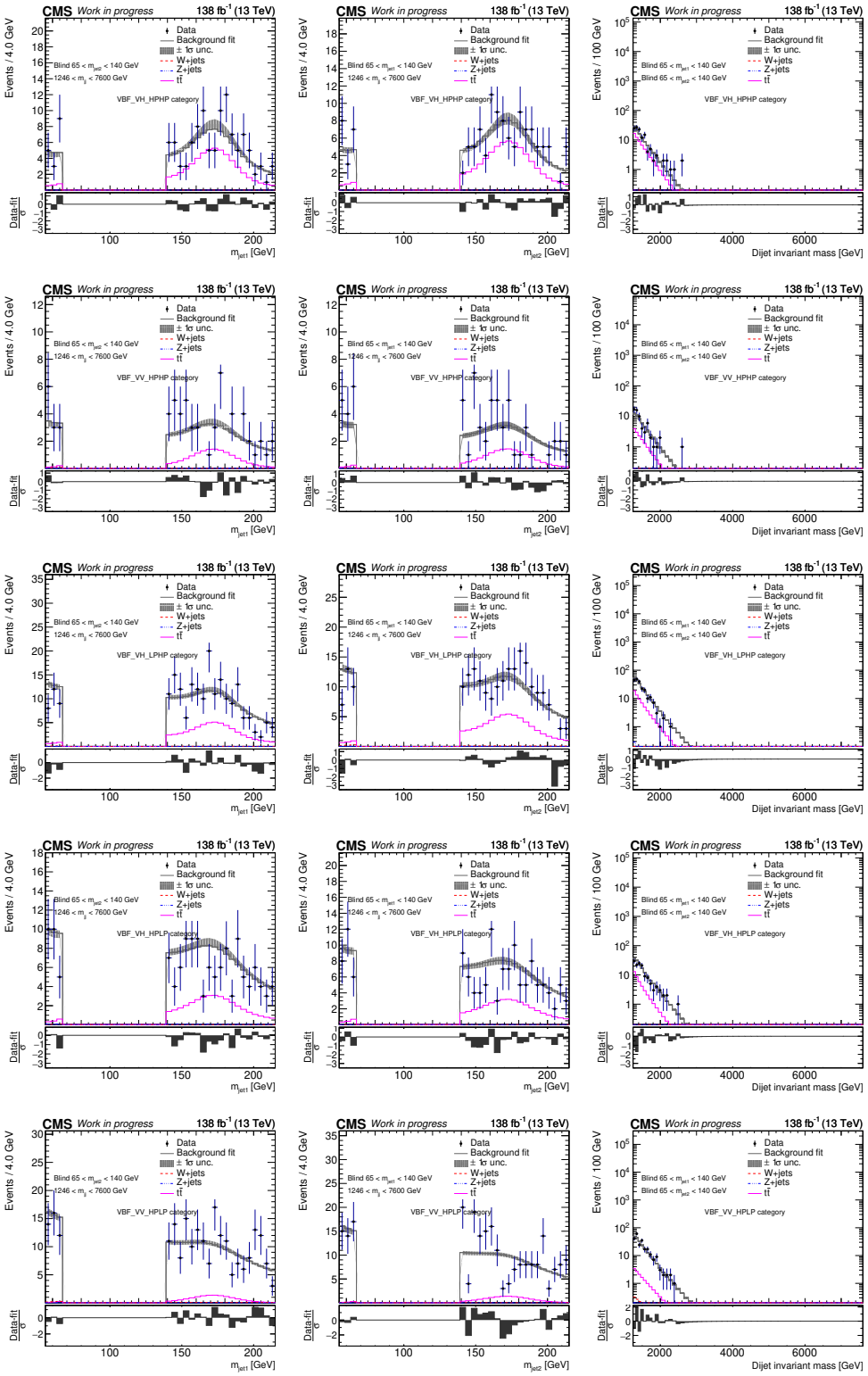


FIGURE C.4: Projections of data and post-fit distributions onto the m_{jet1} (left), m_{jet2} (middle), and m_{jj} (right) dimensions for, from top to bottom, the VBF VH HPHP, VBF VV HPHP, VBF VH LPHP, VBF VH HPLP and VBF VV HPLP categories.

D | Additional results

D.1 Observed significance

Figures D.1–D.6 show the observed significance for all considered signal hypotheses.

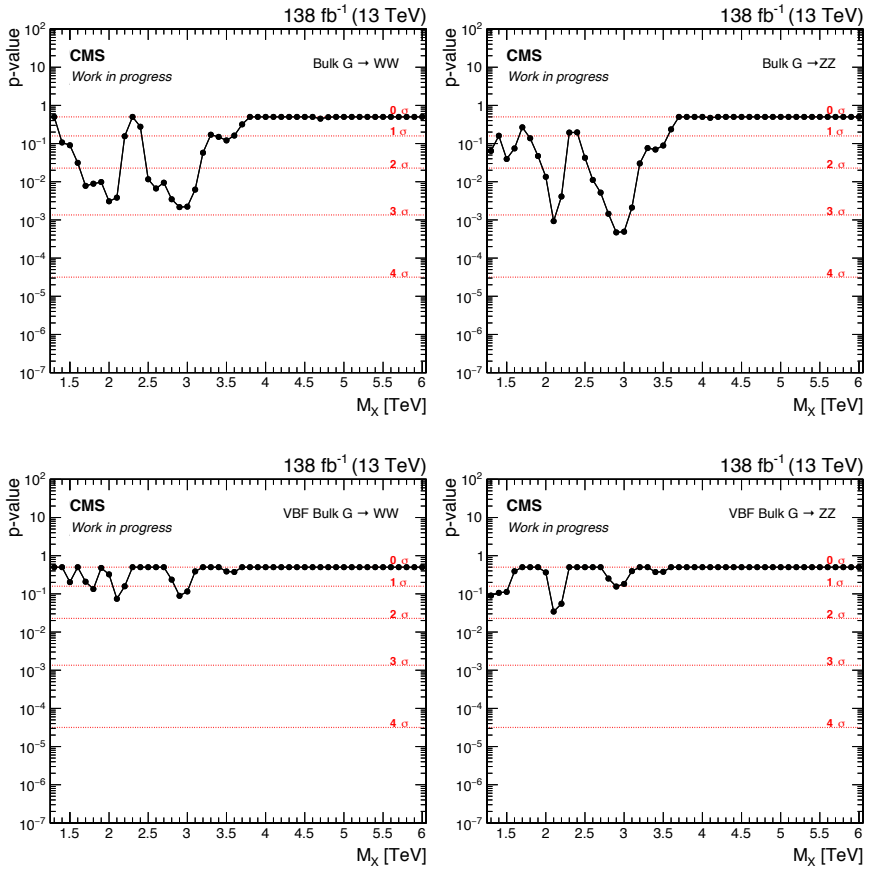


FIGURE D.1: Observed significance using 138fb^{-1} of data for a $G_{bulk} \rightarrow WW$ (left) and a $G_{bulk} \rightarrow ZZ$ (right) signals with the combination of all the analysis categories. On the top the ggF production mode is shown while on the bottom the VBF one is presented.

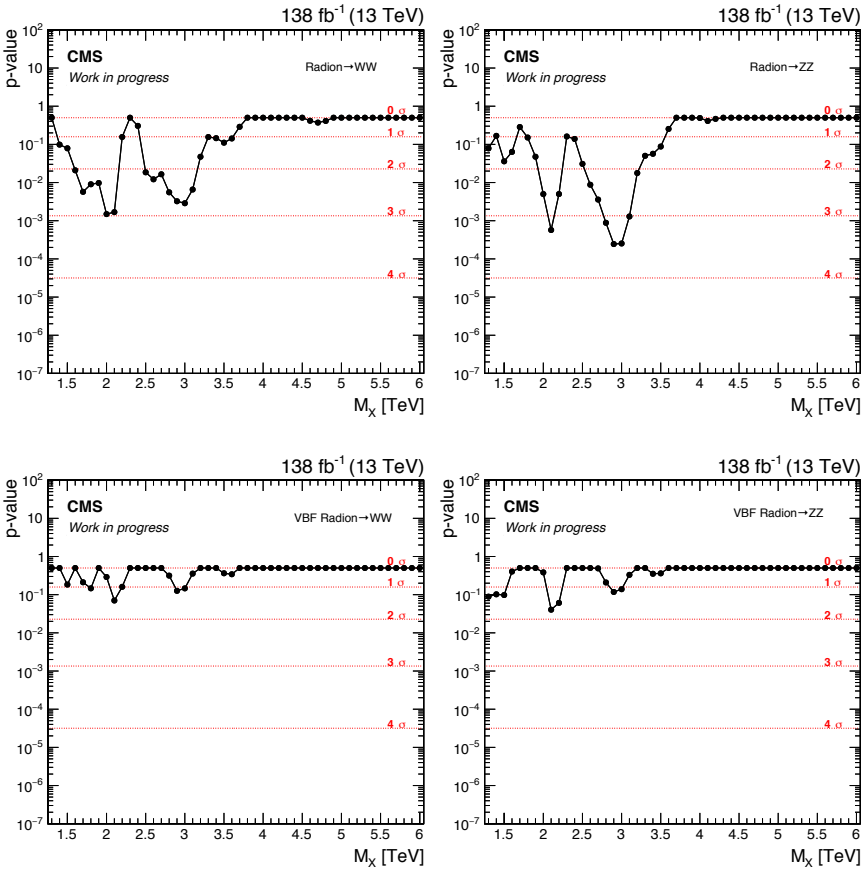


FIGURE D.2: Observed significance using 138fb^{-1} of data for a Radion→ WW (left) and a Radion→ ZZ (right) signals with the combination of all the analysis categories. On the top the ggF production mode is shown while on the bottom the VBF one is presented.

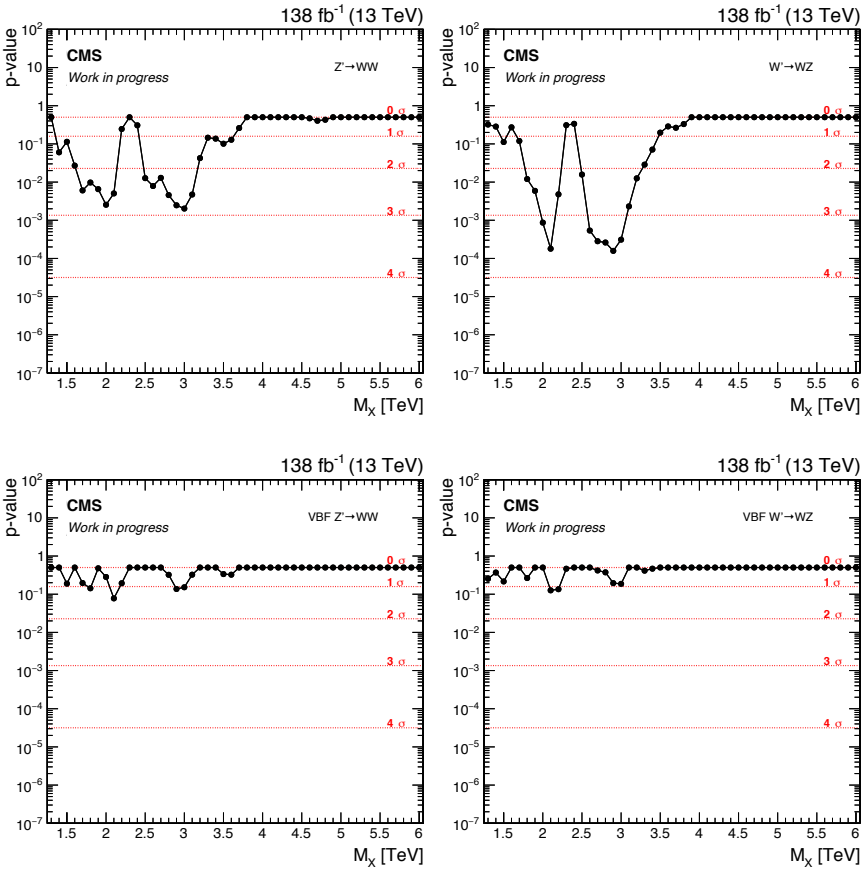


FIGURE D.3: Observed significance using 138fb^{-1} of data respectively for a $Z' \rightarrow WW$ (left) and a $W' \rightarrow WZ$ (right) signals with the combination of all the analysis categories. On the top the DY production mode is shown while on the bottom the VBF one is presented.

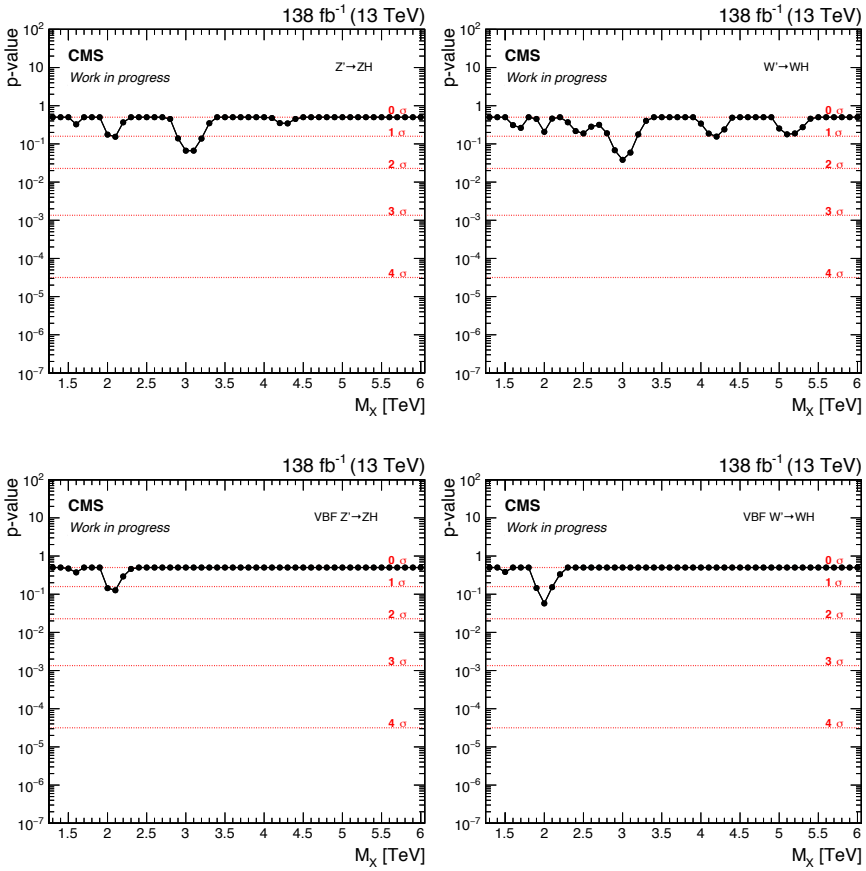


FIGURE D.4: Observed significance using 138fb^{-1} of data for a $Z' \rightarrow ZH$ (left) and a $W' \rightarrow WH$ (right) signals with the combination of all the analysis categories. On the top the DY production mode is shown while on the bottom the VBF one is presented.

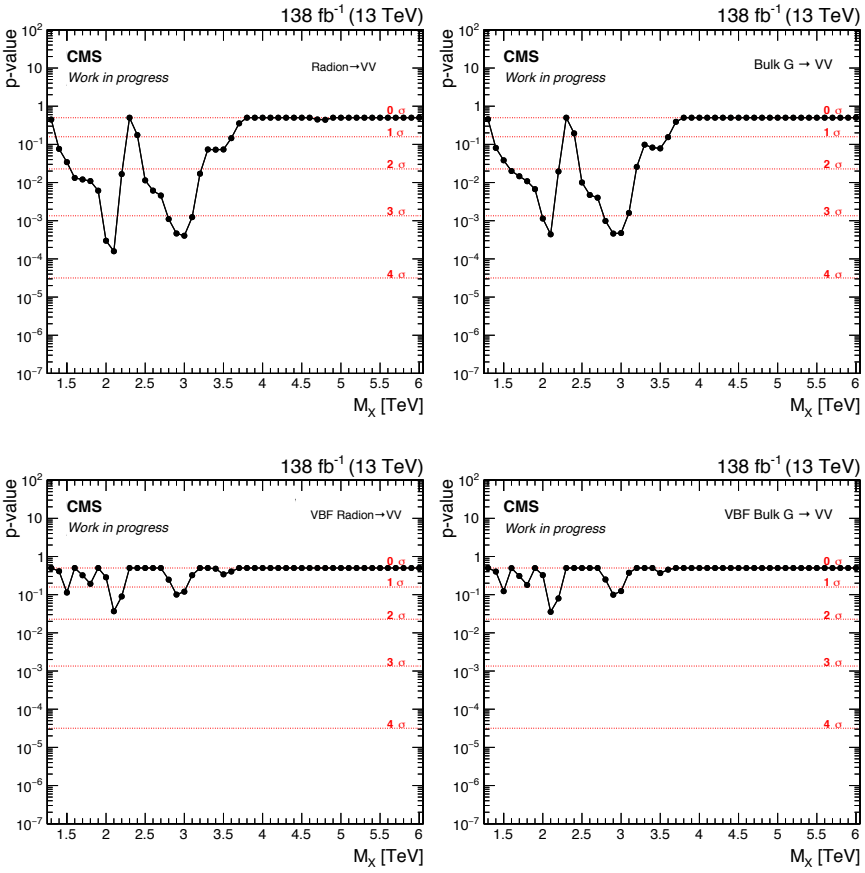


FIGURE D.5: Observed significance using 138fb^{-1} of data for a Radion$\rightarrow$ VV (left) and a $G_{bulk} \rightarrow$ VV (right) signals with the combination of all the analysis categories. On the top the ggF production mode is shown while on the bottom the VBF one is presented.

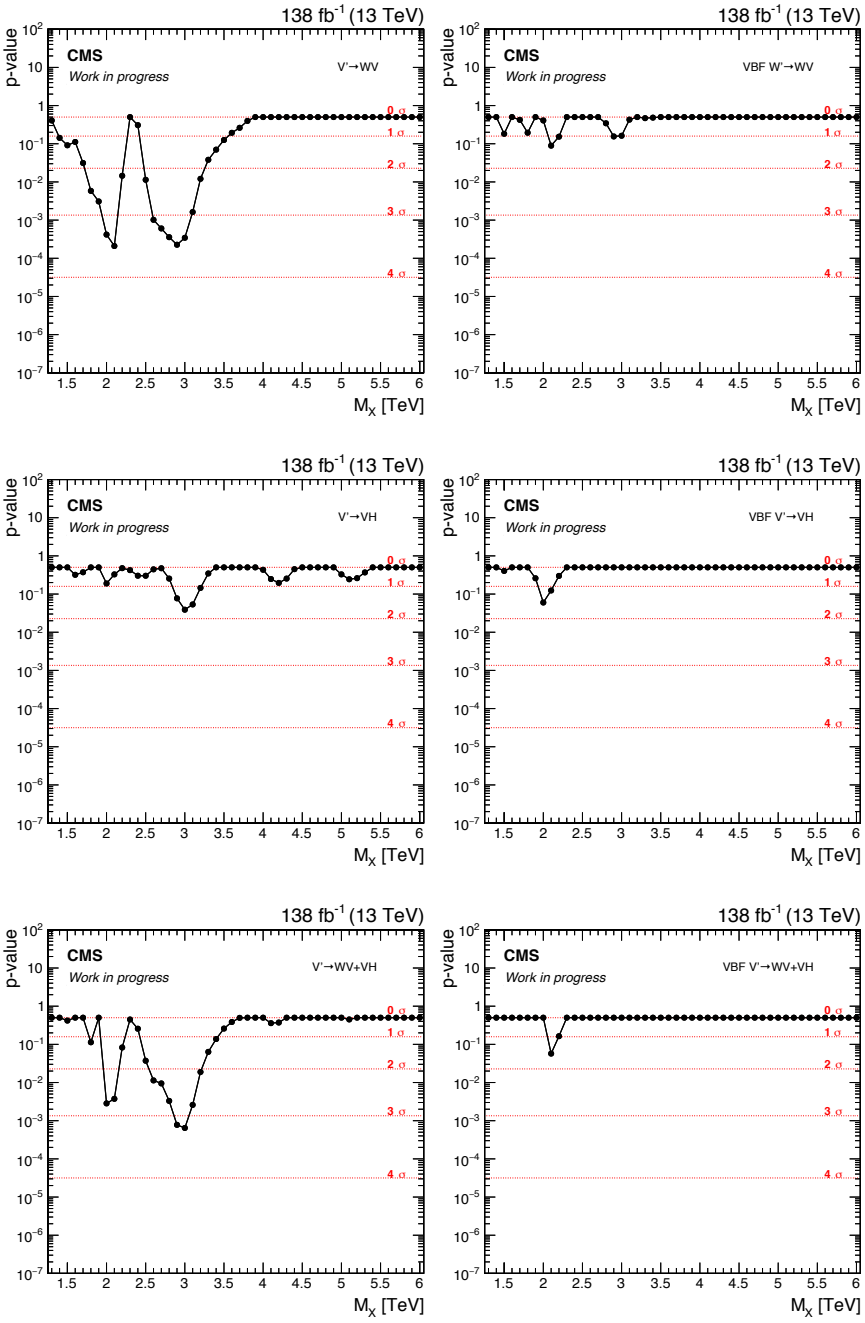


FIGURE D.6: Observed significance using 138fb^{-1} of data for $V' \to WV$ (top), $V' \to VH$ (middle) and $V' \to VV$ combining $V' \to WV$ and $V' \to VH$ (bottom). On the left the DY production mode is shown while on the right the VBF one is presented.

D.2 Observed limits

Figure D.7 shows the resulting 95% confidence level (CL) expected exclusion limits on the signal cross section as a function of the resonance mass for a $G_{bulk} \rightarrow WW$ (left) and a $G_{bulk} \rightarrow ZZ$ (right) signals for the full Run-2 dataset. On the top the ggF production mode is shown while on the bottom the VBF one is presented. The limits are compared with the cross section times the branching fraction to WW or ZZ for a bulk graviton with $k/\overline{M}_{Pl} = 0.5$.

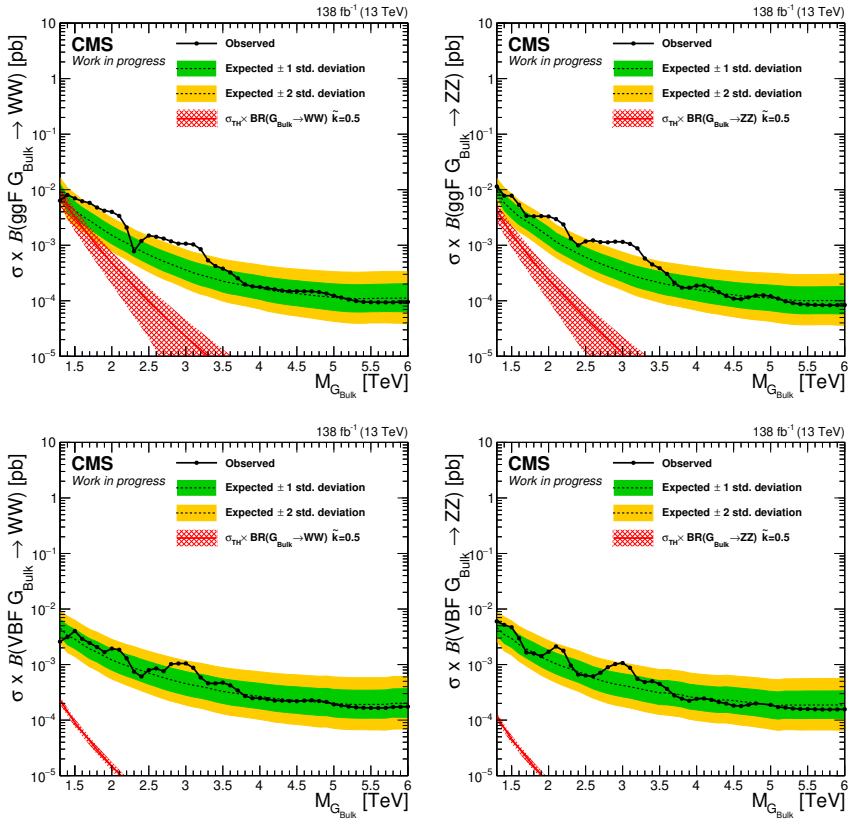


FIGURE D.7: Observed and expected 95% C.L. upper limits on the product of the production cross section (σ) and the branching fraction B, obtained after combining all categories with 138 fb^{-1} of 13 TeV data for a $G_{bulk} \rightarrow WW$ (left) and a $G_{bulk} \rightarrow ZZ$ (right) signals with the combination of all the analysis categories. On the top the ggF production mode is shown while on the bottom the VBF one is presented.

Figure D.8 shows the resulting 95% CL expected exclusion limits on the signal cross section as a function of the resonance mass for a Radion$\rightarrow$ WW (left) and a Radion$\rightarrow$ ZZ (right) signals for the full Run-2 dataset. On the top the ggF production mode is shown while on the bottom the VBF one is presented. The limits are compared with the cross section times the branching fraction to WW or ZZ for a radion with Λ_R= 3 TeV and kl = 35.

Figure D.9 shows the resulting 95% CL expected exclusion limits on the signal cross section as a function of the resonance mass for a $Z' \rightarrow WW$ (left) and a $W' \rightarrow WZ$ (right) signals for the full Run-2 dataset. On the top the DY production mode is shown while on the bottom the VBF one is presented. The limits are compared with the cross section times the branching fraction to WW for the spin-1 Z' particle and to WZ for W' predicted by the HVT model B (top) and model C (bottom).

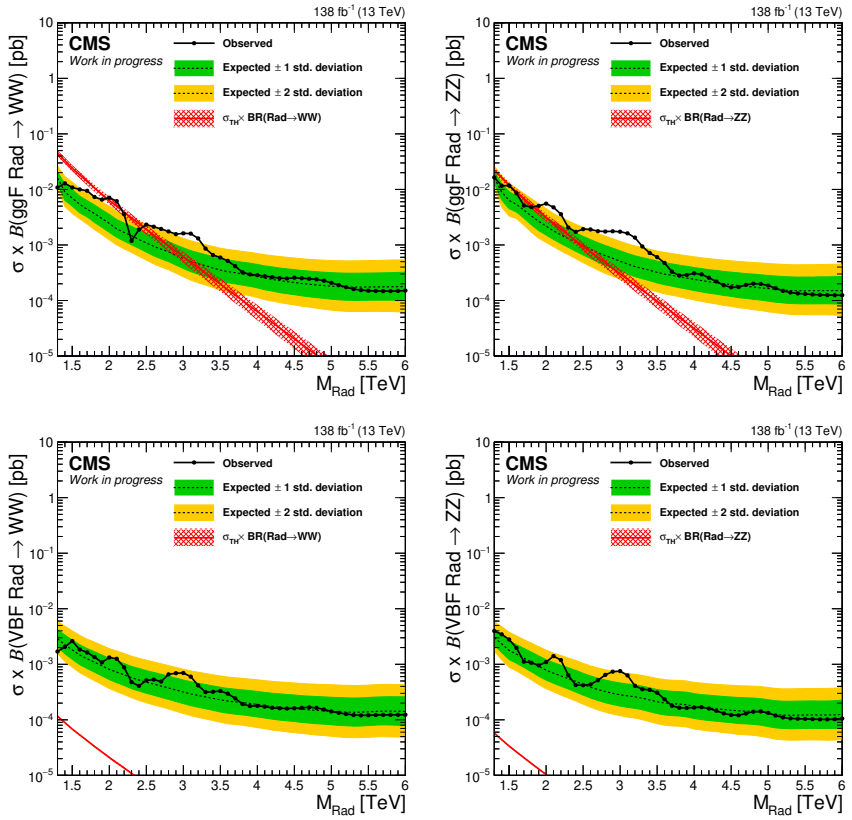


FIGURE D.8: Observed and expected 95% C.L. upper limits on the product of the production cross section (σ) and the branching fraction B, obtained after combining all categories with 138 fb^{-1} of 13 TeV data for a Radion→ WW (left) and a Radion→ ZZ (right) signals with the combination of all the analysis categories. On the top the ggF production mode is shown while on the bottom the VBF one is presented.

Figure D.10 shows the resulting 95% CL expected exclusion limits on the signal cross section as a function of the resonance mass for a $Z' \to ZH$ (left) and a $W' \to WH$ (right) signals for the full Run-2 dataset. On the top the DY production mode is shown while on the bottom the VBF one is presented. The limits are compared with the cross section times the branching fraction to ZH for the spin-1 Z' particle and to WH for W' predicted by the HVT model B (top) and model C (bottom).

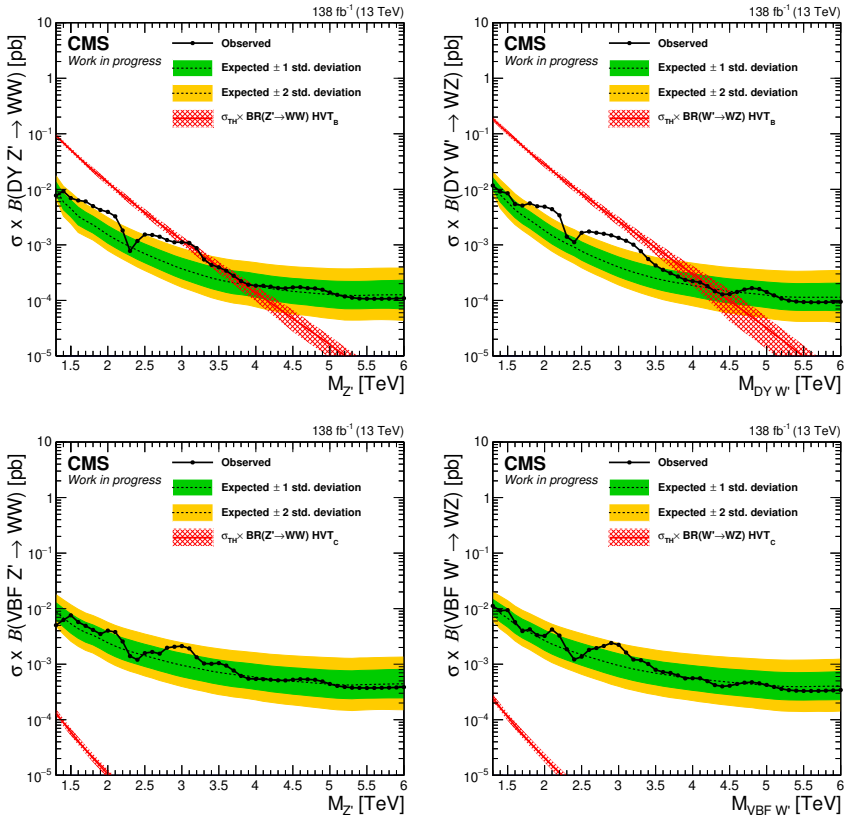


FIGURE D.9: Observed and expected 95% C.L. upper limits on the product of the production cross section (σ) and the branching fraction B, obtained after combining all categories with 138 fb^{-1} of 13 TeV data for a $Z' \to WW$ (left) and a $W' \to WZ$ (right) signals with the combination of all the analysis categories. On the top the DY production mode is shown while on the bottom the VBF one is presented.

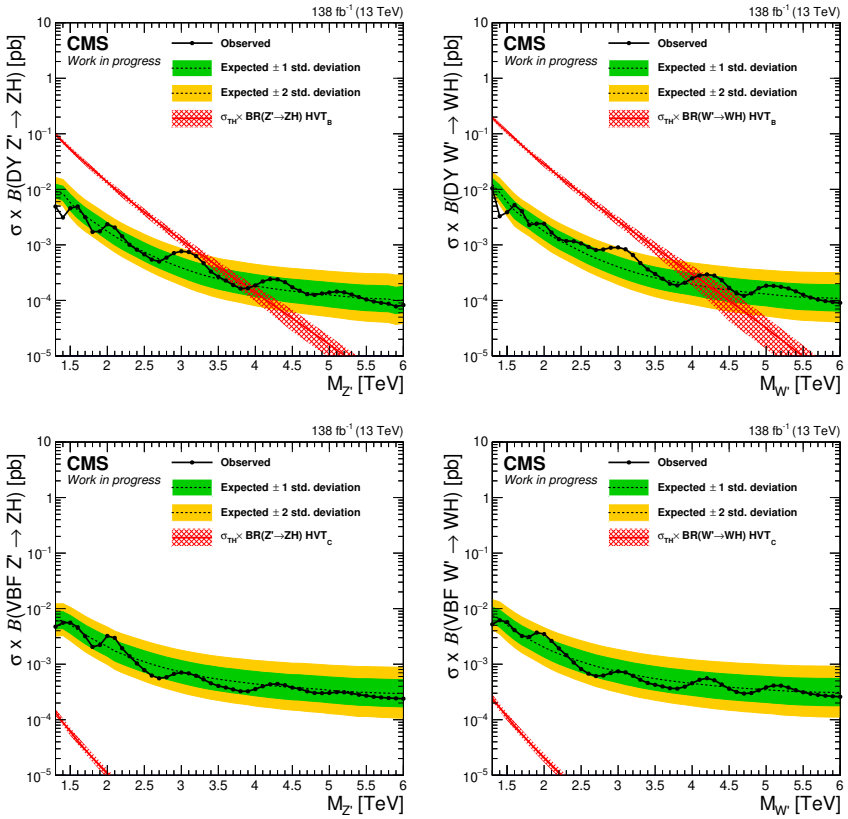


FIGURE D.10: Observed and expected 95% C.L. upper limits on the product of the production cross section (σ) and the branching fraction B, obtained after combining all categories with 138 fb^{-1} of 13 TeV data for a $Z' \to ZH$ (left) and a $W' \to WH$ (right) signals with the combination of all the analysis categories. On the top the DY production mode is shown while on the bottom the VBF one is presented.

E | Simulation setup for substructure studies

The details of the simulation setup for the studies presented in Sec. 11.3 and Sec. 11.4 are:

- CMSSW `9_0_0_pre5-90X`
- conditions: `auto:phase2 realistic`
- geometry: `Extended2023D4`
- era `Phase2C2`
- pileup `AVE_200_BX_25ns`
- signal samples: `/BulkGravToWWToWlepWhad_narrow_M-2000_13TeV-madgraph/`, `/BulkGravToWWToWlepWhad_narrow_M-4000_13TeV-madgraph/`.

The Phase-1 detector simulation added for comparison was obtained with `CMSSW_9_0_0_pre6-PU25ns_90X_mcRun2_asymptotic_v4-v1`.

F FBK sensors and measurements

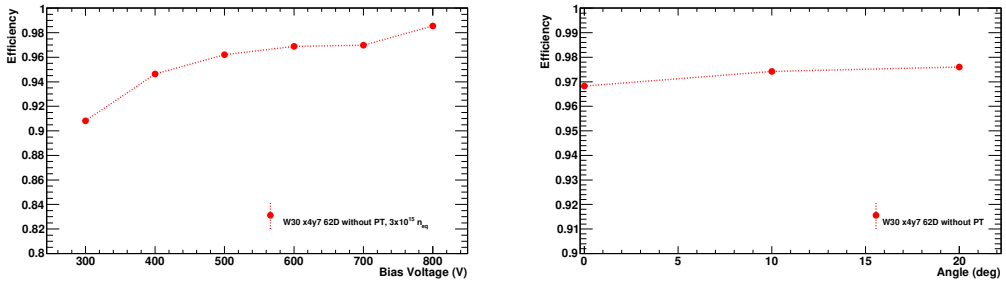
The author of this thesis took part in the first studies performed on planar and "single side process" 3D pixel sensors produced in the context of the *Istituto Nazionale di Fisica Nucleare* (INFN) and the *Fondazione Bruno Kessler* (FBK) R&D campaign[1] for HL-LHC. The results summarized in this chapter were presented at "The European Physical Society Conference on High Energy Physics" in 2017 and published in Ref. [302] by the author of this thesis.

This work was performed under the supervision of Dr. Marco Meschini, in collaboration with Dr. Mauro Dinardo and Prof. Dr. Luigi Moroni. Dr. Maurizio Boscardin, Dr. Gian Franco Dalla Betta, Dr. Gabriele Giacomini, Dr. Marco Meschini, Dr. Alberto Messineo, Dr. Sabina Ronchin and Dr. D.M.S. Sultan designed and produced the sensors. Dr. Lorenzo Uplegger provided assistance in the test beam operations. The author of this thesis, together with Dr. Mauro Dinardo, Dr. Dario Menasce, Dr. Marco Meschini, Prof. Dr. Luigi Moroni, Dr. Caterina Vernieri, Dr. Lorenzo Viliani and Dr. Davide Zuolo, performed the calibration of the tested modules, took part in the data taking and analysis.

F.1 Sensors and setup description

The prototype planar and 3D columnar pixel sensors used for these studies were designed in the INFN R&D campaign to meet the requirements specified in Sec 10.3. The production took place at FBK [303] foundry adopting the new single-side process DRIE technique (Deep Reactive Ion Etching) [304], on Silicon-Silicon Direct Wafer Bonded (SiSi DWB) material with active thicknesses of 100 µm and 130 µm. The sensors have been produced on 6" wafers in different sizes and pitches in order to be individually bump bonded to different readout chips. The designs are compatible with the readout chips available at the time the study was performed, the FE-I4 for ATLAS [305] and the PSI46dig [248] for CMS. The results presented in the following are based on planar and 3D single-chip module pixel sensors bump-bonded to the PSI46dig ($100\times150\,\mu m^2$ pitch). Planar sensors were bump-bonded at Fraunhofer IZM [306] while the Indium bump-bonding of 3D sensors was performed at Leonardo Company [307]. The tests were performed in 2015-17 at the Fermilab Test Beam Facility (FTBF) [308] with a 120 GeV proton beam. The facility houses a tracking telescope [260] with a resolution of about 8 µm on the Detector Under Test (DUT), which can be operated in a humidity and temperature controlled environment. All the devices have been characterised in terms of detection efficiency, which is computed as the ratio of the number of tracks reconstructed by the telescope having a corresponding hit on the detector, and the total number of tracks reconstructed by

[1]This work was supported by the Italian National Institute for Nuclear Physics (INFN). The project received funding also by EU H2020 AIDA-2020, GA no. 654168



(A) Efficiency vs bias voltage (B) Efficiency vs tilt angle

FIGURE F.1: Results for a 100 µm thick planar sensor. Overall detection efficiency at 3×10^{15} n_{eq}/cm^2 as a function of the bias voltage using orthogonal tracks (a) and as a function of the tilting angle at a bias voltage of 600 V (b). Published in Ref. [310].

the telescope and traversing the detector, taking into account track pointing misplacements.

F.2 Planar pixel sensors

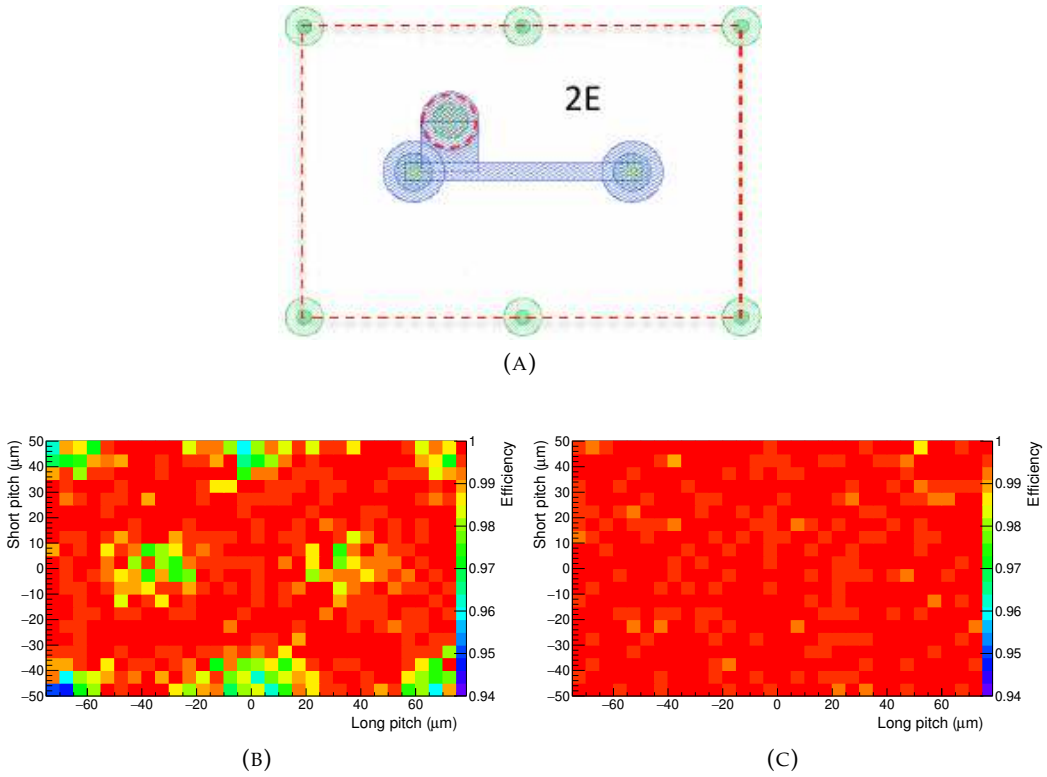
In this section, results from hybrid modules of the first FBK planar batch are presented. The sensors have been produced in various layouts, described in Ref. [309]. The measurements presented in the following are performed on a planar pixel sensor (wafer W30) with 100 µm active thickness and 100×150 µm^2 pitch, chosen as representative example for the perfomance. The sensor design is characterised by a 50 µm large gap between the n^+ pixel implants surrounded by C-shaped p-stop structures and no punch-through structure. To reduce discharges due to the high electric field between sensor and ROC, a frame-shaped BCB (Benzo-Cyclo-Butene) layer has been deposited by IZM on both sensor and ROC periphery. The same device has been tested before and after irradiation at an average fluence of 3×10^{15} n_{eq}/cm^2, performed at IRRAD (CERN) [246] in 2016 with 24 GeV/c protons.

Before irradiation, the sensor was operated at a bias voltage of 40 V at room temperature; the detection efficiency, defined above, has been measured to be always above 99%. The detection efficiency of the irradiated detector (cooled down to $-20\,^{\circ}$C) was measured at different bias voltages (Fig. F.1a) and at different tilting angles with respect to the beam direction (bias voltage of 600 V, Fig. F.1b): the sensor can sustain high voltages without charge multiplication and has a high efficiency, 97% at 600 V.

F.3 3D Columnar Pixel Sensors

The first INFN R&D batch of 3D columnar pixel sensors was produced at FBK foundry in 2015. The 3D sensors have p^+ (ohmic) columns passing through the high resistivity ($\geq$ 3k Ohm·cm) active layer and reaching the low resistivity handle wafer, while the n^+ (junction) columns are stopped at a distance of about 15–20 µm from it, to avoid early breakdown [311]. The bias voltage is applied from the back-side through the handle wafer bulk, highly doped with Boron.

Devices with cells with one, two or three junction columns (1E, 2E, 3E) have been implemented, with the option to have the bump pad just on top of an ohmic column (BO) as an alternative to the standard off-column position (no BO). Morevor, small



(A) (B) (C)

FIGURE F.2: 3D pixel $100\times150\,\mu m^2$ pitch 2E sensor, 130 µm active thickness. Cell scheme with junction columns in blue, ohmic ones in green and bump pads circled in red (a). Cell detection efficiency for orthogonal (b) and 5 degrees inclined (c) tracks. Published in Ref. [11].

pitch sensors were produced, with pixel cell structures of $50\times50\,\mu m^2$ or $100\times25\,\mu m^2$, which have a dedicated adaptation metal network to allow a fraction of small pitch pixels to be readout by the PSI46dig ROC. In the small pitch case only one pixel out of six is read out, all the others being shorted together by the metal grid.

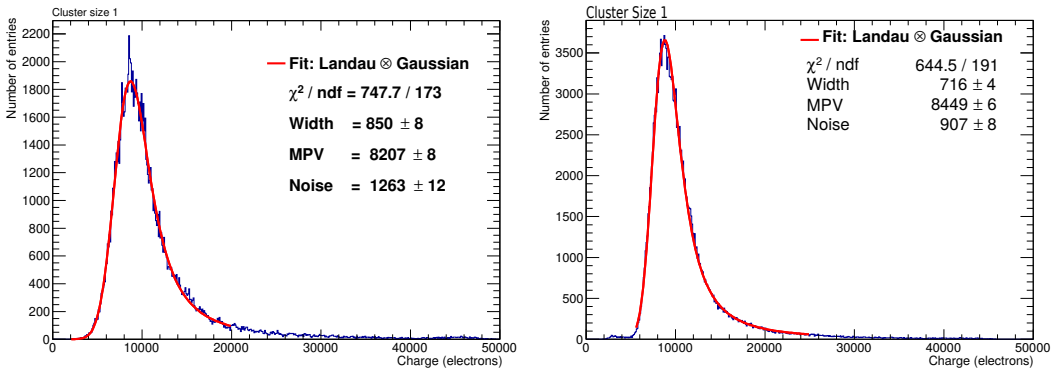
The results presented here refer to non irradiated sensors belonging to two non-thinned wafers (W76, W78) with 130 µm active thickness[2]. The detection efficiency for orthogonal track has been measured to be above 99%.

Detection cell efficiency as a function of the track impact position, averaged on all the cells, is shown in Fig. F.2 for a 3D sensor with $100\times150\,\mu m^2$ pitch and two electrodes (Fig. F.2a). The column positions are still visible for orthogonal tracks (Fig. F.2b) but for 5 degrees incident tracks, 3D pixels recover full efficiency (Fig. F.2c).

Figure F.3 shows the collected charge distributions for a planar sensor (Fig. F.3a) and the above mentioned 3D sensor (Fig. F.3b). A fit with a Landau function convolved with a Gaussian, in red, is also shown. The Most Probable Value ("MPV") of the Landau and the σ of the Gaussian ("Noise") are given. The sensors have the same thickness (130 µm) and collect the same charge but the planar sensor is operated at 100 V while to the 3D sensor a lower voltage of 40 V is applied.

Small pitch 3D pixel sensors have also been tested and representative examples (from wafer W76) are shown in Fig. F.4. Left figures refer to sensors with $100\times25\,\mu m^2$ pitch: Fig. F.4a and Fig. F.4b are two examples of different pixel cell designs and options, as the bump can be either on top of the junction column (BO) or displaced from it; Fig. F.4d shows the distribution of the charge collected by the sensor with

[2] 120 µm effective active thickness due to Boron diffusion from the handle wafer.



(A) Planar (B) 3D

FIGURE F.3: Collected charge distributions for a planar sensor operated at 100V (a) and a 3D sensor (b) operated at 40 V, with same active thickness (130 µm). Published in Ref. [309].

2E BO configuration (bias voltage of 40 V), where the secondary peak at low charge values is due to the contribution of not read out adjacent pixels; the most probable value of the Landau distribution as a function of the bias voltage for three sensors with different layouts is in Fig. F.4f. The three configurations differ for the number of charge collecting electrodes (1E or 2E) and for the bump-bonding position. All three sensors are found to collect a comparable amount of charge within the systematic uncertainties, which are evaluated to be around 5%, mainly due to the readout chip calibration procedures. Right figures are for a $50\times50\,\mu m^2$ pitch sensor: Fig. F.4c shows the pixel cell design, Fig. F.4e the distribution of the charge collected by the sensor and Fig. F.4g the most probable value of the Landau distribution as a function of the bias voltage. Variations in the amount of collected charge between sensors with different pixel sizes are due to the different impact of the not read out cells.

F.4 Summary

This chapter summarizes the first results on planar and "single side process" 3D pixels produced in the INFN-FBK R&D campaign for HL-LHC. The sensors showed an excellent performance. The planar irradiated modules reached a detection efficiency of 97% at a bias voltage of 600 V. The new 3D BO small pitch pixel sensors with the bump pad on the junction column have been produced and were tested on beam for the first time, with outcomes agreeing to the best expectations. The results obtained in beam tests before irradiation and the intrinsic radiation tolerance of 3D sensors confirm that the sensors are very good candidates for the innermost pixel detector layers at the HL-LHC.

At the time this study was performed, the bonding pattern of the available readout chips did not match the reduced sensor pitch size and the readout chips radiation tolerance was not suited for the high fluences of the HL-LHC operation. One of the main challenges of the measurements described here was to operate sensors bump-bonded to readout chips irradiated at the limits of their tolerance, or even beyond, and to decouple effects due to the readout chip patterns with possible problems in the prototypes' design. However, the campaign was successful and further studies at higher fluence and with dedicated readout chips were performed later on. More recent results can be found in Refs. [272, 312–316].

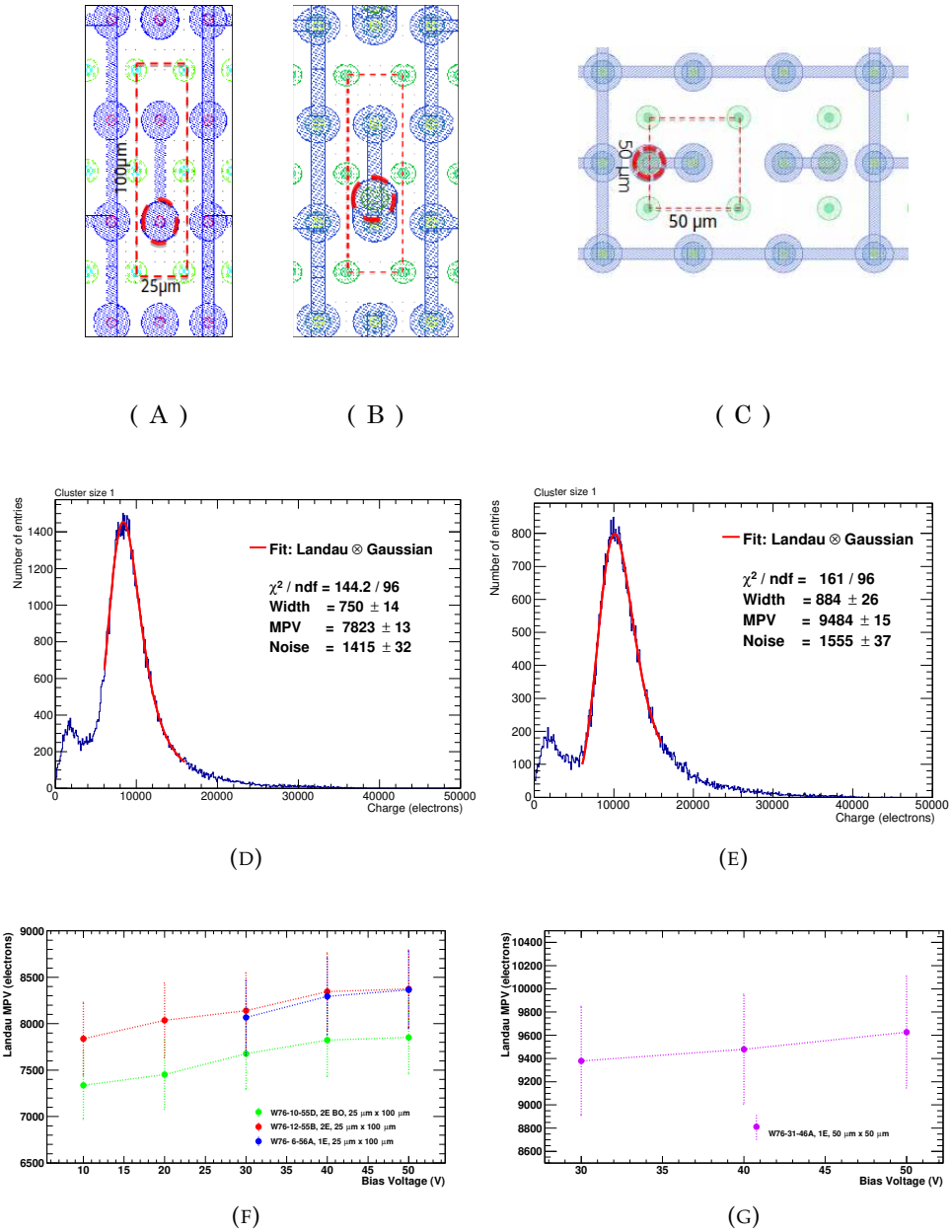


FIGURE F.4: Pixel cell designs and charge collection properties of different small pitch sensors (130 µm thickness), $100\times25\,\mu m^2$ pitch on the left and $50\times50\,\mu m^2$ pitch on the right. From top to bottom: $100\times25\,\mu m^2$ pixel cell design with and without the bump-bonding on top of the column (left) and $50\times50\,\mu m^2$ pixel cell design (right), collected charge distribution and collected charge as a function of the bias voltage. Published in Ref. [309].

Bibliography

[1] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.020.

[2] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021.

[3] Lisa Randall and Raman Sundrum. "A Large mass hierarchy from a small extra dimension". In: *Phys. Rev. Lett.* 83 (1999), pp. 3370–3373. DOI: 10.1103/PhysRevLett.83.3370. arXiv: hep-ph/9905221.

[4] Lisa Randall and Raman Sundrum. "An Alternative to compactification". In: *Phys. Rev. Lett.* 83 (1999), pp. 4690–4693. DOI: 10.1103/PhysRevLett.83.4690. arXiv: hep-th/9906064.

[5] Kaustubh Agashe et al. "Warped Gravitons at the LHC and Beyond". In: *Phys. Rev. D* 76 (2007), p. 036006. DOI: 10.1103/PhysRevD.76.036006. arXiv: hep-ph/0701186.

[6] A. Liam Fitzpatrick et al. "Searching for the Kaluza-Klein Graviton in Bulk RS Models". In: *JHEP* 09 (2007), p. 013. DOI: 10.1088/1126-6708/2007/09/013. arXiv: hep-ph/0701150.

[7] Brando Bellazzini, Csaba Csáki, and Javi Serra. "Composite Higgses". In: *Eur. Phys. J. C* 74.5 (2014), p. 2766. DOI: 10.1140/epjc/s10052-014-2766-x. arXiv: 1401.2457 [hep-ph].

[8] Roberto Contino et al. "On the effect of resonances in composite Higgs phenomenology". In: *JHEP* 10 (2011), p. 081. DOI: 10.1007/JHEP10(2011)081. arXiv: 1109.1570 [hep-ph].

[9] CMS Collaboration. "A multi-dimensional search for new heavy resonances decaying to boosted WW, WZ, or ZZ boson pairs in the dijet final state at 13 TeV". In: *Eur. Phys. J. C* 80 (2019), 237. 49 p. DOI: 10.1140/epjc/s10052-020-7773-5. arXiv: 1906.05977.

[10] Apollinari G. et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*. CERN Yellow Reports: Monographs. CERN, 2017. DOI: 10.23731/CYRM-2017-004. URL: https://cds.cern.ch/record/2284929.

[11] CMS Collaboration. *The Phase-2 Upgrade of the CMS Tracker*. Tech. rep. CERN-LHCC-2017-009. CMS-TDR-014. CERN, 2017. URL: https://cds.cern.ch/record/2272264.

[12] Griffiths, David. *Introduction to elementary particles*. Vch Verlagsgesellschaft Mbh, 2008. ISBN: 978-3-527-40601-2.

[13] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, 1995. ISBN: 978-0-201-50397-5.

[14] Andrew J. Larkoski. *Elementary Particle Physics: An Intuitive Introduction*. Cambridge University Press, 2019. ISBN: 978-1-108-49698-8, 978-1-108-57940-7.

[15] Particle Data Group, P. A. Zyla, et al. "Review of particle physics". In: *Prog. Theor. Exp. Phys.* 2020 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[16] Latham Boyle. *The Standard Model of Particle Physics*. Accessed: 2021-07-04. URL: https://commons.m.wikimedia.org/wiki/File:Standard_Model_Of_Particle_Physics--Most_Complete_Diagram.png.

[17] E. Fermi. "Versuch einer Theorie der β-Strahlen. I". In: *Zeitschrift fur Physik* 88.3-4 (1934), pp. 161–177. DOI: 10.1007/BF01351864.

[18] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[19] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[20] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[21] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton-Hadron Symmetry". In: *Phys. Rev. D* 2 (1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285.

[22] SNO Collaboration. "Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory". In: *Phys. Rev. Lett.* 87 (7 2001), p. 071301. DOI: 10.1103/PhysRevLett.87.071301.

[23] Super-Kamiokande Collaboration. "Evidence for Oscillation of Atmospheric Neutrinos". In: *Phys. Rev. Lett.* 81 (8 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562.

[24] SNO Collaboration. "Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory". In: *Phys. Rev. Lett.* 89 (1 2002), p. 011301. DOI: 10.1103/PhysRevLett.89.011301.

[25] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[26] Makoto Kobayashi and Toshihide Maskawa. "CP Violation in the Renormalizable Theory of Weak Interaction". In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[27] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[28] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[29] Johannes Haller et al. "Update of the global electroweak fit and constraints on two-Higgs-doublet models". In: *Eur. Phys. J. C* 78.8 (2018), p. 675. DOI: 10.1140/epjc/s10052-018-6131-3. arXiv: 1803.01853 [hep-ph].

[30] V. C. Rubin, Jr. Ford W. K., and N. Thonnard. "Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa -> Sc." In: *Astrophysical Journal, Letters* 225 (1978), pp. L107–L111. DOI: 10.1086/182804.

[31] Douglas Clowe et al. "A Direct Empirical Proof of the Existence of Dark Matter". In: *The Astrophysical Journal* 648.2 (2006), pp. L109–L113. DOI: 10.1086/508162.

[32] S. Perlmutter et al. "Measurements of Ω and Λ from 42 High-Redshift Supernovae". In: *The Astrophysical Journal* 517.2 (1999), pp. 565–586. DOI: 10.1086/307221.

[33] Adam G. Riess et al. "Observational evidence from supernovae for an accelerating universe and a cosmological constant". In: *Astron. J.* 116 (1998), pp. 1009–1038. DOI: 10.1086/300499. arXiv: astro-ph/9805201.

[34] KATRIN Collaboration. "Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN". In: *Phys. Rev. Lett.* 123 (22 2019), p. 221802. DOI: 10.1103/PhysRevLett.123.221802.

[35] M. B. Gavela et al. "Standard model CP violation and baryon asymmetry. Part 2: Finite temperature". In: *Nucl. Phys. B* 430 (1994), pp. 382–426. DOI: 10.1016/0550-3213(94)00410-2. arXiv: hep-ph/9406289.

[36] M.E. Shaposhnikov. "Baryon asymmetry of the universe in standard electroweak theory". In: *Nuclear Physics B* 287 (1987), pp. 757–775. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(87)90127-1.

[37] A. D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe". In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.

[38] Gerard 't Hooft et al. "Recent Developments in Gauge Theories." In: *Proceedings of the NATO Advanced Study Institutes Series* (Cargese, France). Ed. by Gerard 't Hooft. Vol. 59. Springer, 1980, pp.1–438. ISBN: 978-1-4684-7573-9. DOI: 10.1007/978-1-4684-7571-5.

[39] Gian F. Giudice. "Naturalness after LHC8". In: *PoS* EPS-HEP2013 (2013), p. 163. DOI: 10.22323/1.180.0163. arXiv: 1307.7879 [hep-ph].

[40] Michael E. Peskin. "On the Trail of the Higgs Boson". In: *Annalen Phys.* 528.1-2 (2016), pp. 20–34. DOI: 10.1002/andp.201500225. arXiv: 1506.08185 [hep-ph].

[41] Nima Arkani–Hamed, Savas Dimopoulos, and Gia Dvali. "The hierarchy problem and new dimensions at a millimeter". In: *Physics Letters B* 429.3 (1998), pp. 263–272. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(98)00466-3.

[42] C. D. Hoyle et al. "Submillimeter tests of the gravitational inverse-square law". In: *Phys. Rev. D* 70 (4 2004), p. 042004. DOI: 10.1103/PhysRevD.70.042004.

[43] Alexandra Oliveira. *Gravity particles from Warped Extra Dimensions, a review. Part I - KK Graviton*. Tech. rep. 2014. arXiv: 1404.0102. URL: https://cds.cern.ch/record/1692776.

[44] A. Carvahlo. *Cross sections for bulk graviton produced through gluon fusion*. Git repository. 2018. URL: https://github.com/CrossSectionsLHC/WED/blob/master/KKGraviton_Bulk/GF_NLO_13TeV_ktilda_0p1.txt.

[45] A. Carvahlo. *Cross sections for bulk graviton produced through vector boson fusion*. Git repository. 2018. URL: https://github.com/CrossSectionsLHC/WED/blob/master/KKGraviton_Bulk/Xjj_VBF_LO_13TeV_ktilda_0p1.txt.

[46] A. Carvahlo. *Cross sections for bulk radion produced through gluon fusion*. Git repository. 2018. URL: https://github.com/CrossSectionsLHC/WED/blob/master/Radion_Bulk/GF_NLO_13TeV_LR_3TeV_kl_35.txt.

[47] A. Carvahlo. *Cross sections for bulk graviton produced through vector boson fusion*. Git repository. 2018. URL: https://github.com/CrossSectionsLHC/WED/blob/master/Radion_RS1/VBF_HXSWG2016_13TeV_LR_3TeV.txt.

[48] E. Accomando et al. "Charged di-boson production at the LHC in a 4-site model with a composite Higgs boson". In: *Phys. Rev. D* 86 (2012), p. 115006. DOI: 10.1103/PhysRevD.86.115006. arXiv: 1208.0268 [hep-ph].

[49] Vernon D. Barger, Wai-Yee Keung, and Ernest Ma. "A Gauge Model With Light W and Z Bosons". In: *Phys. Rev. D* 22 (1980), p. 727. DOI: 10.1103/PhysRevD.22.727.

[50] Christophe Grojean, Ennio Salvioni, and Riccardo Torre. "A weakly constrained W' at the early LHC". In: *JHEP* 07 (2011), p. 002. DOI: 10.1007/JHEP07(2011)002. arXiv: 1103.2761 [hep-ph].

[51] Duccio Pappadopulo et al. "Heavy Vector Triplets: Bridging Theory and Data". In: *JHEP* 09 (2014), p. 060. DOI: 10.1007/JHEP09(2014)060. arXiv: 1402.4431 [hep-ph].

[52] Sally Dawson. "The Effective W Approximation". In: *Nucl. Phys. B* 249 (1985), pp. 42–60. DOI: 10.1016/0550-3213(85)90038-0.

[53] CMS Collaboration. "Combination of CMS searches for heavy resonances decaying to pairs of bosons or leptons". In: *Phys. Lett. B* 798 (2019), 134952. 37 p. DOI: 10.1016/j.physletb.2019.134952. arXiv: 1906.00057.

[54] A. Zucchetta. "Cross sections for HVT model C for VBF production". Git repository. URL: https://github.com/zucchett/HVT.git.

[55] Tommaso Dorigo. "Hadron collider searches for diboson resonances". In: *Progress in Particle and Nuclear Physics* 100 (2018), pp. 211–261. ISSN: 0146-6410. DOI: https://doi.org/10.1016/j.ppnp.2018.01.009.

[56] Roman Kogler. *Advances in Jet Substructure at the LHC: Algorithms, Measurements and Searches for New Physical Phenomena*. Vol. 284. Springer, 2021. ISBN: 978-3-030-72857-1, 978-3-030-72858-8. DOI: 10.1007/978-3-030-72858-8.

[57] Thea Aarrestad. "Searching for diboson resonances in the boosted all-hadronic final state at $\sqrt{s}$ = 13 TeV with CMS". In: *Mod. Phys. Lett. A* 35.32 (2020), p. 2030014. DOI: 10.1142/S0217732320300141. arXiv: 2008.06282 [hep-ex].

[58] ATLAS Collaboration. "Search for diboson resonances in hadronic final states in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 1909 (2019), 091. 41 p. DOI: 10.1007/JHEP09(2019)091. arXiv: 1906.08589.

[59] CMS Collaboration. "Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at sqrt(s)=13 TeV". In: *Eur. Phys. J. C* 77 (2017), 636. 34 p. DOI: 10.1140/epjc/s10052-017-5192-z. arXiv: 1707.01303.

[60] ATLAS Collaboration. "Search for resonances decaying into a weak vector boson and a Higgs boson in the fully hadronic final state produced in proton−proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 102 (2020), 112008. 27 p. DOI: 10.1103/PhysRevD.102.112008. arXiv: 2007.05293.

[61] ATLAS Collaboration. "Search for heavy diboson resonances in semileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Eur. Phys. J. C* 80.12 (2020), p. 1165. DOI: 10.1140/epjc/s10052-020-08554-y. arXiv: 2004.14636 [hep-ex].

[62] CMS Collaboration. *Search for weak vector boson and gluon-gluon fusion production of heavy resonances decaying to* $Z(\nu\bar{\nu})V(qq)$. CMS Physics Analysis Summary. CERN, 2021. URL: https://cds.cern.ch/record/2759299.

[63] CMS Collaboration. *Search for heavy resonances decaying to WW, WZ, or WH boson pairs in the lepton plus merged jet final state at* $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary. CERN, 2021. URL: https://cds.cern.ch/record/2759298.

[64] CMS Collaboration. *Search for a heavy vector resonance decaying to a Z boson and a Higgs boson in proton-proton collisions at* $\sqrt{s}$ *13 TeV*. Tech. rep. Submitted to EPJC. CERN, 2021. arXiv: 2102.08198. URL: https://cds.cern.ch/record/2752268.

[65] ATLAS Collaboration. "Search for heavy resonances decaying into a W or Z boson and a Higgs boson in final states with leptons and b-jets in 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector". In: *JHEP* 03 (2018). [Erratum: JHEP 11, 051 (2018)], p. 174. DOI: 10.1007/JHEP03(2018)174. arXiv: 1712.06518 [hep-ex].

[66] ATLAS Collaboration. "Combination of searches for heavy resonances decaying into bosonic and leptonic final states using 36 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 98 (5 2018), p. 052008. DOI: 10.1103/PhysRevD.98.052008.

[67] Michael Benedikt et al. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, 2004. DOI: 10.5170/CERN-2004-003-V-3. URL: http://cds.cern.ch/record/823808.

[68] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001.

[69] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003.

[70] CMS Collaborationi. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004.

[71] ALICE Collaboration. "The ALICE experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (2008), S08002–S08002. DOI: 10.1088/1748-0221/3/08/s08002.

[72] LHCb Collaboration. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005.

[73] O Adriani et al. *LHCf experiment: Technical Design Report*. Technical design report. LHCf. CERN, 2006. URL: http://cds.cern.ch/record/926196.

[74] Vasiliki A. Mitsou. "The MoEDAL experiment at the LHC: status and results". In: *J. Phys. Conf. Ser.* 873.1 (2017). Ed. by Bohdan Grzadkowski, Jan Kalinowski, and Maria Krawczyk, p. 012010. DOI: 10.1088/1742-6596/873/1/012010. arXiv: 1703.07141 [hep-ex].

[75] TOTEM Collaboration. "The TOTEM Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (2008), S08007–S08007. DOI: 10.1088/1748-0221/3/08/s08007.

[76] W. Herr and B. Muratori. "Concept of luminosity". In: *CERN Accelerator School and DESY Zeuthen: Accelerator Physics*. Sept. 2003.

[77] Vernon D. Barger and R. J. N. Phillips. *COLLIDER PHYSICS*. Addison-Wesley Publishing Company, 1987. ISBN: 978-0-201-14945-6.

[78] CMs Collaboration. "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS". In: (Apr. 2021). arXiv: 2104.01927 [hep-ex].

[79] CMS Collaboration. *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} =$ 13 TeV*. CMS Physics Analysis Summary. CERN, 2018. URL: https://cds.cern.ch/record/2621960.

[80] CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} =$ 13 TeV*. CMS Physics Analysis Summary. CERN, 2019. URL: https://cds.cern.ch/record/2676164.

[81] CMS Collaboration. "Pileup mitigation at CMS in 13 TeV data". In: *JINST* 15.09 (2020), P09018. DOI: 10.1088/1748-0221/15/09/P09018. arXiv: 2003.00503 [hep-ex].

[82] Siona Ruth Davis. *CMS slice image view (transverse/longitudinal/3-D)*. 2016. URL: http://cds.cern.ch/record/2204863.

[83] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, 1997. URL: https://cds.cern.ch/record/349375.

[84] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, 1997. URL: https://cds.cern.ch/record/357153.

[85] CMS Collaboration. *Results related to the Phase1 HE upgrade*. CMS Detector Performance Note. 2018. URL: https://cds.cern.ch/record/2320857.

[86] CMS Collaboration. *The CMS magnet project: Technical Design Report*. Technical design report. CMS. CERN, 1997. URL: https://cds.cern.ch/record/331056.

[87] J. G. Layter. *The CMS muon project: Technical Design Report*. Technical design report. CMS. CERN, 1997. URL: https://cds.cern.ch/record/343814.

[88] CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical design report. CMS. URL: http://cds.cern.ch/record/706847.

[89] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical design report. CMS. CERN, 2002. URL: https://cds.cern.ch/record/578006.

[90] V Karimäki et al. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. CERN, 1997. URL: https://cds.cern.ch/record/368412.

[91] CMS Collaboration. *The CMS tracker: addendum to the Technical Design Report*. Technical design report. CMS. CERN, 2000. URL: https://cds.cern.ch/record/490194.

[92] Serguei Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[93] CMS Collaboration. *Technical proposal for the upgrade of the CMS detector through 2020*. Tech. rep. 2011. URL: https://cds.cern.ch/record/1355706.

[94] A. Dominguez et al. *CMS Technical Design Report for the Pixel Detector Upgrade*. Tech. rep. 2012. URL: https://cds.cern.ch/record/1481838.

[95] Jory Sonneveld. "Commissioning and first results from the CMS phase 1 upgrade pixel detector". In: *PoS* Vertex 2017 (2018), p. 018. DOI: 10.22323/1.309.0018. arXiv: 1807.08987 [physics.ins-det].

[96] H. Chr. Kästli et al. "Design and performance of the CMS pixel detector readout chip". In: *NIM A* 565.1 (2006). Proceedings of the International Workshop on Semiconductor Pixel Detectors for Particles and Imaging, pp. 188 –194. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2006.05.038.

[97] A. Starodumov, P. Berger, and M. Meinhard. "High rate capability and radiation tolerance of the PROC600 readout chip for the CMS pixel detector". In: *Journal of Instrumentation* 12.01 (2017), pp. C01078–C01078. DOI: 10.1088/1748-0221/12/01/c01078.

[98] CMS Tracker Group of the CMS Collaboration. "The CMS Phase-1 Pixel Detector Upgrade". In: *JINST* 16 (2020), P02027. 84 p. DOI: 10.1088/1748-0221/16/02/P02027. arXiv: 2012.14304.

[99] CMS Collaboration. *CMS private page: Pixel Performance Results*. Accessed: 2021-10-12. URL: https://twiki.cern.ch/twiki/bin/view/CMS/PixelOfflinePlots.

[100] CMS Collaboration. *The Performance plots for Phase 1 Pixel Detector 2017*. Accessed: 2021-06-03. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/PixelOfflinePlotsAugust2017#Pixel_Charge_Profiles.

[101] M Swartz et al. "A new technique for the reconstruction, validation, and simulation of hits in the CMS Pixel Detector". In: *PoS* Vertex 2007 (2007), 035. 37 p. DOI: 10.22323/1.057.0035. URL: https://cds.cern.ch/record/1073691.

[102] Morris Swartz. *A Detailed Simulation of the CMS Pixel Sensor*. CMS Note. CERN, 2002. URL: https://cds.cern.ch/record/687440.

[103] Paul Schütze. "Silicon Pixel Detectors - Performance after Irradiation and Application in Three-dimensional Imaging". PhD thesis. Hamburg U., 2019. DOI: 10.3204/PUBDB-2019-04814.

[104] CMS Collaboration. *The Performance plots for Phase 1 Pixel Detector 2017*. Accessed: 2021-06-04. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuidePixelDigitization#Cluster_Charge_Re_weighting_CCR.

[105] Irene Zoi for the CMS Collaboration. *Pixel Charge Profiles in Run-2 Data and Simulation*. CMS Detector Performance Note. 2020. URL: https://cds.cern.ch/record/2720551.

[106] R. P. Feynman. "The behavior of hadron collisions at extreme energies". In: *Conf. Proc. C* 690905 (1969), pp. 237–258.

[107] Jonathan M. Butterworth, Guenther Dissertori, and Gavin P. Salam. "Hard Processes in Proton-Proton Collisions at the Large Hadron Collider". In: *Ann. Rev. Nucl. Part. Sci.* 62 (2012), pp. 387–405. DOI: 10.1146/annurev-nucl-102711-094913. arXiv: 1202.0583 [hep-ex].

[108] G. Altarelli and G. Parisi. "Asymptotic Freedom in Parton Language". In: *Nucl. Phys. B* 126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4.

[109] Yuri L. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics." In: *Sov. Phys. JETP* 46 (1977), pp. 641–653.

[110] V. N. Gribov and L. N. Lipatov. "Deep inelastic electron scattering in perturbation theory". In: *Phys. Lett. B* 37 (1971), pp. 78–80. DOI: 10.1016/0370-2693(71)90576-4.

[111] Richard D. Ball et al. "Parton distributions from high-precision collider data". In: *Eur. Phys. J. C* 77.10 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].

[112] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[113] Andy Buckley et al. "General-purpose event generators for LHC physics". In: *Phys. Rept.* 504 (2011), pp. 145–233. DOI: 10.1016/j.physrep.2011.03.005. arXiv: 1101.2599 [hep-ph].

[114] Stefan Höche. "Introduction to parton-shower event generators". In: *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*. Nov. 2014. DOI: 10.1142/9789814678766_0005. arXiv: 1411.4085 [hep-ph].

[115] Vardan Khachatryan et al. "Measurement of the underlying event activity using charged-particle jets in proton-proton collisions at sqrt(s) = 2.76 TeV". In: *JHEP* 09 (2015), p. 137. DOI: 10.1007/JHEP09(2015)137. arXiv: 1507.07229 [hep-ex].

[116] B. Andersson et al. "Parton fragmentation and string dynamics". In: *Physics Reports* 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(83)90080-7.

[117] Bo Andersson. *The Lund model*. Vol. 7. Cambridge University Press, 2005. ISBN: 978-0-521-01734-3, 978-0-521-42094-5, 978-0-511-88149-7.

[118] D. Amati and G. Veneziano. "Preconfinement as a property of perturbative QCD". In: *Physics Letters B* 83.1 (1979), pp. 87–92. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(79)90896-7.

[119] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), 079. 157 p. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301.

[120] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[121] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[122] M. Bahr et al. "Herwig++ Physics and Manual". In: *Eur. Phys. J. C* 58 (2008), pp. 639–707. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[123] Vardan Khachatryan et al. "Event generator tunes obtained from underlying event and multiparton scattering measurements". In: *Eur. Phys. J. C* 76 (2016), p. 155. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex].

[124] Albert M Sirunyan et al. "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements". In: *Eur. Phys. J. C* 80.1 (2020), p. 4. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex].

[125] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *J. Phys. G* 43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].

[126] Sayipjamal Dulat et al. "New parton distribution functions from a global analysis of quantum chromodynamics". In: *Phys. Rev. D* 93 (3 2016), p. 033006. DOI: 10.1103/PhysRevD.93.033006. arXiv: 1506.07443 [hep-ph].

[127] L. A. Harland-Lang et al. "Parton distributions in the LHC era: MMHT 2014 PDFs". In: *Eur. Phys. J. C* 75 (2015), p. 204. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].

[128] Jun Gao and Pavel Nadolsky. "A meta-analysis of parton distribution functions". In: *JHEP* 07 (2014), p. 035. DOI: 10.1007/JHEP07(2014)035. arXiv: 1401.0013 [hep-ph].

[129] Stefano Carrazza et al. "An Unbiased Hessian Representation for Monte Carlo PDFs". In: *Eur. Phys. J. C* 75 (2015), p. 369. DOI: 10.1140/epjc/s10052-015-3590-7. arXiv: 1505.06736 [hep-ph].

[130] Johan Alwall et al. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". In: *Eur. Phys. J. C* 53 (2008), p. 473. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569 [hep-ph].

[131] Simone Alioli, Sven-Olaf Moch, and Peter Uwer. "Hadronic top-quark pair-production with one jet and parton showering". In: *JHEP* 01 (2012), p. 137. DOI: 10.1007/JHEP01(2012)137. arXiv: 1110.5251 [hep-ph].

[132] Stefan Kallweit et al. "NLO electroweak automation and precise predictions for W+multijet production at the LHC". In: *JHEP* 04 (2015), p. 012. DOI: 10.1007/JHEP04(2015)012. arXiv: 1412.5157 [hep-ph].

[133] Stefan Kallweit et al. "NLO QCD+EW predictions for V+jets including off-shell vector-boson decays and multijet merging". In: *JHEP* 04 (2016), p. 021. DOI: 10.1007/JHEP04(2016)021. arXiv: 1511.08692 [hep-ph].

[134] S. Agostinelli et al. "GEANT4—a simulation toolkit". In: *NIM A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[135] John Allison et al. "GEANT4 developments and applications". In: *IEEE Trans. Nucl. Sci.* 53 (2006), p. 270. DOI: 10.1109/TNS.2006.869826.

[136] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector." In: *JINST* 12 (2017), P10003. 82 p. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965.

[137] Florian Beaudette. *Performance of the Prticle Flow Algorithm in CMS*. Accessed: 2021-06-10. URL: http://indico.cern.ch/event/73513/session/45/contribution/825.

[138] CMS collaboration. "Alignment of the CMS tracker with LHC and cosmic ray data". In: *Journal of Instrumentation* 9.06 (2014), P06009–P06009. DOI: 10.1088/1748-0221/9/06/p06009.

[139] CMS Collaboration. *Additional Run 2 CMS Tracker Alignment Performance Results*. CMS Detector Performance Note. 2020. URL: http://cds.cern.ch/record/2727090.

[140] Wolfgang Adam et al. *Track Reconstruction in the CMS tracker*. CMS Note. CERN, 2006. URL: http://cds.cern.ch/record/934067.

[141] R. Frühwirth. "Application of Kalman filtering to track and vertex fitting". In: *NIM A* 262.2 (1987), pp. 444–450. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(87)90887-4.

[142] Thomas Speer et al. *Vertex Fitting in the CMS Tracker*. CMS Note. CERN, 2006. URL: https://cds.cern.ch/record/927395.

[143] R Frühwirth, Wolfgang Waltenberger, and Pascal Vanlaer. *Adaptive Vertex Fitting*. CMS Note. CERN, 2007. URL: https://cds.cern.ch/record/1027031.

[144] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ex].

[145] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet user manual". In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[146] CMS Collaboration. *Pileup Removal Algorithms*. CMS Physics Analysis Summary. CERN, 2014. URL: http://cds.cern.ch/record/1751454.

[147] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The Catchment Area of Jets". In: *JHEP* 04 (2008), p. 005. DOI: 10.1088/1126-6708/2008/04/005. arXiv: 0802.1188 [hep-ph].

[148] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". In: *Phys. Lett. B* 659 (2008), pp. 119–126. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph].

[149] Vardan Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12.02 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[150] Daniele Bertolini et al. "Pileup Per Particle Identification". In: *JHEP* 10 (2014), p. 059. DOI: 10.1007/JHEP10(2014)059. arXiv: 1407.6013 [hep-ph].

[151] Anna Benecke. "Searches for new heavy bosons and vector-like quarks with the CMS experiment at $\sqrt{s}$ = 13 TeV and novel pileup mitigation techniques". PhD thesis. Hamburg U., 2020. URL: https://ediss.sub.uni-hamburg.de/handle/ediss/8934?mode=full.

[152] Yuri L. Dokshitzer et al. "Better jet clustering algorithms". In: *JHEP* 08 (1997), p. 001. DOI: 10.1088/1126-6708/1997/08/001. arXiv: hep-ph/9707323.

[153] Serguei Chatrchyan et al. "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS". In: *JINST* 6 (2011), P11002. DOI: 10.1088/1748-0221/6/11/P11002. arXiv: 1107.4277 [physics.ins-det].

[154] CMS Collaboration. *Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018*. CMS Detector Performance Note. 2020. URL: https://cds.cern.ch/record/2715872.

[155] Andreas Hoecker. "Physics at the LHC Run-2 and Beyond". In: *2016 European School of High-Energy Physics*. Nov. 2016. arXiv: 1611.07864 [hep-ex].

[156] Andrew J. Larkoski et al. "Soft Drop". In: *JHEP* 05 (2014), p. 146. DOI: 10.1007/JHEP05(2014)146. arXiv: 1402.2657 [hep-ph].

[157] Mrinal Dasgupta et al. "Towards an understanding of jet substructure". In: *JHEP* 09 (2013), p. 029. DOI: 10.1007/JHEP09(2013)029. arXiv: 1307.0007 [hep-ph].

[158] Albert M Sirunyan et al. "Search for massive resonances decaying into WW, WZ, ZZ, qW, and qZ with dijet final states at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 97.7 (2018), p. 072006. DOI: 10.1103/PhysRevD.97.072006. arXiv: 1708.05379 [hep-ex].

[159] Jesse Thaler and Ken Van Tilburg. "Identifying Boosted Objects with N-subjettiness". In: *JHEP* 03 (2011), p. 015. DOI: 10.1007/JHEP03(2011)015. arXiv: 1011.2268 [hep-ph].

[160] S. Catani et al. "Longitudinally-invariant $k_\perp$-clustering algorithms for hadron-hadron collisions". In: *Nucl. Phys. B* 406 (1993), 187–224. 38 p. DOI: 10.1016/0550-3213(93)90166-M. URL: https://cds.cern.ch/record/246812.

[161] Stephen D. Ellis and Davison E. Soper. "Successive combination jet algorithm for hadron collisions". In: *Phys. Rev. D* 48 (1993), pp. 3160–3166. DOI: 10.1103/PhysRevD.48.3160. arXiv: hep-ph/9305266.

[162] ATLAS Collaboration. "Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector". In: *Physics Letters B* 786 (2018), pp. 59–86. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2018.09.013.

[163] A. M. Sirunyan et al. "Observation of Higgs boson decay to bottom quarks". In: *Phys. Rev. Lett.* 121.12 (2018), p. 121801. DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex].

[164] A. M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[165] Albert M Sirunyan et al. "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques". In: *JINST* 15.06 (2020), P06005. DOI: 10.1088/1748-0221/15/06/P06005. arXiv: 2004.08262 [hep-ex].

[166] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90. arXiv: 1512.03385.

[167] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *J. Machine Learning Res.* 15 (2014), pp. 1929–1958.

[168] James Dolen et al. "Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure". In: *JHEP* 05 (2016), p. 156. DOI: 10.1007/JHEP05(2016)156. arXiv: 1603.00027 [hep-ph].

[169] Gilles Louppe, Michael Kagan, and Kyle Cranmer. *Learning to Pivot with Adversarial Networks*. 2017. arXiv: 1611.01046 [stat.ML].

[170] Günter Klambauer et al. "Self-Normalizing Neural Networks". In: *CoRR* abs/1706.02515 (2017). arXiv: 1706.02515.

[171] Albert M Sirunyan et al. "Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 100.11 (2019), p. 112007. DOI: 10.1103/PhysRevD.100.112007. arXiv: 1909.04114 [hep-ex].

[172] Anna Benecke. Illustration provided in private communication.

[173] Irene Zoi et al. *A multi-dimensional search for new heavy resonances decaying to boosted WW, WZ, ZZ, WH or ZH boson pairs in the all-jets final state at 13 TeV*. CMS private documentation 2019/131. 2021.

[174] Serguei Chatrchyan et al. "Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 10 (2011), p. 132. DOI: 10.1007/JHEP10(2011)132. arXiv: 1107.4789 [hep-ex].

[175] L. Gouskos. Private communication between the author and the DeepAK8 developer.

[176] Albert M Sirunyan et al. *A search for the standard model Higgs boson decaying to charm quarks*. Tech. rep. 2020, p. 131. DOI: 10.1007/JHEP03(2020)131. arXiv: 1912.01662 [hep-ex].

[177] Daniela Schäfer. "Search for Diboson Resonances in the All-Hadronic Final State with the CMS Detector at $\sqrt{s} = 13$ TeV." PhD thesis. KIT, 2019. URL: http://cds.cern.ch/record/2711965.

[178] Thea Aarrestad. "A Novel Multidimensional Search for Diboson Resonances in the Boosted Dijet Final State and Encoding Jet Substructure with Machine Learning". PhD thesis. Zurich U., 2019. URL: https://cds.cern.ch/record/2690792.

[179] David Krohn, Jesse Thaler, and Lian-Tao Wang. "Jet Trimming". In: *JHEP* 02 (2010), p. 084. DOI: 10.1007/JHEP02(2010)084. arXiv: 0912.1342 [hep-ph].

[180] Albert M Sirunyan et al. "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JINST* 15 (2020), P10017. DOI: 10.1088/1748-0221/15/10/P10017. arXiv: 2006.10165 [hep-ex].

[181] CMS Collaboration. *Jet algorithms performance in 13 TeV data*. CMS Physics Analysis Summary CMS-PAS-JME-16-003. 2016. URL: http://cds.cern.ch/record/2256875.

[182] CMS Collaboration. *Tracking and Primary Vertex Results in First 7 TeV Collisions*. CMS Physics Analysis Summary CMS-PAS-TRK-10-005. 2010. URL: http://cds.cern.ch/record/1279383.

[183] CMS Collaborationi. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* 12.10 (2017), P10003–P10003. DOI: 10.1088/1748-0221/12/10/p10003.

[184] I. Zoi et al. *Search for VBF produced diboson resonances in all hadronic final states*. CMS private document 2017/314. 2017.

[185] Albert M Sirunyan et al. "Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J.* C77 (2017), p. 636. DOI: 10.1140/epjc/s10052-017-5192-z. arXiv: 1707.01303 [hep-ex].

[186] Giovanni Punzi. "Sensitivity of searches for new signals and its optimization". In: *eConf* C030908 (2003). Ed. by L. Lyons, R. P. Mount, and R. Reitmeyer, MODT002. arXiv: physics/0308063.

[187] M. Oreglia. "A Study of the Reactions $\psi' \to \gamma\gamma\psi$". PhD thesis. Stanford U., 1980.

[188] CMS Collaboration. *Measurement of the differential cross section for* $\mathrm{t\bar{t}}$ *production in the dilepton final state at* $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-TOP-16-011. 2016.

[189] Vardan Khachatryan et al. "Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV". In: *Phys. Rev. D* 95.9 (2017), p. 092001. DOI: 10.1103/PhysRevD.95.092001. arXiv: 1610.04191 [hep-ex].

[190] Max Baak et al. "Interpolation between multi-dimensional histograms using a new non-linear moment morphing method". In: *NIM A* 771 (2015), pp. 39–48. DOI: 10.1016/j.nima.2014.10.033. arXiv: 1410.7388 [physics.data-an].

[191] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[192] G. Cowan. *Statistical data analysis*. Oxford science publications, 1998. ISBN: 978-0-19-850156-5.

[193] Robert D. Cousins. *Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models, with Application to Histograms*. URL: http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.

[194] L Demortier and L Lyons. *Everything you always wanted to know about pulls*. Tech. rep. CDF/ANAL/PUBLIC/5776. CDF, 2002. URL: http://physics.rockefeller.edu/luc/technical_reports/cdf5776_pulls.pdf.

[195] Eilam Gross and Ofer Vitells. "Trial factors for the look elsewhere effect in high energy physics". In: *Eur. Phys. J. C* 70 (2010), pp. 525–530. DOI: 10.1140/epjc/s10052-010-1470-8. arXiv: 1005.1891 [physics.data-an].

[196] Higgs PAG. *CMS private page: Look Elsewhere Effect*. Accessed: 2021-08-01. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/HiggsWG/SWGuideNonStandardCombineUses#Look_Elsewhere_Effect.

[197] Elena Accomando et al. *Exclusion and discovery via Drell-Yan in the 4DCHM*. 2015. arXiv: 1507.04245 [hep-ph].

[198] Huilin Qu and Loukas Gouskos. "Jet tagging via particle clouds". In: *Physical Review D* 101.5 (2020). ISSN: 2470-0029. DOI: 10.1103/physrevd.101.056019.

[199] CMS Collaboration. *Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques*. CMS Detector Performance Note. 2020. URL: https://cds.cern.ch/record/2707946.

[200] Béjar Alonso I. et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report Vol. 10*. CERN Yellow Reports: Monographs. CERN, 2020. DOI: 10.23731/CYRM-2020-0010. URL: https://e-publishing.cern.ch/index.php/CYRM/issue/view/127.

[201] Andrea Dainese et al. *Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC*. Tech. rep. CERN-2019-007. 2019. DOI: 10.23731/CYRM-2019-007. URL: https://cds.cern.ch/record/2703572.

[202] The ATLAS and CMS Collaborations. *Addendum to the report on the physics at the HL-LHC, and perspectives for the HE-LHC: Collection of notes from ATLAS and CMS*. Tech. rep. arXiv:1902.10229. CERN, 2019. DOI: 10.23731/CYRM-2019-007.Addendum.

[203] M. Cepeda et al. "Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC". In: *CERN Yellow Rep. Monogr.* 7.arXiv:1902.00134 (2018), 221–584. 364 p. DOI: 10.23731/CYRM-2019-007.221.

[204] P. Azzi et al. *Report from Working Group 1: Standard Model Physics at the HL-LHC and HE-LHC*. Tech. rep. arXiv:1902.04070. CERN, 2018. DOI: 10.23731/CYRM-2019-007.1.

[205] A. Cerri et al. *Report from Working Group 4: Opportunities in Flavour Physics at the HL-LHC and HE-LHC. Flavour Physics at the HL-LHC and HE-LHC*. Tech. rep. arXiv:1812.07638. Report from Working Group 4 on the Physics of the HL-LHC, and Perspectives at the HE-LHC, 292 pages. CERN, 2018. DOI: 10.23731/CYRM-2019-007.867.

[206] Xabier Cid Vidal et al. *Report from Working Group 3: Beyond the Standard Model physics at the HL-LHC and HE-LHC*. Tech. rep. arXiv:1812.07831. CERN, 2018. DOI: 10.23731/CYRM-2019-007.585.

[207] HL-LHC Project. *LHC project schedule (2020)*. Accessed: 2021-01-07. URL: https://project-hl-lhc-industry.web.cern.ch/content/project-schedule.

[208] D Contardo et al. *Technical Proposal for the Phase-II Upgrade of the CMS Detector*. Tech. rep. CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02. 2015. URL: https://cds.cern.ch/record/2020886.

[209] CMS Collaboration. *The Phase-2 Upgrade of the CMS Endcap Calorimeter*. Tech. rep. CERN-LHCC-2017-023. CMS-TDR-019. CERN, 2017. URL: https://cds.cern.ch/record/2293646.

[210] CMS Collaboration. *The Phase-2 Upgrade of the CMS Barrel Calorimeters*. Tech. rep. CERN-LHCC-2017-011. CMS-TDR-015. CERN, 2017. URL: https://cds.cern.ch/record/2283187.

[211] CMS Collaboration. *A MIP Timing Detector for the CMS Phase-2 Upgrade*. Tech. rep. CERN-LHCC-2019-003. CMS-TDR-020. CERN, 2019. URL: https://cds.cern.ch/record/2667167.

[212] CMS Collaboration. *The Phase-2 Upgrade of the CMS Muon Detectors*. Tech. rep. CERN-LHCC-2017-012. CMS-TDR-016. CERN, 2017. URL: https://cds.cern.ch/record/2283189.

[213] S. Calzaferri. "Production and installation of the first GEM station in CMS". In: *Journal of Instrumentation* 15.09 (2020), pp. C09040–C09040. DOI: 10.1088/1748-0221/15/09/c09040.

[214] Archana Sharma. *FIRST GEM STATION (GE11) INSTALLED IN CMS*. Accessed: 2021-01-08. URL: https://cms.cern/news/first-gem-station-ge11-installed-cms.

[215] CMS Collaboration. *The Phase-2 Upgrade of the CMS L1 Trigger Interim Technical Design Report*. Tech. rep. CERN-LHCC-2017-013. CMS-TDR-017. CERN, 2017. URL: https://cds.cern.ch/record/2283192.

[216] CMS Collaboration. *The Phase-2 Upgrade of the CMS Level-1 Trigger*. Tech. rep. CERN-LHCC-2020-004. CMS-TDR-021. CERN, 2020. URL: https://cds.cern.ch/record/2714892.

[217] J. Chistiansen and M. Garcia-Sciveres. *RD Collaboration Proposal: Development of pixel readout integrated circuits for extreme rate and radiation*. Tech. rep. CERN-LHCC-2013-008. LHCC-P-006. CERN, 2013. URL: http://cds.cern.ch/record/1553467.

[218] S. Orfanelli. "The Phase 2 Upgrade of the CMS Inner Tracker". In: *NIM A* 980 (2020), p. 164396. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.164396.

[219] Frank Hartmann. *Evolution of Silicon Sensor Technology in Particle Physics*. Vol. 231. Springer Tracts in Modern Physics. Springer, 2009. ISBN: 978-3-319-87794-5. DOI: 10.1007/978-3-319-64436-3.

[220] Frank Hartmann. *Evolution of Silicon Sensor Technology in Particle Physics*. Vol. 275. Springer Tracts in Modern Physics. Springer, 2017. ISBN: 978-3-319-64434-9, 978-3-319-64436-3. DOI: 10.1007/978-3-319-64436-3.

[221] L. Rossi et al. *Pixel Detectors: From Fundamentals to Applications*. Particle Acceleration and Detection. Springer-Verlag, 2006. ISBN: 978-3-540-28332-4, 978-3-540-28333-1. DOI: 10.1007/3-540-28333-1.

[222] Glen Knoll. *Radiation Detection and Measurement (4th ed.)* John Wiley, 2010. ISBN: 978-0-470-13148-0.

[223] Michael Moll. "Radiation damage in silicon particle detectors: microscopic defects and macroscopic properties". PhD thesis. Hamburg U., 1999. URL: https://bib-pubdb1.desy.de/record/300958.

[224] J. Lindhard and V. Nielsen. "Nuclear collisions and ionisation fluctuations in charged particle detectors". In: *Phys. Letters* Vol: 2 (1962). DOI: 10.1016/0031-9163(62)90229-9. URL: https://www.osti.gov/biblio/4722888.

[225] Pietro P Altermatt et al. "Reassessment of the intrinsic carrier density in crystalline silicon in view of band-gap narrowing". In: *Journal of Applied Physics* 93 (2003), p. 1598. ISSN: 00218979. DOI: 10.1063/1.1529297.

[226] M. Carlà. *Appunti di Elettronica per Fisici*. 2013. URL: https://studenti.fisica.unifi.it/~carla/appunti/2013-14/.

[227] Simon Ramo. "Currents induced by electron motion". In: *Proc. Ire.* 27 ("1939"), pp. 584–585. DOI: 10.1109/JRPROC.1939.228757.

[228] M. Krammer and F. Hartmann. *Silicon detectors*. URL: https://indico.cern.ch/event/124392/contributions/1339904/attachments/74582/106976/IntroSilicon.pdf.

[229] Maurice Garcia-Sciveres and Norbert Wermes. "A review of advances in pixel detectors for experiments with high rate and radiation". In: *Rept. Prog. Phys.* 81.6 (2018), p. 066101. DOI: 10.1088/1361-6633/aab064. arXiv: 1705.10150 [physics.ins-det].

[230] V. A. J. van Lint et al. "Mechanisms of radiation effects in electronic materials. Volume 1". In: *NASA STI/Recon Technical Report A* 81 (1980), p. 13073.

[231] M. Huhtinen. "Simulation of non-ionising energy loss and defect formation in silicon". In: *NIM A* 491 ("2002"), pp. 194–215. DOI: 10.1016/S0168-9002(02)01227-5.

[232] J. J. Loferski and P. Rappaport. "Radiation Damage in Ge and Si Detected by Carrier Lifetime Changes: Damage Thresholds". In: *Phys. Rev.* 111 (2 1958), pp. 432–439. DOI: 10.1103/PhysRev.111.432.

[233] J. R. Srour et al. "Radiation Damage Coefficients for Silicon Depletion Regions". In: *IEEE Transactions on Nuclear Science* 26.6 (1979), pp. 4783–4791. DOI: 10.1109/TNS.1979.4330228.

[234] A. Chilingarov. "Temperature dependence of the current generated in Si bulk". In: *Journal of Instrumentation* 8.10 (2013), P10003–P10003. DOI: 10.1088/1748-0221/8/10/p10003.

[235] R. Wunstorf. "Systematische Untersuchungen zur Strahlenresistenz von Silizium-Detektoren für die Verwendung in Hochenergiephysik-Experimenten". PhD thesis. Hamburg U., 1992. URL: https://bib-pubdb1.desy.de/record/153817.

[236] M. Moll, E. Fretwurst, and G. Lindström. "Investigation on the improved radiation hardness of silicon detectors with high oxygen concentration". In: *NIM A* 439.2 (2000), pp. 282–292. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(99)00842-6.

[237] M. Moll, E. Fretwurst, and G. Lindström. "Leakage current of hadron irradiated silicon detectors – material dependence". In: *NIM A* 426.1 (1999), pp. 87–93. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(98)01475-2.

[238] Michael Moll. "Displacement damage in silicon detectors for high energy physics. Displacement Damage in Silicon Detectors for High Energy Physics". In: *IEEE Trans. Nucl. Sci.* 65 (2018), 1561–1582. 22 p. DOI: 10.1109/TNS.2018.2819506. URL: https://cds.cern.ch/record/2640820.

[239] J. Schwandt. "CMS Pixel detector development for the HL-LHC". In: *NIM A* 924 (2019), pp. 59–63. DOI: 10.1016/j.nima.2018.08.121.

[240] *Hamamatsu Photonics K.K.* Accessed: 2021-01-23. URL: http://www.hamamatsu.com/.

[241] Finn Feindt. *Beam Test Measurements on Planar Pixel Sensors for the CMS Phase 2 Upgrade*. Tech. rep. CMS-CR-2019-298. CERN, 2019. URL: http://cds.cern.ch/record/2713732.

[242] F. Feindt. "Silicon Pixel Sensors in the Inner Tracking System of the CMS Experiment". In progress. PhD thesis. Hamburg U.

[243] G. Steinbrück. "Development of planar pixel sensors for the CMS Inner Tracker at the High-Luminosity LHC". In: *NIM A* 978 (2020), p. 164438. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.164438.

[244] Tilman Rohe. *ROC4SENS – a generic readout chip for sensor studies.* 2017. URL: https://indico.cern.ch/event/663851/contributions/2788211/attachments/1562585/2460787/R4S_RD50-22-11-2017.pdf.

[245] Maurice Garcia-Sciveres. *The RD53A Integrated Circuit*. Tech. rep. CERN-RD53-PUB-17-001. CERN, 2017. URL: https://cds.cern.ch/record/2287593.

[246] Federico Ravotti et al. "A New High-Intensity Proton Irradiation Facility at the CERN PS East Area". In: *PoS* TIPP2014 (2014), p. 354. DOI: 10.22323/1.213.0354.

[247] Cindro V. *Transnational acces to TRIGA MarkII reactor*. AIDA meeting, Bologna, 2018.

[248] T. Rohe et al. "Radiation hardness of CMS pixel barrel modules". In: *NIM A* 624.2 (2010). New Developments in Radiation Detectors, pp. 414 –418. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2010.03.157.

[249] I. Zoi et al. *Position resolution with 25 µm pitch pixel sensors before and after irradiation*. CMS Note. Accepted by *NIM A*. CERN, 2021. arXiv: 2107.04640 [physics.ins-det]. URL: https://cds.cern.ch/record/2776477.

[250] P. Le Boterf P. Soussan B. Majeed and P. Bouillon. "Development and evaluation of lead free reflow soldering techniques for the flip chip bonding of large gaas pixel detectors on Si readout chip". In: *Proc. EPTC 2014, the 16th Electronic Packaging and Technology Conf. (Orlando, USA 2014)* (2014), pp. 453–7.

[251] H. Pernegger et al. "Radiation hard monolithic CMOS sensors with small electrodes for High Luminosity LHC". In: *NIM A* 986 (2021), p. 164381. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.164381.

[252] Fuyue Wang, Benjamin Nachman, and Maurice Garcia-Sciveres. "Ultimate position resolution of pixel clusters with binary readout for particle tracking". In: *NIM A* 899 (2018), pp. 10–15. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.04.053.

[253] M. Boronat et al. "Physical Limitations to the Spatial Resolution of Solid-State Detectors". In: *IEEE Transactions on Nuclear Science* 62.1 (2015), pp. 381–386. DOI: 10.1109/TNS.2014.2376941.

[254] E H M Heijne et al. "Vectors and submicron precision: redundancy and 3D stacking in silicon pixel detectors". In: *Journal of Instrumentation* 5.06 (2010), pp. C06004–C06004. DOI: 10.1088/1748-0221/5/06/c06004.

[255] Ryan M. Field and Erik H.M. Heijne. "Microscopic imaging of muons and 120GeV/c pion interactions in a single, low-noise 256×256 Si pixel detector". In: *NIM A* 577.3 (2007), pp. 595–603. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2007.03.028.

[256] Simon Spannagel. "Test Beam Measurements for the Upgrade of the CMS Pixel Detector and Measurement of the Top Quark Mass from Differential Cross Sections". PhD thesis. Hamburg U., 2016. URL: https://cds.cern.ch/record/2162902.

[257] Hendrik Jansen et al. "Performance of the EUDET-type beam telescopes". In: *EPJ Tech. Instrum.* 3.1 (2016), p. 7. DOI: 10.1140/epjti/s40485-016-0033-2. arXiv: 1603.09669 [physics.ins-det].

[258] Julien-Christopher Beyer. "Optimisation of pixel modules for the ATLAS inner tracker at the high-luminosity LHC". PhD thesis. 2019. URL: http://nbn-resolving.de/urn:nbn:de:bvb:19-239390.

[259] Caterina Vernieri et al. "Pixel sensors with slim edges and small pitches for the CMS upgrades for HL-LHC". In: *NIM A* 845 (2017). Proceedings of the Vienna Conference on Instrumentation 2016, pp. 189–193. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2016.06.020.

[260] Simon Kwan et al. "The pixel tracking telescope at the Fermilab Test Beam Facility". In: *NIM A* 811 (2016), pp. 162–169. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2015.12.003.

[261] Irene Zoi. "Silicon pixel detectors for the CMS Tracker upgrade at HL-LHC". Master thesis. Università degli studi di Firenze, 2016. URL: http://www.infn.it/thesis/thesis_dettaglio.php?tid=11085.

[262] M. Boronat. "The spatial resolution of DEPFET active pixel detectors". Master thesis. Universitat de Valencia, 2012.

[263] I. Rubinskiy. "An EUDET/AIDA Pixel Beam Telescope for Detector Development". In: *Phys. Procedia* 37 (2012). Ed. by Ted Liu, pp. 923–931. DOI: 10.1016/j.phpro.2012.02.434.

[264] Kazuyoshi Akiba et al. "Charged particle tracking with the Timepix ASIC". In: *NIM A* 661.1 (2012), pp. 31–49. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2011.09.021.

[265] E. Buchanan. "The LHCb Vertex Locator (VELO) Pixel Detector Upgrade". In: *Journal of Instrumentation* 12.01 (2017), pp. C01013–C01013. DOI: 10.1088/1748-0221/12/01/c01013.

[266] I. Gorelov et al. "A measurement of Lorentz angle and spatial resolution of radiation hard silicon pixel sensors". In: *NIM A* 481.1 (2002), pp. 204–221. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(01)01413-9.

[267] S. Spannagel. "Technologies for future vertex and tracking detectors at CLIC". In: *NIM A* 936 (2019). Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors, pp. 612–615. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.08.103.

[268] L. Andricek et al. "Intrinsic resolutions of DEPFET detector prototypes measured at beam tests". In: *NIM A* 638.1 (2011), pp. 24–32. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2011.02.015.

[269] Roma Bugiel et al. "Test-beam results of a SOI pixel-detector prototype". In: *NIM A* 901 (2018), pp. 173–179. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.06.017.

[270] A. Bulgheroni. *Results from the EUDET telescope with high resolution planes*. Tech. rep. EUDET-Report-2009-02. URL: https://www.eudet.org/e26/e27/e50990/eudet-report-09-02.pdf.

[271] Ruth Magdalena Munker et al. *Study of the ALICE Investigator chip in view of the requirements at CLIC*. Tech. rep. 2017. URL: https://cds.cern.ch/record/2284145.

[272] M. Meschini et al. "Radiation resistant innovative 3D pixel sensors for the CMS upgrade at the High Luminosity LHC". In: *NIM A* 978 (2020), p. 164429. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.164429.

[273] Simon Spannagel. "Silicon Vertex & Tracking Detectors for the Compact Linear Collider". In: *PoS* Vertex2019 (2019), p. 044. DOI: 10.22323/1.373.0044. arXiv: 2002.10169 [physics.ins-det].

[274] Ladislav Andricek et al. "Spatial resolution analysis of micron resolution silicon pixel detectors based on beam and laser tests". In: *NIM A* 604.1 (2009). PSD8, pp. 385–389. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2009.01.097.

[275] Ivan Perić et al. "High-voltage pixel detectors in commercial CMOS technologies for ATLAS, CLIC and Mu3e experiments". In: *NIM A* 731 (2013). PIXEL 2012, pp. 131–136. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2013.05.006.

[276] Shun Ono et al. "Development of a pixel sensor with fine space-time resolution based on SOI technology for the ILC vertex detector". In: *NIM A* 845 (2017). Proceedings of the Vienna Conference on Instrumentation 2016, pp. 139–142. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2016.04.119.

[277] Marco Battaglia et al. "Characterisation of a pixel sensor in 0.20µm SOI technology for charged particle tracking". In: *NIM A* 654.1 (2011), pp. 258–265. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2011.05.081.

[278] Daisuke Sekigawa et al. "Fine-Pixel Detector FPIX Realizing Sub-micron Spatial Resolution Developed Based on FD-SOI Technology". In: *Springer Proc. Phys.* 213 (2018). Ed. by Zhen-An Liu, pp. 331–338. DOI: 10.1007/978-981-13-1316-5_62.

[279] T Poikela et al. "Timepix3: a 65K channel hybrid pixel readout chip with simultaneous ToA/ToT and sparse readout". In: *Journal of Instrumentation* 9.05 (2014), pp. C05013–C05013. DOI: 10.1088/1748-0221/9/05/c05013.

[280] Roma Bugiel et al. *High spatial resolution monolithic pixel detector in SOI technology*. Tech. rep. CLICdp-Pub-2020-004. CERN, 2020. DOI: http://dx.doi.org/10.1016/j.nima.2020.164897. URL: https://cds.cern.ch/record/2727594.

[281] Peter Kodys. "High resolution DEPFET active pixel sensors for the Belle II experiment". In: *2nd International Conference on Advancements in Nuclear Instrumentation, Measurement Methods and their Applications*. 2011. DOI: 10.1109/ANIMMA.2011.6172845.

[282] R. Diener et al. "The DESY II test beam facility". In: *NIM A* 922 (2019), pp. 265 –286. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.11.133.

[283] E. Belau et al. "The Charge Collection in Silicon Strip Detectors". In: *NIM* 214.MPI-PAE/Exp El-114 (1983), p. 253. DOI: 10.1016/0167-5087(83)90591-4.

[284] R. Turchetta. "Spatial resolution of silicon microstrip detectors". In: *NIM A* 335 (1993), pp. 44–58. DOI: 10.1016/0168-9002(93)90255-G.

[285] Susanna Cucciarelli, Danek Kotlinski, and Teddy Todorov. *Position Determination of Pixel Hits*. CMS Note CMS-NOTE-2002-049, CERN-CMS-NOTE-2002-049. 2002.

[286] A. Ebrahimi et al. "Position reconstruction for segmented detectors". Accepted by NIM A. 2021.

[287] G. Kramberger et al. "Investigation of Irradiated Silicon Detectors by Edge-TCT". In: *IEEE Trans. Nucl. Sci.* 57.4 (2010), pp. 2294–2302. DOI: 10.1109/TNS.2010.2051957.

[288] R. Klanner et al. "Determination of the electric field in highly-irradiated silicon sensors using edge-TCT measurements". In: *NIM A* 951 (2020), p. 162987. DOI: 10.1016/j.nima.2019.162987. arXiv: 1909.04953 [physics.ins-det].

[289] Gerald R. Lynch and Orin I. Dahl. "Approximations to multiple Coulomb scattering". In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 58.1 (1991), pp. 6–10. ISSN: 0168-583X. DOI: https://doi.org/10.1016/0168-583X(91)95671-Y.

[290] Michelangelo L. Mangano et al. "Matching matrix elements and shower evolution for top-quark production in hadronic collisions". In: *JHEP* 01 (2007), p. 013. DOI: 10.1088/1126-6708/2007/01/013. arXiv: hep-ph/0611129.

[291] CMS Collaboration. *Search for a spin-1 heavy resonance that decays to a Z boson and Higgs boson in the semileptonic final states with Run-2 data*. CMS Physics Analysis Summary CMS-PAS-B2G-19-006. CERN, 2020. URL: https://cds.cern.ch/record/2725672.

[292] Albert A. et al. *Search for dark matter produced with an energetic jet, or ahadronically decaying W or Z boson, at* $\sqrt{s} = 13$ *TeV with the full Run2 dataset*. CMS Internal Note 2019/137. 2019. URL: http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2019_137_v7.pdf.

[293] J. M. Lindert et al. "Precise predictions for $V+$ jets dark matter backgrounds". In: *Eur. Phys. J. C* 77 (2017), p. 829. DOI: 10.1140/epjc/s10052-017-5389-1. arXiv: 1705.04664 [hep-ph].

[294] GEN group. "CMS Private document: Cross sections analyzer". URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/HowToGenXSecAnalyzer#Automated_scripts_to_compute_the.

[295] Albert M Sirunyan et al. *Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV*. Tech. rep. 11. 2019, p. 112007. DOI: 10.1103/PhysRevD.100.112007. arXiv: 1909.04114 [hep-ex].

[296] John M. Campbell, R. Keith Ellis, and Ciaran Williams. "Vector boson pair production at the LHC". In: *JHEP* 07 (2011), p. 018. DOI: 10.1007/JHEP07(2011)018. arXiv: 1105.0020 [hep-ph].

[297] F. Cascioli et al. "ZZ production at hadron colliders in NNLO QCD". In: *Phys. Lett. B* 735 (2014), pp. 311–313. DOI: 10.1016/j.physletb.2014.06.056. arXiv: 1405.2219 [hep-ph].

[298] T. Gehrmann et al. "W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD". In: *Phys. Rev. Lett.* 113.21 (2014), p. 212001. DOI: 10.1103/PhysRevLett.113.212001. arXiv: 1408.5243 [hep-ph].

[299] Albert M Sirunyan et al. *Inclusive search for highly boosted Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV*. Tech. rep. 2020, p. 085. DOI: 10.1007/JHEP12(2020)085. arXiv: 2006.13251 [hep-ex].

[300] Nikolaos Kidonakis. "Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-". In: *Phys. Rev. D* 82 (2010), p. 054018. DOI: 10.1103/PhysRevD.82.054018. arXiv: 1005.4451 [hep-ph].

[301] Nikolaos Kidonakis. "Top Quark Production". In: *Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons*. Nov. 2013. DOI: 10.3204/DESY-PROC-2013-03/Kidonakis. arXiv: 1311.0283 [hep-ph].

[302] Irene Zoi et al. "Beam Test Results of Thin n-in-p 3D and Planar Pixel Sensors for the High Luminosity LHC Tracker Upgrade at CMS". In: *PoS* EPS-HEP2017 (2017). Ed. by Paolo Checchia et al., p. 809. DOI: 10.22323/1.314.0809.

[303] *Fondazione Bruno Kessler, Trento (IT)*. Accessed: 2021-07-29. URL: https://www.fbk.eu/en/.

[304] Gian-Franco Dalla Betta et al. "Development of a new generation of 3D pixel sensors for HL-LHC". In: *NIM A* 824 (2016). Ed. by Maria Giuseppina Bisogni et al., pp. 386–387. DOI: 10.1016/j.nima.2015.08.032. arXiv: 1612.00624 [physics.ins-det].

[305] J Albert et al. "Prototype ATLAS IBL Modules using the FE-I4A Front-End Readout Chip". In: *JINST* 7 (2012), P11010. DOI: 10.1088/1748-0221/7/11/P11010. arXiv: 1209.1906 [physics.ins-det].

[306] *Fraunhofer-Institut für Zuverlässigkeit und Mikrointegration IZM*. Accessed: 2021-07-29. URL: https://www.izm.fraunhofer.de.

[307] *Leonardo Company*. Accessed: 2021-07-29. URL: https://www.leonardocompany.com/it/home.

[308] *Fermilab Test Beam Facility (FTBF)*. Accessed: 2021-07-29. URL: https://ftbf.fnal.gov.

[309] M. Meschini et al. "The INFN-FBK pixel R&D program for HL-LHC". In: *NIM A* 831 (2016). Ed. by Yoshinobu Unno et al., pp. 116–121. DOI: 10.1016/j.nima.2016.05.009.

[310] Marco Meschini et al. *Pixel Detector Developments for Tracker Upgrades of the High Luminosity LHC*. Tech. rep. CERN, 2017. DOI: 10.1007/978-981-13-1316-5_65. URL: https://cds.cern.ch/record/2276248.

[311] Gian-Franco Dalla Betta et al. "The INFN–FBK "Phase-2" R & D; program". In: *NIM A* 824 (2016). Ed. by Maria Giuseppina Bisogni et al., pp. 388–391. DOI: 10.1016/j.nima.2015.08.074. arXiv: 1612.00626 [physics.ins-det].

[312] Davide Zuolo. "Development of silicon pixel sensors for the High Luminosity upgrade of the CMS experiment at LHC and search for Higgs boson pair production in the $b\bar{b}\tau^{+}\tau^{-}$ final state at $\sqrt{s}$ = 13 TeV". PhD thesis. Milan Bicocca U., 2021.

[313] Rudy Ceccarelli et al. "Characterization of planar and 3D Silicon pixel sensors for the high luminosity phase of the CMS experiment at LHC". In: *PoS* EPS-HEP2019 (2020), p. 117. DOI: 10.22323/1.364.0117.

[314] A. Cassese et al. "Performances of highly irradiated 3d and planar pixel sensors interconnected to the RD53A readout chip". In: *JINST* 15.02 (2020). Ed. by Marzio Nessi, p. C02016. DOI: 10.1088/1748-0221/15/02/C02016.

[315] D. Zuolo et al. "R&D for new silicon pixel sensors for the High Luminosity phase of the CMS experiment at LHC". In: *Nuovo Cim. C* 42.4 (2019), p. 185. DOI: 10.1393/ncc/i2019-19185-7.

[316] M. Meschini et al. "First Results on 3D Pixel Sensors Interconnected to the RD53A Readout Chip after Irradiation to 1×10^{16}neq cm^{-2}". In: *JINST* 14.06 (2019), p. C06018. DOI: 10.1088/1748-0221/14/06/C06018. arXiv: 1903.01963 [physics.ins-det].

Acknowledgements

I would like to express my gratitude to the many people who supported me during the years of the PhD. First of all, I would like to thank Andreas Hinzmann and Erika Garutti for allowing me to join their groups and pursue my research interests. It has been a long journey and I am extremely grateful for your support and guidance, for the opportunity to benefit from your expertise while carrying out this work and learn so much.

To Andreas, as my main supervisor, I would like to express my appreciation for the many advices in all the stages and aspects of the work and in the development of my career, for the enthusiasm brought to the projects, and the stimulating and positive environment.

I would like to thank also Elisabetta Gallo, Dieter Horns and Kai Schmidt-Hoberg for being part of my defence committee.

Among the many amazing friends and colleagues met during these years, I would like to thank some of them in particular. I am thankful to Anna Benecke for showing me the way around Hamburg and the university (and much more) from day one until now and for the contribution to this work, always ready to help. I also want to thank my office mates, Arne Reimers, Andrea Malara and Alexander Fröhlich for the pleasant and funny atmosphere, for the reciprocal help and for putting up with all my noises. I would like to express a big thank you to Finn Feindt for the fruitful collaboration in the pixel work and the words of support during the hard moments. I would like to thank Robin Aggleton for helping me solve many weird technical challenges and for the nice chats. I am grateful also to Sam Bein for the words of encouragement, the original and interesting questions you always ask and the good vibes.

I am also deeply grateful to Jennifer Ngadiuba, for her fundamental assistance and guidance in the development of the analysis presented in this thesis and for being there in many crucial moments. Grazie, davvero di cuore.

And at this point I would like to thank also all the other CMS members that together with Andreas, Jennifer and Anna contributed to this challenging work with a very much appreciated collaboration. Especially I would like to thank Clemens Lange for the useful discussions during meetings, Gerrit van Onsem for performing part of the statistical validation when I need time for writing, Daniela Schäfer for her contribution to different pieces of this work and Thea K. Årrestad.

I am also grateful to Daniel Pitzl for answering many questions about the Dreimaster analysis and the nice chats during data taking. I would like to thank also Georg Steinbrück and Jörn Schwandt for their valuable insights in the development of the pixel work.

I dedicate a really special and warm thank you to Sara Cerioli, for the friendship we developed, for being there in the good and in the hard moments, for telling me держись when I needed it. I am also grateful to Andrea Bellandi, Vanessa Grattoni and Matteo Defranchis for being part of my small Hamburg family.

And спасибо to the person that has been with me for almost all my time in Hamburg and my PhD, until life separated us almost at the end of it, for what we shared and how we have been growing.

Ich möchte mich auch bei meiner Lehrerin Martina Bölck bedanken für die gute Stimmung im Deutschkurs, oft gebraucht. Ich habe mehr als nur die deutsche Sprache gelernt. Ausserdem, möchte ich die Frauen 2. Mannschaft des Altona 93 Fußballvereins meinen Dank aussprechen, für die zusammen spielen und trainieren.

A big thank you (or grazie) goes to all friends, because no matter where they are or where I am, I know I can always rely on them.

Dulcis in fundo, vorrei ringraziare la mia famiglia, in particolare mio babbo, mia mamma e mio fratello per il supporto, l'amore e l'affetto che ho ricevuto e per aver sempre trovato il modo di esserci vicini anche quando le distanze sembravano enormi e i mesi lunghissimi.