Observation of Higgs boson decay to bottom quarks

Stephane Brunet Cooperstein

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Physics
Adviser: Professor James Olsen

June 2019

Abstract

The observation of the Standard Model Higgs boson decay to a bottom quark-antiquark pair is presented. The primary contribution to this result is from processes in which the Higgs boson is produced in association with a W or Z boson. The latest measurement of these processes is described, using 41.3 fb^{-1} of proton-proton collision data at center-of-mass energy $\sqrt{s} = 13$ TeV, collected by the CMS experiment in 2017. The significance of the observed excess in data over Standard Model backgrounds is 3.3 standard deviations. The result is combined with similar measurements performed by CMS on previous datasets, resulting in an observed significance of 5.6 standard deviations. The measured signal is well consistent with the Standard Model expectation for a Higgs boson with $m_{\text{H}} = 125$ GeV decaying to bottom quarks, with a precision of 20%.

Acknowledgements

I have been very fortunate to take part in the global scientific effort at CERN as a member of the CMS Collaboration. This has given me the opportunity to work with and learn from many highly talented scientists. The completion of this thesis would not have been possible without the help and support of many individuals.

I would first and foremost like to thank Professor James Olsen, my supervisor, for his constant support and for giving me the opportunity to pursue this highly interesting research in full force. I am very grateful for having had the opportunity to learn and to work under his guidance. Jim has always been extremely supportive while also expecting a high standard in all aspects of research, which is the perfect environment for the development of a young scientist.

I would also like to thank Professor Sridhara Dasu, who believed in me as a young undergraduate student and gave me an incredible opportunity to take part in CMS research at an early stage in my career. Thank you very much Sridhara for your patience and guidance, and for all that you have taught me.

Thank you to Paolo Azzurri for introducing me to Standard Model precision measurements on an analysis which is unfortunately beyond the scope of this thesis. This work has introduced me to new experiences from which I have learned a great deal, and for that I am very grateful.

Thank you as well to Chris Palmer for helping me to develop as a scientist and for his dedicated feedback over many years. Thank you to Jacobo Konigsberg, Rainer Mankel, and Andrea Rizzi for their careful review and many important insights in all phases of the analysis from the initial stages to the final publication.

Thank you to Professor Isobel Ojalvo for taking the time to review this thesis and for her help and advice over many years. Thank you very much as well to Professor Dan Marlow and Professor Mariangela Listanti for taking the time to serve on my thesis committee.

I would also like to thank my parents for nurturing my interest in science and math from a very early age. Their encouragement and support throughought my life has given me every possible opportunity to succeed.

And thank you to Nadya, who has shared with me every moment of these years. I am extremely thankful for all that we have shared together. With your stength and support I feel that I can sustain any challenge in life.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Since the first proton-proton collision data was collected at the Large Hadron Collider (LHC) nearly ten years ago, enormous progress has been made in deepening our knowledge of the fundamental constituents of Nature. We have already not only discovered the Higgs boson, the last particle predicted by the Standard Model that had yet to be observed, but measured its mass to nearly per-mille precision and observed its coupling to both bosons and fermions in multiple decay channels. The Standard Model description of electroweak symmetry breaking via the Higgs mechanism has thus been so far powerfully validated. And yet we know that our description of Nature through the Standard Model must be incomplete, and that there must be unexplored physics at some energy scale that remains unknown. Despite the agreement so far between Standard Model predictions and the six years of Higgs boson measurements since its discovery, further exploring the Higgs sector is one of the most promising avenues towards the discovery of physics beyond the Standard Model. Although great progress has been made since the initial discovery in measuring the properties of the Higgs boson, it should be only the beginning of a scientific era using this new particle to probe Nature.

We are still in an early phase in our study of the Higgs boson. The first observation of Yukawa couplings, a fundamental aspect of the Standard Model Lagrangian responsible for the masses of all fermions (Sec. 2.4), was achieved in only 2016 via observation of H$\to \tau\tau$ decay through the combination of ATLAS and CMS measurements [9]. The direct confirmation of the Yukawa coupling to top quarks was just accomplished in 2018 via the observation of ttH production [10, 11]. Despite these achievements, there remained a fundamental missing piece in our experimental tests of the Higgs boson couplings to third generation fermions: the coupling to bottom quarks.

The Standard Model predicts that the Higgs boson decays to a bottom quark-antiquark pair roughly 58% of the time. The precision on this decay mode is the limiting factor in the indirect constraint on the branching fraction of the Higgs boson to beyond the Standard Model (BSM) particles. Despite the relatively large number of H $\to$ b$\bar{\text{b}}$ events expected at the LHC compared to other Higgs boson decay channels, it is an extremely challenging process to measure at a hadron collider due to overwhelming backgrounds from the production of bottom quarks via strong interactions. It was not originally expected to be able to measure H $\to$ b$\bar{\text{b}}$ at the LHC due to these experimental challenges. The ATLAS technical design report, for example, described the prospects for H $\to$ b$\bar{\text{b}}$ measurement at the LHC as "very difficult, even under the most optimistic assumptions" [12]. An important breakthrough was the understanding that the sensitivity to H $\to$ b$\bar{\text{b}}$ at the LHC is highly enhanced in a kinematic regime where the Higgs boson is produced in association with a high-momentum W or Z boson [13]. This strategy, as well as the use of sophisticated analysis techniques including multiple uses of the latest machine learning technology, has made the observation of H $\to$ b$\bar{\text{b}}$ possible much earlier than was originally expected. This thesis presents the analysis which finally made possible the observation of Higgs boson decay to bottom quarks at CMS. It is the culmination of many years of

dedicated $H \to b\bar{b}$ searches that began at the Large Electron-Positron (LEP) collider at CERN, continued at the Tevatron at Fermilab, and now with LHC data yields the first observation of a Yukawa coupling to a down-type quark [14].

Chapter 2

Theoretical overview

The Standard Model (SM) is a renormalizable quantum field theory describing the interactions between the known fundamental particles via the electromagnetic, weak, and strong forces. A crucial element of the SM is the assumption of local gauge invariance under particular group transformations. A direct consequence of imposing local gauge invariance is the prediction of the existence of the force mediator gauge bosons. The resulting SM Lagrangian is described in terms of a set of fundamental parameters which, once experimentally measured, allow for the quantitative description of all interactions between SM particles. This chapter will describe separately each of the primary features of the SM Lagrangian, then present the combination of these components into the single SM Lagrangian.

2.1 Electroweak theory

The electromagnetic interaction was the first to be described in terms of a quantum field theory, quantum electrodynamics (QED) [15]. In QED, the electromagnetic force is mediated by the massless photon, with the coupling strength to each particle proportional to electric charge. It had been observed experimentally (e.g. beta decay) that the weak force couples exclusively to left-handed fermions (spin-1/2 particles). A

mathematical description of the weak force must therefore violate parity symmetry by differentiating between the left-handed and right-handed fermion components. The fermions are split into two categories: leptons, such as the electron, muon, and the neutrinos, and quarks, such as the up and the down quark, the constituent particles of the proton and neutron. The left-handed components of the lepton and quark fields are paired in SU(2) complex doublets while the right-handed fermion components are represented by U(1) singlets. The particle structure of the electroweak interaction is depicted in Table 2.1. This structure is repeated per fermion "family", of which there are three: the electron, the muon, and the tau lepton, paired with corresponding neutrinos to form three families of leptons. There are similarly three families of quarks, with an "up-type" and "down-type" quark for each family or "generation".

Table 2.1: The groupings of the fermion fields in the electroweak theory. The left-handed components of the leptons and quarks are paired in complex SU(2) doublets, whereas the right-handed components are represented by U(1) singlets.

Fields	Ψ_1	Ψ_2	Ψ_3
Quarks	$\begin{pmatrix} u \\ d \end{pmatrix}_L$	u_R	d_R
Leptons	$\begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L$	$\nu_{\ell,\mathrm{R}}$	ℓ_R

The electroweak free massless Lagrangian in this representation can be written as

$$\mathcal{L}^0_{\mathrm{EW}} = \sum_{j=1}^{3} i\bar{\Psi}_j(x)\gamma^\mu\partial_\mu\Psi_j(x), \tag{2.1}$$

where the Ψ_j are the fields given in Table 2.1. This Lagrangian $\mathcal{L}^0_{\mathrm{EW}}$ describes a theory of free massless fermions which do not interact. $\mathcal{L}^0_{\mathrm{EW}}$ has a clear global symmetry under rotations in either $\mathrm{SU(2)_L}$ or $\mathrm{U(1)_Y}$, where the index L is specified for the SU(2)

transformations because only the left-handed fermion components are represented by SU(2) doublets. Similarly the index Y corresponds to a conserved quantity, the weak hypercharge, which will be described below. The global rotations have the generic form given by

$$\Psi(x) \rightarrow e^{iy_1\beta}\Psi(x), \Psi(x) \rightarrow e^{iy_2\beta}e^{\frac{i}{2}\vec{\sigma}\cdot\vec{\alpha}}\Psi(x), \tag{2.2}$$

where β, y_1, and y_2 are arbitrary constants, σ_i are the Pauli matrices which generate the SU(2) group and $\vec{\alpha}$ is an arbitrary three-vector specifying the rotation. The left transformation corresponds to the global symmetry under $\mathrm{U(1)_Y}$, while the right transformation corresponds to the global symmetry under $\mathrm{SU(2)_L}$. The key ingredient of the electroweak formalism, and in fact the SM in general, is to then assume that these global symmetries are also preserved locally such that with the generalization $\beta \rightarrow \beta(x)$, $\vec{\alpha} \rightarrow \vec{\alpha}(x)$ the gauge symmetries are preserved. The assumption of local gauge symmetry is a statement that two different observers can assume distinct transformations without changing the physical predictions of the theory. The electroweak free massless Lagrangian $\mathcal{L}^0_{\mathrm{EW}}$ as written in Equation 2.1, however, is not invariant under the local gauge transformations due to the derivative term propagating to $\beta(x)$ and $\vec{\alpha}(x)$. In order to preserve local gauge invariance it is necessary to introduce the covariant derivative given by (for the right-handed $\mathrm{U(1)_Y}$ singlets)

$$D_\mu\Psi(x) \equiv [\partial_\mu + ig'y_2B_\mu(x)]\Psi(x), \tag{2.3}$$

where g' is an additional arbitrary constant and $B_\mu(x)$ is the U(1) gauge field. Similarly for the left-handed SU(2) doublets a covariant derivative is introduced given by

$$D_\mu\Psi(x) \equiv [\partial_\mu + ig'y_1B_\mu(x) + ig\frac{\vec{\sigma}}{2}\vec{W}_\mu]\Psi(x), \tag{2.4}$$

where each of the three components of $\vec{W}_\mu$ is a gauge field of SU(2). Note that it is not possible to include an additional arbitrary y_k constant for the $\mathrm{SU(2)_L}$ transformations as was done for the transformations under $\mathrm{U(1)_Y}$. This is due to the non-commutative nature of SU(2). It effectively means that while individual fermions can carry unique "charge" under the $\mathrm{U(1)_Y}$ transformations, each fermion is simply either charged or not under $\mathrm{SU(2)_L}$. The preservation of local $\mathrm{SU(2)_L}$ x $\mathrm{U(1)_Y}$ gauge symmetry has required the addition of four massless gauge fields. This is a generic feature of imposing local gauge symmetry, and the form of the resulting gauge fields is a feature of the group. It is important to note that the introduction of a mass term for any of the fields would break the local gauge symmetry, such that the fields must all be massless. The electroweak Lagrangian can therefore be compactly written as

$$\mathcal{L}_{\mathrm{EW}} = \frac{-1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{2} \mathrm{Tr}[\widetilde{W_{\mu\nu}} \widetilde{W^{\mu\nu}}] + \sum_{k=1}^{3} i \bar{\Psi}_k \gamma_\mu D_\mu \Psi_k, \tag{2.5}$$

where

$$\begin{aligned} B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu, \\ \widetilde{W_{\mu\nu}} = \frac{\vec{\sigma}}{2} \vec{W}_{\mu\nu} &= \frac{-i}{g} [(\partial_\mu + i g \widetilde{W_\mu}), (\partial_\nu + i g \widetilde{W_\nu})]. \end{aligned} \tag{2.6}$$

The inherent global $\mathrm{SU(2)_L}$ x $\mathrm{U(1)_Y}$ symmetry of the electroweak Lagrangian and the commutation of the $\mathrm{SU(2)_L}$ generators with $\mathrm{U(1)_Y}$ generators implies two conserved quantities, the weak isospin T and the hypercharge Y. As previously mentioned, the non-commutative nature of $\mathrm{SU(2)_L}$ does not allow for an arbitrary constant per fermion field in the covariant derivative. This means that all left-handed fermions have quantum number T = 1/2 and the right-handed fermions have T = 0 since they are singlets in $\mathrm{SU(2)_L}$. This can be considered in close analogy with the quantum mechanics spin operator for a particle with total spin S = 1/2. Due to the

noncommutativity of the projection operators $S_i = \frac{1}{2}\sigma_i$, it is not possible to measure all three components of spin simultaneously. Particles are instead labeled by the total spin S and the third component S_3. In analogy, the left-handed fermions possess the quantum number $T_3 = \pm 1/2$, which must be conserved in all weak interactions. The neutrinos and up-type quarks have $T_3 = +1/2$, while the electron, muons, taus, and down-type quarks have T_3 = -1/2.

The four gauge fields describe the four gauge boson mediator particles of the electroweak theory: the electrically neutral photon and Z boson, and the $W^{\pm}$ bosons. Two of the four gauge fields, W_μ^3 and B_μ, are electrically neutral, meaning that both fields couple particles with their corresponding antiparticles. It is tempting to assume that one field corresponds to the photon while the other field describes the Z boson. It is, however, very well established that the electromagnetic force and therefore the photon does not differentiate between the left-handed and right-handed components of the fermion fields, as is the case for both W_μ^3 and B_μ. A transformation of the W_μ^3 and B_μ fields is performed in order to represent the neutral currents in terms of the physical currents associated with the photon and Z boson:

$$\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix}. \tag{2.7}$$

It relevant to note that under $\mathrm{SU(2)_L}$ x $\mathrm{U(1)_Y}$ symmetry this transformation is actually forbidden because W_μ^3 and B_μ derive from the separate $\mathrm{SU(2)_L}$ and $\mathrm{U(1)_Y}$ gauge symmetries. It is however known that this symmetry is broken due to the experimentally observed nonzero mass of the Z boson. The mechanism for the electroweak symmetry breaking which gives rise to a mass term for the Z boson will be discussed in Section 2.3. For this section it suffices to note that $m_Z > 0$ allows for

this transformation. With this transformation, A_μ describes the photon field from electromagnetism given that

$$g \sin\theta_W = g' \cos\theta_W = e, \tag{2.8}$$

where e is the standard unit of electric charge from electrodynamics, the electric charge of one electron. The relation between weak hypercharge, electric charge, and weak isospin is given by Y = Q - T_3. Since the weak hypercharge Y is arbitrary per fermion, it can be assigned such that this relation holds for all fermions and reproduces the known fermion electric charges. With this specified, the full set of electroweak interactions between fermions, the electroweak gauge bosons Z, $W^{\pm}$, and the photon is quantitatively predicted. There is however the remaining issue that the introduction of any mass term for either the fermion or gauge fields would break local gauge invariance. This will be addressed in Section 2.3.

2.2 Quantum chromodynamics

The quarks are subject to an additional force, the strong force, which is responsible for the confinement of quarks within the nucleus of the atom despite electromagnetic repulsion forces between the quarks. The strong force is described by the theory of quantum chromodynamics (QCD). In QCD, quarks have an additional "color" charge, red, green, or blue, such that any bound state of quarks must form a color singlet. QCD is a quantum field theory with $SU(3)_C$ triplets grouping each quark flavor with its three color representations.

The free-field Lagrangian for massless quarks in QCD can be written as

$$\mathcal{L}^0_{\mathrm{QCD}} = \bar{\Psi}(x) i\gamma^\mu \partial_\mu \Psi(x), \tag{2.9}$$

where $\Psi(x)$ are now $\mathrm{SU(3)_C}$ triplets for the quark fields. Similarly to the electroweak theory, it is assumed that QCD preserves not only global but also local gauge invariance under transformations in $\mathrm{SU(3)_C}$ given by

$$\Psi(x) \rightarrow e^{ig_s \frac{\lambda_a}{2} \theta_a(x)} \Psi(x), \tag{2.10}$$

where λ_a are the eight Gell-Mann matrices that generate SU(3), similar to the three Pauli matrices σ_b that generate SU(2). The QCD coupling strength, g_s, is analogous to the coupling strengths g and g' from the electroweak theory (Sec. 2.1). Note that $\alpha_s \equiv \frac{g_s^2}{4\pi}$ is often referred to rather than g_s. Due to the non-commutative nature of SU(3), fields cannot carry a unique charge but rather are either charged under $\mathrm{SU(3)_C}$ or not, similar to the weak interaction under $\mathrm{SU(2)_L}$. In order to preserve gauge invariance it is necessary to introduce the covariant derivative

$$D_\mu = \partial_\mu - i g_s G_\mu^a(x) \frac{\lambda_a}{2}, \tag{2.11}$$

where the eight gauge fields $G_\mu^a(x)$ correspond to the eight gluons that mediate the strong interaction. The introduction of these additional gauge fields, as in the electroweak theory, is necessary in order to preserve local gauge symmetry. The full Lagrangian for QCD is thus

$$\mathcal{L}_{\mathrm{QCD}} = \bar{\Psi}(x) i\gamma^\mu \partial_\mu \Psi(x) - g_s \bar{\Psi}(x) \gamma^\mu \frac{\lambda^a}{2} \Psi(x) G_\mu^a - \frac{1}{4} G_a^{\mu\nu} G_{\mu\nu}^a, \tag{2.12}$$

where

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{\mathrm{abc}} G_\mu^b G_\nu^c. \tag{2.13}$$

The final term in $\mathcal{L}_{\mathrm{QCD}}$ is the free kinetic term for the eight gluon gauge fields. The structure constants for SU(3), f^{abc}, are defined by the commutation relation $[\lambda^a, \lambda^b] =$

$if^{abc}\lambda^c$. Its parallel for SU(2) in the electroweak theory is the Levi-Civita symbol ϵ_{ijk}. By definition the structure constant of a group is nonzero for any non-commutative (non-Abelian) group. Thus self interactions between the gauge mediator fields are a direct consequence of the non-Abelian nature of SU(2) and SU(3). Moreover, there is therefore no photon self-interaction in the SM due to the commutative nature of U(1).

It is interesting to note that the introduction of a quark mass term would not violate local gauge symmetry under $SU(3)_C$ transformation, and could in principle be included in QCD to account for the known masses of the quarks. The introduction of any quark mass term, however, would break local gauge invariance under $SU(2)_L$ transformation once combined with the electroweak theory.

2.3 Spontaneous electroweak symmetry breaking

As described in Section 2.1, the introduction of a mass term for any fermion or gauge field would not preserve local gauge symmetry under $SU(2)_L$ x $U(1)_Y$, crucial to the SM description of the electroweak interaction. This is in clear tension with the experimental observation of nonzero masses for the fermions as well as the W and Z bosons. It is therefore necessary that $SU(2)_L$ x $U(1)_Y$ is a broken symmetry in order for the SM to describe Nature. Massive vector bosons have three degrees of polarization whereas massless vector bosons have only two. Adding a mechanism which yields mass terms for the $W^{\pm}$ and Z bosons therefore requires the addition of at least three degrees of freedom. A minimal choice is the addition of a complex scalar $SU(2)_L$ doublet $\Phi(x)$ to the SM Lagrangian, which contains four degrees of freedom [16, 17, 18, 19, 20]. Three degrees of freedom will be absorbed by the $W^{\pm}$ and Z boson longitudinal polarizations, while the fourth degree of freedom will be

shown to correspond to the mass of the scalar field. The general form of $\Phi(x)$ can be written as

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1(x) + i\phi^2(x) \\ \phi^3(x) + i\phi^4(x) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix}, \tag{2.14}$$

where a hypercharge $Y_\Phi = 1/2$ has been assigned to Φ such that the component $\phi^+(x)$ corresponds to a charged scalar field and $\phi^0(x)$ corresponds to a neutral scalar field. $\Phi(x)$ is introduced to the SM Lagrangian with the term

$$\mathcal{L}_H = D_\mu \Phi(x)^\dagger D^\mu \Phi(x) - V(\Phi(x)), \tag{2.15}$$

where the first component is the kinetic term for the scalar doublet and the second term

$$V(\Phi(x)) = -\mu^2 \Phi^\dagger(x)\Phi(x) + \lambda(\Phi^\dagger(x)\Phi(x))^2 \tag{2.16}$$

describes a potential energy term for the scalar field with $\lambda > 0$ and $\mu^2 > 0$. This potential has a set of degenerate minima whenever $\sqrt{\Phi^\dagger\Phi} = |\Phi| = \sqrt{\frac{\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}} > 0$, where $v \equiv \sqrt{\frac{\mu^2}{\lambda}}$ is referred to as the vacuum expectation value. Electroweak local gauge symmetry is broken by assuming a particular minimum and describing the field H(x) in terms of excitations about the minimum

$$\begin{pmatrix} \Phi^+ \\ \Phi_0 \end{pmatrix} \rightarrow e^{i\frac{\sigma_i}{2}\theta_i(x)} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{2.17}$$

where the three arbitrary rotation degrees of freedom $\theta_i(x)$ correspond to the three massless Goldstone bosons generated by the spontaneous symmetry breaking. The $\mathrm{SU(2)_L}$ symmetry allows for a rotation such that any dependence on the $\theta_i(x)$ is removed. As mentioned above, these three degrees of freedom are absorbed by the

W and Z boson transverse polarizations. The resulting $\Phi(x)$ can therefore be simply expressed as

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{2.18}$$

Note that this particular minimum has been chosen such that the resulting scalar field has zero electric charge. A different choice for Φ would change the definition of the electric charge, but would result in exactly the same physical predictions. This choice of Φ spontaneously breaks the $\mathrm{SU(2)_L}$ x $\mathrm{U(1)_Y}$ into a residual $\mathrm{U(1)_{EM}}$ symmetry. The Φ kinetic term can be expanded

$$\begin{aligned} D_\mu \Phi^\dagger(x) D^\mu \Phi(x) &= |[\partial_\mu - ig' B_\mu - ig \frac{\vec{\sigma}}{2} \vec{W}_\mu] \Phi(x)|^2 \\ &= \frac{1}{2} | \begin{pmatrix} \partial_\mu - \frac{i}{2}(g W_\mu^3 + g' B_\mu) & -\frac{i}{2} g (W_\mu^1 - i W_\mu^2) \\ -\frac{i}{2} g (W_\mu^1 + i W_\mu^2) & \partial_\mu + \frac{i}{2}(g W_\mu^3 - g' B_\mu) \end{pmatrix} \begin{pmatrix} 0 \\ v + H \end{pmatrix} |^2 \\ &= \frac{1}{2}(\partial_\mu H)(\partial^\mu H) + \frac{1}{8} g^2 (v + H)^2 (W_\mu^1 + i W_\mu^2)(W_1^\mu - i W_2^\mu) \\ &+ \frac{1}{8}(v + H)^2 (g W_\mu^3 - g' B_\mu)(g W_3^\mu - g' B^\mu), \end{aligned} \tag{2.19}$$

which can be rewritten in terms of the four orthogonal fields

$$W_\mu^\pm = W_\mu^1 \mp W_\mu^2 \tag{2.20}$$

and Z_μ, A_μ as defined in equation 2.7. The resulting Φ kinetic term

$$\begin{aligned} D_\mu\Phi^\dagger(x)D^\mu\Phi(x) &= \frac{1}{2}(\partial_\mu H)(\partial^\mu H) + \frac{1}{8}g^2(v+H)^2W_\mu^- W^{+\mu} \\ &+ \frac{1}{2}(g^2+g'^2)(v+H)^2 Z_\mu Z^\mu \\ &= \frac{1}{2}(\partial_\mu H)(\partial^\mu H) + \frac{1}{8}g^2v^2W_\mu^- W^{+\mu} + \frac{1}{8}(g^2+g'^2)v^2 Z_\mu Z^\mu \\ &+ \frac{1}{4}g^2vHW_\mu^- W^{+\mu} + \frac{1}{4}(g^2+g'^2)vHZ_\mu Z^\mu \\ &+ \frac{1}{8}g^2H^2W_\mu^- W^{+\mu} + \frac{1}{8}(g^2+g'^2)H^2 Z_\mu Z^\mu \end{aligned} \tag{2.21}$$

has very important implications. The second and third terms give masses for the W and Z bosons, which was forbidden under $SU(2)_L$ x $U(1)_Y$ symmetry. Thus by introducing Φ and describing Φ in terms of excitations about a particular minimum v, the $SU(2)_L$ x $U(1)_Y$ symmetry is broken into a $U(1)_{EM}$ symmetry with massive W and Z gauge bosons. The conserved quantity for the remaining $U(1)_{EM}$ symmetry can be interpreted as the electric charge from electromagnetism. The masses of the W and Z bosons are given by

$$m_W^2 = \frac{1}{4}g^2v^2, m_Z^2 = \frac{1}{4}(g^2+g'^2)v^2. \tag{2.22}$$

Equation 2.21 also predicts interactions between the scalar particle H and the W and Z bosons. The fourth and fifth terms describe vertices with a single H and two Z bosons or a single H, one W^+, and one W^-. The sixth and seventh terms predict quartic vertices with two H's and either two Z bosons or one W^+ and one W^-. Note that the photon field A_μ does not appear in equation 2.21. Because $U(1)_{EM}$ is a preserved local symmetry, the photon remains massless and does not couple to H.

The scalar potential (Equation 2.16) can be expanded in terms of v and H(x) as

$$V(\Phi(x)) = -\frac{1}{4}\lambda v^4 + \lambda v^2H^2 + \lambda vH^3 + \frac{1}{4}\lambda H^4, \tag{2.23}$$

where the H^3 and H^4 terms describe Higgs trilinear and quartic self-interactions, respectively. The H^2 component is a Higgs boson mass term, with $m_H = \sqrt{2\lambda v^2}$.

2.4 Yukawa couplings

The process of spontaneous symmetry breaking (SSB), as described in Section 2.3, yields mass terms for the W and Z gauge bosons. The fermions, however, remain massless without further additions to the SM Lagrangian. Dirac mass terms for the fermions are forbidden by the $\mathrm{SU(2)_L}$ symmetry. The addition of the complex scalar doublet Φ allows for additional terms in the SM Lagrangian including both Φ and the fermion fields. Let Q_{Li} denote the three SU(2) left-handed quark doublets, L_{Li} the left-handed lepton SU(2) doublets, U_{Ri} the right-handed up-type quark singlets, D_{Ri} the right-handed down-type quark singlets, and E_{Ri} the right-handed lepton singlets. Generic terms can be added to the SM Lagrangian of the form

$$\mathcal{L}_{\mathrm{Yuk}} = Y^u_{ij}\overline{Q_{Li}}U_{Rj}\tilde{\Phi} + Y^d_{ij}\overline{Q_{Li}}D_{Rj}\Phi + Y^e_{ij}\overline{L_{Li}}E_{Rj}\Phi + h.c., \tag{2.24}$$

where Y^u, Y^d, and Y^e are general 3 x 3 complex matrices of dimensionless couplings. These terms preserve $\mathrm{SU(2)_L}$ x $\mathrm{U(1)_Y}$ local symmetry. They are in fact the most generic interaction terms between Φ and the fermions allowable in the SM Lagrangian. Other combinations of the above operators would not conserve either T_3 or Y and therefore break $\mathrm{SU(2)_L}$ x $\mathrm{U(1)_Y}$ symmetry or would not be renormalizable. Without loss of generality, it is possible to choose a basis such that the matrices are diagonalized

$$\begin{aligned} Y^e &\to V_{eL}Y^eV^\dagger_{eR} = \hat{Y}^e = \mathrm{diag}(y_e, y_\mu, y_\tau), \\ Y^u &\to V_{uL}Y^uV^\dagger_{uR} = \hat{Y}^u = \mathrm{diag}(y_u, y_c, y_t), \\ Y^d &\to V_{dL}Y^dV^\dagger_{dR} = \hat{Y}^d = \mathrm{diag}(y_d, y_s, y_b). \end{aligned} \tag{2.25}$$

Note that the electroweak interaction eigenstates do not in general have to be the same as the quark mass eigenstates. In particular, this means that unless $V_{uL} = V_{dL}$, the bases are in fact different. This has important implications for the electroweak interaction and implies the possibility of W boson vertices involving multiple quark generations. This quark mixing is described by the CKM matrix $V_{CKM} = V_{uL}^{\dagger} V_{dL}$.

After SSB, $\mathcal{L}_{\text{Yuk}}$ has the form

$$\begin{aligned}
\mathcal{L}_{\text{Yuk}} \supset & \frac{y_u}{\sqrt{2}} \bar{u_L} u_R (v + H) + \frac{y_d}{\sqrt{2}} \bar{d_L} d_R (v + H) + \frac{y_e}{\sqrt{2}} \bar{e_L} e_R (v + H) \\
= & \frac{v y_u}{\sqrt{2}} \bar{u_L} u_R + \frac{v y_d}{\sqrt{2}} \bar{d_L} d_R + \frac{v y_e}{\sqrt{2}} \bar{e_L} e_R \\
& + \frac{y_u}{\sqrt{2}} \bar{u_L} u_R H + \frac{y_d}{\sqrt{2}} \bar{d_L} d_R H + \frac{y_e}{\sqrt{2}} \bar{e_L} e_R H,
\end{aligned} \tag{2.26}$$

with similar terms for all three fermion generations. With the suggestive definition $m_f \equiv \frac{y_f v}{\sqrt{2}}$, $\mathcal{L}_{\text{Yuk}}$ can be written as

$$\mathcal{L}_{\text{Yuk}} = (1 + \frac{H}{v}) m_f \bar{f_L} f_R, \tag{2.27}$$

with the implied sum over all fermion types f. Thus each fermion acquires mass m_f with additional $Hf\bar{f}$ interaction vertices with coupling strength proportional to m_f. An important point to note is that the fermion masses m_f are free parameters of the theory, unlike the W and Z boson masses, which are directly predicted by the values of g, g', and v. However, once the fermion mass has been determined the coupling strength of the interaction vertex $Hf\bar{f}$ is fixed.

2.5 Standard Model Lagrangian

The electroweak theory and QCD can be unified in a single Lagrangian with SU(3)_{C} x SU(2)_{L} x U(1)_{Y} local symmetry. The leptons, which do not carry color charge,

are described as $SU(3)_C$ singlets and the left-handed quarks are considered triplets in $SU(3)_C$ and doublets in $SU(2)_L$. Table 2.2 summarizes the particle constituents of the SM, including a summary of the notation used to denote each field representation. The full SM Lagrangian can be written as

$$\begin{aligned}
\mathcal{L}_{SM} &= -\frac{1}{4}G_a^{\mu\nu}G_{a\mu\nu} - \frac{1}{4}W_b^{\mu\nu}W_{b\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \\
&+ (D^\mu\Phi)^\dagger(D_\mu\Phi) + i\overline{Q_{Li}}\gamma^\mu D_\mu Q_{Li} + i\overline{U_{Ri}}\gamma^\mu D_\mu U_{Ri} \\
&+ i\overline{D_{Ri}}\gamma^\mu D_\mu D_{Ri} + i\overline{L_{Li}}\gamma^\mu D_\mu L_{Li} + i\overline{E_{Ri}}\gamma^\mu D_\mu E_{Ri} \\
&+ Y_{ij}^u\overline{Q_{Li}}U_{Rij}\tilde{\Phi} + Y_{ij}^d\overline{Q_{Li}}D_{Rij}\Phi + Y_{ij}^e\overline{L_{Li}}E_{Ri}\Phi + h.c. \\
&- \lambda(\Phi^\dagger\Phi - \frac{v^2}{2})^2,
\end{aligned} \tag{2.28}$$

where the first line describes the kinetic term for the gauge fields and the second and third lines describe the kinetic term for the fermion fields. The fourth line describes the Yukawa interactions, as discussed in Section 2.4. Finally, the fifth line describes the scalar field potential. The Lagrangian in Equation 2.28 is invariant under local $SU(3)_C$ x $SU(2)_L$ x $U(1)_Y$ transformations provided that the covariant derivative is defined as

$$D^\mu = \partial^\mu + ig_sG_a^\mu L_a + igW_b^\mu T_b + ig'YB^\mu, \tag{2.29}$$

where $L_a = \frac{1}{2}\lambda_a$ (0) for $SU(3)_C$ triplets (singlets), $T_b = \frac{1}{2}\sigma_b$ (0) for $SU(2)_L$ doublets (singlets), and Y is the hypercharge of the field as given in Table 2.2. Although simple in form, this Lagrangian fully describes the interactions between all SM particles.

Table 2.2: Summary of particle fields in the SM.

	symbol	T_3	Y	Q	$SU(2)_L$ rep.	$SU(3)_C$ rep.
Fermions						
e_L, μ_L, τ_L	L_L	-1/2	-1/2	-1	doublet	singlet
$\nu_{eL}, \nu_{\mu L}, \nu_{\tau L}$		+1/2	-1/2	0		
e_R, μ_R, τ_R	E_R	0	-1	-1	singlet	singlet
u_L, c_L, t_L	Q_L	+1/2	+1/6	+2/3	doublet	triplet
d_L, s_L, b_L		-1/2	+1/6	-1/3		
u_R, c_R, t_R	U_R	0	+2/3	+2/3	singlet	triplet
d_R, s_R, b_R	D_R	0	-1/3	-1/3	singlet	triplet
Gauge Bosons						
W^+	W^+_μ	+1	0	+1		
W^-	W^-_μ	-1	0	-1		
Z	Z_μ	0	0	0		
photon	A_μ	0	0	0		
gluon (eight gluons)	G_μ	0	0	0		
Higgs	Φ or H	-1/2	+1/2	0		

2.6 Physics beyond the Standard Model

The SM as described in Section 2.5 has 18 free parameters: the three charged lepton masses, g, g', m_H, v, the six quark masses, g_s, and three angles in the quark mixing matrix V_{CKM} plus one complex phase. Once these parameters have been measured, the SM can be experimentally tested by measuring any physical observable, which in the SM can be calculated as a function of the given parameters. The SM has been remarkably successful in predicting the observed experimental data for an enormous range of high precision measurements. One (of very many) examples is the ability of the SM to precisely predict the couplings and production cross sections of the newly discovered Higgs boson now that m_H has been measured experimentally.

There are, however, phenomena observed in Nature that the SM cannot explain. It is well established from astrophysical measurements that most of the mass in the

universe is in the form of dark matter, which has so far only been observed to interact with SM particles via the gravitational force. The SM, moreover, does not incorporate gravity, which is many orders of magnitude weaker than the electroweak and strong forces. Additionally, the Higgs boson mass is subject to quantum loop corrections which diverge as a function of the energy scale cutoff, Λ. This divergence is not present for other SM particles, which as fermions or gauge bosons always have opposite correction terms which cancel these divergences.

It is therefore clear that the SM is not a full description of Nature, but rather that it is most likely an effective theory valid up to some energy scale Λ at which additional particles and potentially forces are manifested that are not described by the SM. The Higgs boson is in particular an excellent probe with which to test the SM due to its universal role in SSB and its subsequent couplings to all massive particles. The remarkable progress in measuring the properties of the Higgs boson in the seven years since its discovery is a testament to the impressive performance of the LHC and its experiments. There are, however, many remaining possible extensions of the SM involving the Higgs sector including composite Higgs bosons, multiple Higgs bosons, the Higgs boson as a portal to dark matter, or the Higgs boson as a Goldstone boson of an additional fundamental symmetry that is broken at a high energy scale Λ. All these possible models can affect the Higgs boson couplings to SM particles. Despite the significant experimental challenges, measuring the first Yukawa coupling to down-type quarks via Higgs boson decay to bottom quarks is therefore an important test of the SM description of the Higgs sector.

Chapter 3

Experimental history

3.1 LEP

3.1.1 LEP overview

The Large Electron-Positron Collider (LEP) was a circular collider built at CERN in the same tunnel used by the LHC today (described in Section 4.1). It remains, to date, the highest energy lepton collider ever built. LEP was used between 1989 and 2000, then dismantled to allow for construction of the LHC. After several years of operation on the Z boson resonance at beam energies of about 45 GeV per beam, the beam energy was increased in gradual steps to over 100 GeV. The maximum center-of-mass energy (twice the beam energy for a circular energy-symmetric collider) achieved towards the end of LEP operation was 209 GeV. As described in Chapter 2, the Higgs boson mass is a free parameter in the SM. Higgs boson searches at LEP therefore considered a large range of potential m_{H} hypotheses up to 115 GeV, after which the Higgs boson production cross section is highly suppressed at LEP beam energies. The datasets collected at LEP allowed for highly precise measurements of the SM electroweak sector and the exclusion of a Higgs boson with a mass less than 115 GeV at 95% confidence level.

3.1.2 $H \to b\bar{b}$ at LEP

At LEP the primary expected Higgs boson production mechanism was through the "Higgsstrahlung" process $e^+e^- \to ZH$, shown at tree level in Figure 3.1.

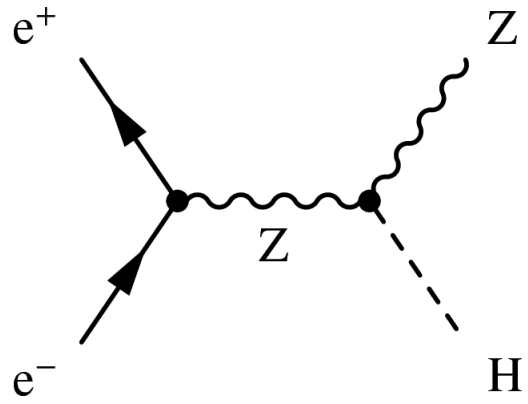


Figure 3.1: Leading order Feynmann diagram for the "Higgstrahlung" process, the dominant Higgs boson production mode at LEP.

The SM Higgs boson is predicted to decay to $b\bar{b}$ with a branching fraction of 74% for $m_H = 115$ GeV. For $m_H < 115$ GeV, the Higgs branching fraction to $b\bar{b}$ is even greater. In addition, the background from production of final states with multiple b quarks is very small at electron-positron colliders. The search for the Higgs boson at LEP therefore focused primarily on $H \to b\bar{b}$ candidates produced in association with a Z boson. Independent Higgs boson searches were performed targeting each of the Z boson decay channels. Similar search categories would later be considered at the Tevatron and at the LHC, however these searches were subject to much larger backgrounds due to the abundance of multijet events produced at hadron colliders.

3.1.3 LEP Higgs boson combination

Dedicated Higgs boson searches by each of the four LEP experiments, ALEPH, DELPHI, L3, and OPAL, using the full LEP datasets, were combined in 2003 [1]. The input from each experiment was the observed number of data events as well as the

expected signal and background contributions in selected bins of the reconstructed Higgs boson candidate mass, m_H^{rec}, and a global discriminating variable $\mathcal{G}$ which combined many event features including b-tagging variables, likelihood functions, and neural network outputs. Similar analysis techniques would be later adopted at the Tevatron and the LHC. A set of Higgs boson mass hypotheses was considered ranging up to 115 GeV, the extent of the LEP kinematic reach. Figure 3.2 shows the combined m_H^{rec} distribution for data as well as the expected background and expected contribution from a SM Higgs boson signal with $m_H = 115$ GeV.

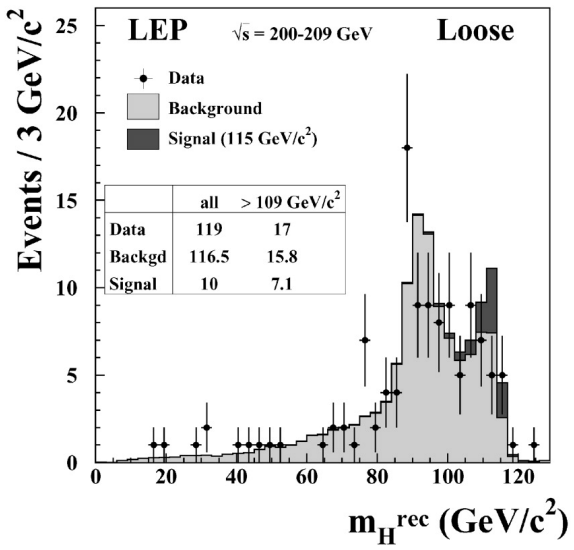


Figure 3.2: Reconstructed Higgs boson candidate mass in a region of intermediate signal purity for the combination of LEP searches [1].

The combined result was interpreted in terms of a likelihood comparison between the background-only and the signal plus background hypothesis as a function of the assumed m_H. No statistically significant excess of events over the background was observed, and a 95% confidence level (CL) lower limit was set on m_H of 114.4 GeV.

Excess in ALEPH data

Due to an observed excess in the ALEPH data during the last year of data taking, the LEP experiments requested an extension of the LEP program for six months [21].

The statistical significance of the excess with respect to the background-only hypothesis was about three standard deviations and compatible with a Higgs boson mass of roughly 115 GeV. The request, however, was denied in order to not delay the construction of the LHC, which would be built in the same tunnel as LEP. A Higgs boson with mass near 115 GeV was later strongly excluded at the LHC.

3.1.4 Constraints on m_H from global electroweak fit

The high-precision measurements of SM electroweak parameters achieved at LEP allowed for indirect constraints on m_H and m_t through expected loop corrections [2]. Since the leading m_t dependence is quadratic while the leading m_H dependence is logarithmic, the indirect constraints on m_H were much weaker than those on m_t. A global fit was performed to LEP data with m_H and m_t considered as free parameters. Figure 3.3 shows the resulting constraints on m_H and m_t compared with the already excluded m_H range and the direct measurement of m_t achieved at the Tevatron (Sec. 3.2). Good agreement was observed for m_t between the direct measurement and the indirect constraints. An additional fit was performed with only m_H as a free parameter, which yielded the indirect constraint $m_H < 193$ GeV at 95% CL. The allowed range for m_H at 95% CL was therefore $[114.4, 193]$ GeV. Later direct searches at the Tevatron further excluded the range $[158, 175]$ GeV [22], leaving a relatively small window of potential m_H values at the time of the first significant LHC datasets in 2010.

3.2 $H \to b\bar{b}$ at the Tevatron

The Tevatron was a circular proton-antiproton collider built at the Fermi National Accelerator Labratory, located near Batavia, Illinois. At the time of its operation, the Tevatron was the highest energy particle collider ever built. The Tevatron ring

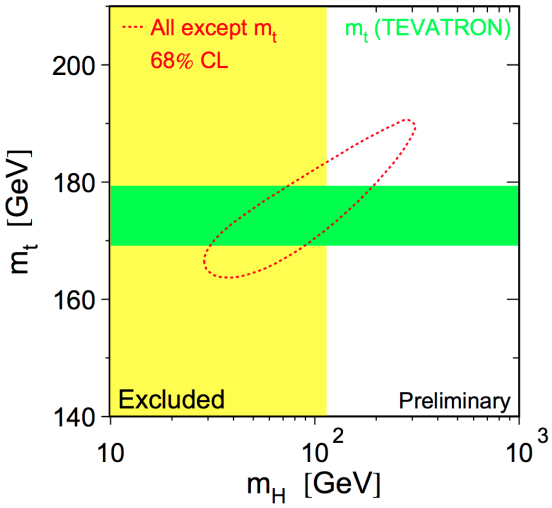


Figure 3.3: The 68% confidence level contour (dashed red) in m_H and m_t obtained from the fit to LEP data. The yellow shaded area corresponds to regions of m_H that had already been excluded, while the green shaded area shows the m_t range determined by direct measurement at the Tevatron [2].

was 6.28 km in circumference and ran at increasing center-of-mass energies up to $\sqrt{s}$ = 2.0 TeV with an instantaneous luminosity up to 4 x $10^{32}cm^{-2}s^{-1}$. Two detectors were installed in the Tevatron ring, CDF and DØ. The Tevatron is well known for the joint discovery of the top quark by the CDF and DØ collaborations in 1995 [23, 24] and subsequent measurement of the top quark mass to a precision of nearly 1%. The Tevatron also provided a rich dataset for novel results in flavor physics including the first measurement of B_s oscillations by the CDF Collaboration in 2006 [25]. Exclusion limits on the Higgs boson mass were tightened with respect to those obtained at LEP. For the lower mass scenarios, the most sensitive Higgs boson search channel was $H \rightarrow b\bar{b}$, where the Higgs boson is produced in association with a W or Z boson. The primary Higgs boson production modes at hadron colliders will be described in Section 4.1.2. The results of the final Higgs boson search at the Tevatron using the full dataset collected at $\sqrt{s}$ = 2.0 TeV were released in July 2012, one week before the announcement of the Higgs boson discovery at the LHC. Figure 3.4 shows the Higgs boson candidate invariant mass with nonresonant backgrounds subtracted for

the sum of all CDF and DØ search channels [3]. An excess in the data over the background-only hypothesis was observed with a maximal global significance of 3.1 standard deviations for $m_{\mathrm{H}} = 135$ GeV. The measured $\mathrm{H} \to \mathrm{b\bar{b}}$ rate was twice the prediction of the SM for a Higgs boson with mass in this range, although consistent with the SM within uncertainties. For a SM Higgs boson with $m_H = 125$ GeV, the observed significance was 2.8 standard deviations. Measurement of $\mathrm{H} \to \mathrm{b\bar{b}}$ at the LHC following the Higgs boson discovery, however, remained a difficult challenge due to very large background rates.

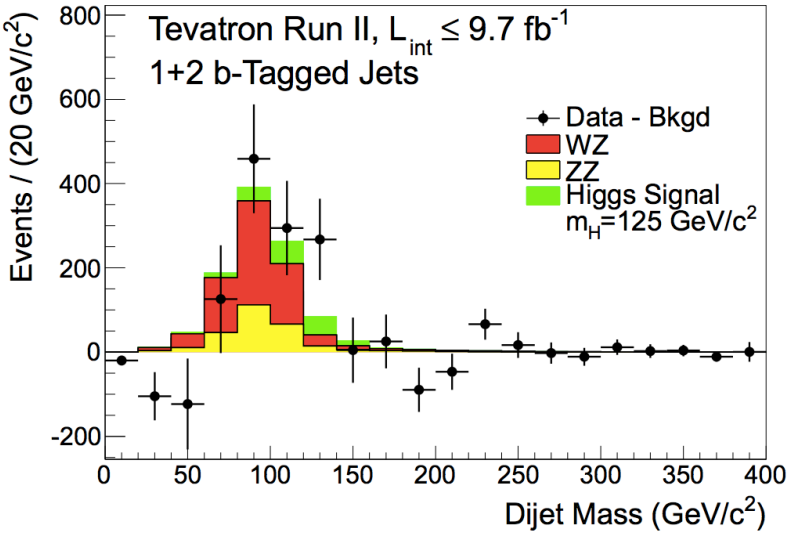


Figure 3.4: Reconstructed Higgs boson candidate mass with the nonresonant backgrounds subtracted for the combination of all CDF and DØ input channels. The expectation for a SM Higgs boson with mass 125 GeV is shown in light green [3].

Chapter 4

Experimental apparatus

4.1 The Large Hadron Collider

The LHC is a particle accelerator at CERN (European Organization for Nuclear Research) designed to collide protons at a center-of-mass energy of $\sqrt{s}$ = 14 TeV with an instantaneous luminosity of $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$ as well as lead ions at $\sqrt{s}$ = 5.52 TeV at an instantaneous luminosity of $\mathcal{L} = 10^{27}\text{cm}^{-2}\text{s}^{-1}$. The LHC tunnel is 100 m underground and 26.7 km in circumference, traversing the border between Switzerland and France near Geneva, Switzerland. As described in Section 3.1.2, the same tunnel was previously used for the LEP collider. The LHC was first proposed in 1984 and began operations in 2008.

Figure 4.1 shows a schematic of the various stages of the CERN accelerator complex, of which the LHC is the final stage. Hydrogen atoms from H_2 gas are first stripped of electrons, then the protons are accelerated to 50 MeV in the linear accelerator LINAC 2 over a distance of 33 m. The proton beam is then injected into the PS booster, the first synchrotron in the acceleration chain, which consists of four superimposed rings of circumference 157 m. In the PS booster, the proton beam is accelerated from 50 MeV to 1.4 GeV over 1.2 seconds. The beam is next injected

into the Proton-Synchrotron (PS), which has a circumference exactly four times that of the PS booster at 628 m. The PS accelerates the protons from 1.4 GeV to 26 GeV over 3.6 seconds. This cycle is repeated over four separate injections from the PS booster such that the full PS is filled with proton bunches. The beam is then injected into the Super Proton Synchrotron (SPS), the first underground synchrotron in the acceleration sequence (30 m underground). The SPS is eleven times the circumference of the PS, or 6.9 km, and accelerates the protons to 450 GeV. The beam is then injected into the LHC tunnel.

The LHC consists of 1232 superconducting dipole magnets of length 15 m and field strength of 8.33 T, which bend the proton beam around the LHC circumference. Quadrupole magnets are used to focus the proton beams, with a total of 858 installed between the dipole bending magnets. An additional 6000 corrector magnets are installed to make adjustments to preserve the beam quality. Two separate beams circulate in opposite directions which are then focused and crossed at the four active collision points in the LHC. The ATLAS, CMS, LHCb, and ALICE detectors are each situated underground at one of the LHC collision points. ATLAS and CMS are general-purpose hermetic detectors primarily designed to discover the Higgs boson, perform precision electroweak measurements, and search for BSM particles. LHCb is a forward detector dedicated to precision B physics measurements, and ALICE is designed for the study of heavy ion collisions.

4.1.1 LHC operations

After more than a decade of construction and installation, the LHC began operation on September 10, 2008. Unfortunately a magnet quench incident caused extensive damage to over 50 superconducting magnets, their mountings, and the vacuum pipe. It required 14 months to fully repair the damage, such that the LHC did not resume operations again until November 2009. After several months of testing and commis-

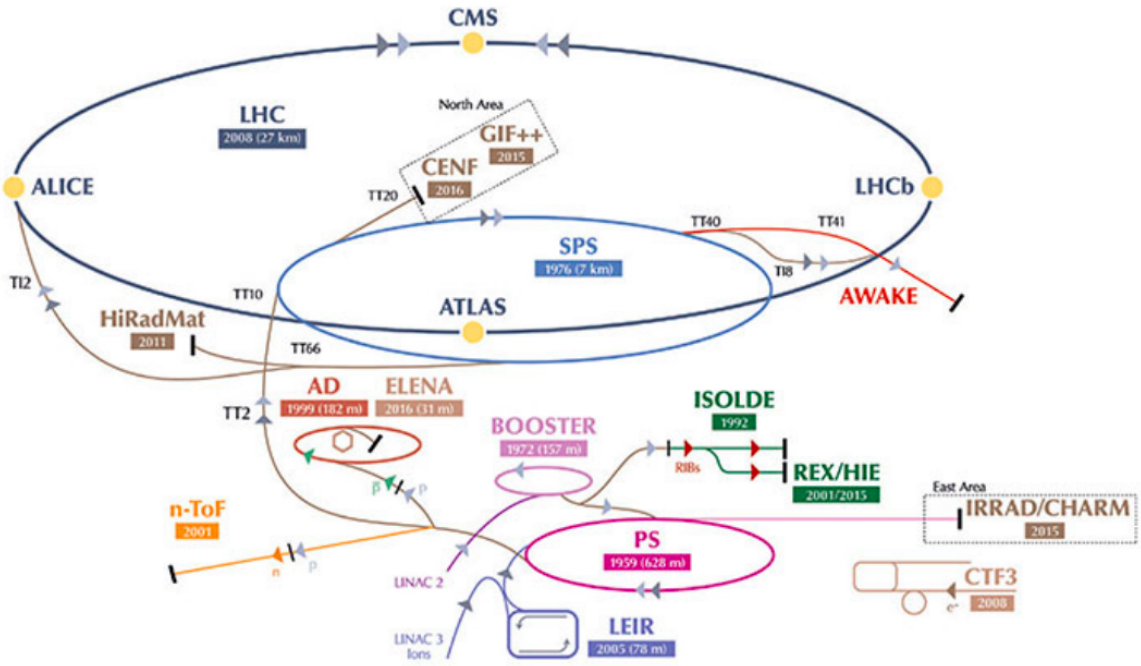


Figure 4.1: Schematic view of the LHC accelerator complex [4].

sioning at lower energy, the LHC began the first high energy collisions on March 30, 2010, thus beginning "Run-1" of the LHC. It was decided to run at a center-of-mass energy lower than the design energy in order to protect the magnets. About 45 pb^{-1} and 6 fb^{-1} of proton-proton collision data was collected by CMS in 2010 and 2011, respectively, at a center-of-mass energy of $\sqrt{s} = 7$ TeV. In 2012 an additional 23 fb^{-1} was collected at $\sqrt{s} = 8$ TeV. The discovery of the Higgs boson in July 2012 was achieved using the 2011 dataset combined with 5.3 fb^{-1} of 2012 data at $\sqrt{s} = 8$ TeV [26, 27].

After three years of successful operations, the LHC began Long Shutdown 1 (LS1) and halted operations for two years. During this time, the LHC magnets were trained to withstand higher currents in preparation for colliding beams at higher energy. "Run-2" of the LHC began in spring 2015 at $\sqrt{s} = 13$ TeV and continued through November 2018. During this period, a total of 160 fb^{-1} of proton-proton collision

data was delivered to CMS, roughly five times the dataset collected in Run-1 and at nearly twice the center-of-mass energy. Figure 4.2 summarizes the datasets delivered to CMS by year. This thesis will focus on the dataset collected in 2017, which will be described in further detail in Section 7.4.

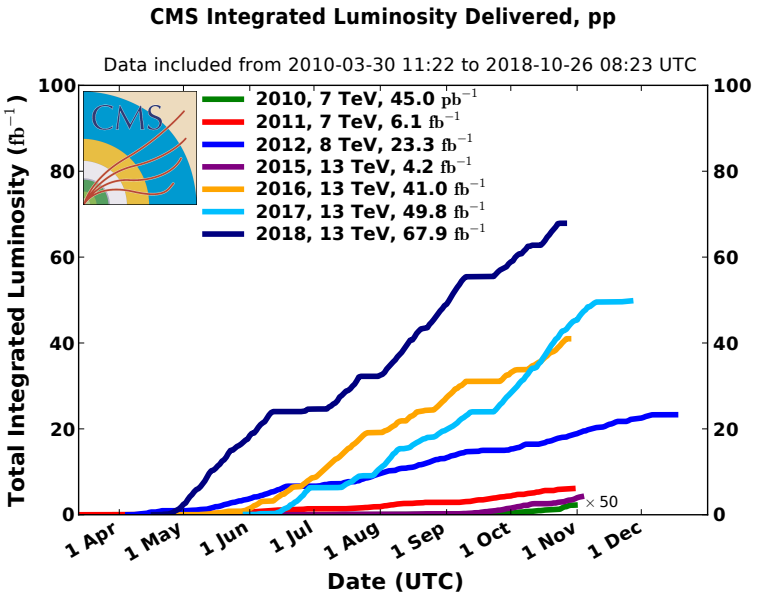


Figure 4.2: Integrated luminosity delivered to CMS for proton-proton collisions, split by year.

4.1.2 Higgs boson production at the LHC

At the LHC, the dominant production mode for a Higgs boson with mass $m_{\mathrm{H}} = 125$ GeV is gluon fusion (ggH), where two initial-state gluons produce the Higgs boson via a virtual fermion loop. Figure 4.3 (left) shows the tree-level Feynmann diagram for this process. The main contribution is from top quarks in the fermion loop because the matrix element is proportional to the mass of the fermion in the loop. The overall predicted cross section for gluon fusion production of a Higgs boson with $m_H = 125$ GeV is 43.92 pb at $\sqrt{\mathrm{s}} = 13$ TeV, in agreement with experimental measurements within the current precision of 10-15%. The Higgs boson discovery

primarily considered Higgs bosons produced via ggH production and decaying to the experimentally very clean final states H$\to \gamma\gamma$ and H$\to$ZZ$^*\to 4\ell$. Although ggH has the largest cross section of the Higgs boson production modes at the LHC, it is a difficult production mode for Higgs measurements without experimentally distinctive Higgs boson decay products such as H $\to$ b$\bar{\text{b}}$. This is due to an enormous background of events with multiple quarks produced via strong interactions.

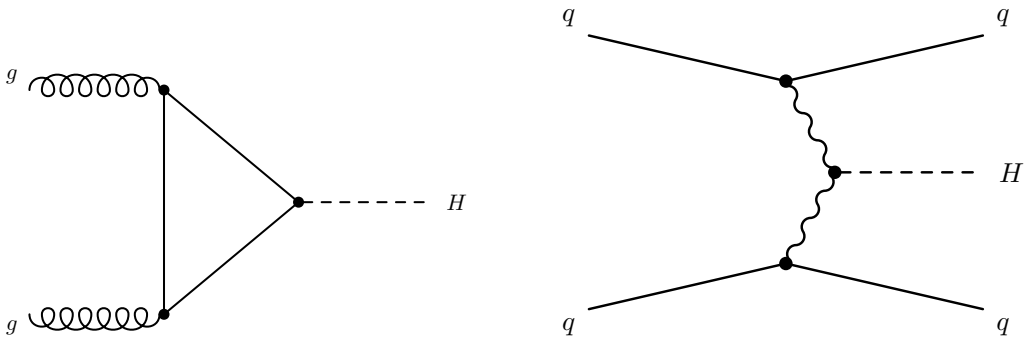


Figure 4.3: Tree-level Feynmann diagrams for the two dominant Higgs boson production modes at the LHC, gluon fusion (left), and vector boson fusion (right).

The sub-leading production mode of Higgs bosons at the LHC is vector boson fusion (VBF), where two initial-state quarks produce a Higgs boson via the exchange of two virtual W or Z bosons. Figure 4.3 (right) shows the VBF tree-level diagram. A characteristic feature of VBF production is the presence of two final-state quarks, in addition to the Higgs boson decay products, which tend to have large angular separation as well as a large dijet invariant mass with respect to other SM processes. Despite the lower cross section (3.75 pb), the presence of the two additional quark jets in the final state yields a much more experimentally distinct signature than gluon fusion production. For H $\to$ b$\bar{\text{b}}$ searches, however, the resulting four-jet final state still has large challenging multijet backgrounds and is difficult to identify with reasonable efficiency during data taking while maintaining an acceptable rate of recorded events.

With a cross section of 2.25 pb, the production of a Higgs boson in association with a W or Z boson (VH) is a relatively small component of the overall number

of Higgs bosons produced at the LHC. Figure 4.4 shows the tree level diagram for VH production. Note that for ZH about 12% of the production cross section comes from diagrams with gluons in the initial state (ggZH) rather than quarks. Despite the relatively small cross section for VH, the additional presence of the W or Z boson provides a very distinct experimental signature. In the search for $H \to b\bar{b}$ at the LHC, where backgrounds with multiple b jets are produced with a cross section seven to nine orders of magnitude larger than Higgs boson production, VH is by far the most effective search channel. This is due to the large reduction in background rate achieved by requiring a leptonically decaying W or Z boson in the final state in addition to the Higgs boson candidate. This strategy will be discussed in more detail in Section 7.1.

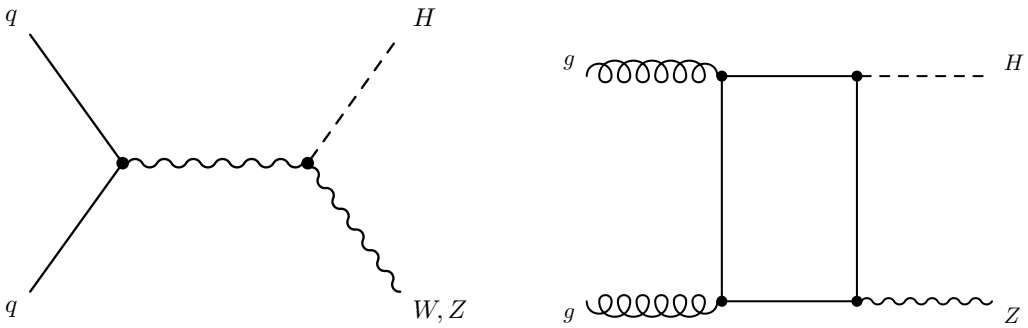


Figure 4.4: Tree-level Feynmann diagrams for the production of a Higgs boson in association with a W or Z boson with initial state quarks (left) and with initial state gluons (ggZH, right).

Another significant production mode of the Higgs boson at the LHC is in association with a top quark-antiquark pair (ttH). Figure 4.5 shows the tree-level Feynmann diagram for ttH production. The relatively small cross section (0.509 pb) and the presence of many particles in the final state makes ttH measurements experimentally very difficult at the LHC. Since the decay $H \to t\bar{t}$ is kinematically forbidden ($m_t > \frac{m_H}{2}$), the best way to directly measure the Higgs Yukawa coupling to top quarks at the LHC is via measurement of ttH production. This process was recently observed by the CMS and ATLAS Collaborations [10, 11].

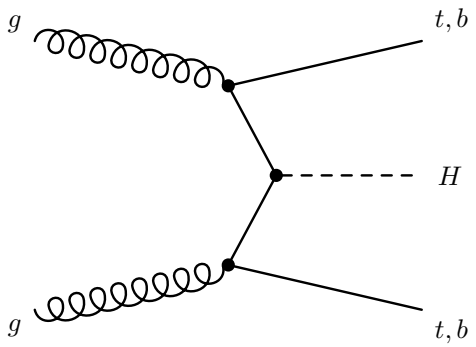


Figure 4.5: Tree-level Feynmann diagram for the production of a Higgs boson in association with a top quark-antiquark pair (ttH).

4.2 The CMS experiment

4.2.1 Overview

The Compact Muon Solenoid (CMS) detector is situated at one of the four collision points in the LHC, about 120 m underground, near Cessy, France. The complete technical proposal for CMS was completed in 1994. The final assembly of the CMS detector took place on the ground surface. CMS was then lowered underground into the experimental cavern. Figure 4.6 shows a schematic view of the CMS detector. The central feature of CMS is a 3.8 T superconducting solenoidal magnet, which makes it possible to precisely measure the momentum of charged particles due to the resulting curved particle trajectories. Inside the solenoidal magnet and closest to the beam pipe is a silicon tracker, surrounded by an electromagnetic calorimeter (ECAL) of lead tungstate crystal followed by a sampling hadronic calorimeter (HCAL) of brass scintillator. The solenoidal magnet is surrounded by a large iron return yoke, inside of which are situated gas-ionization chambers for muon detection. This section will describe each of these detector subsystems individually.

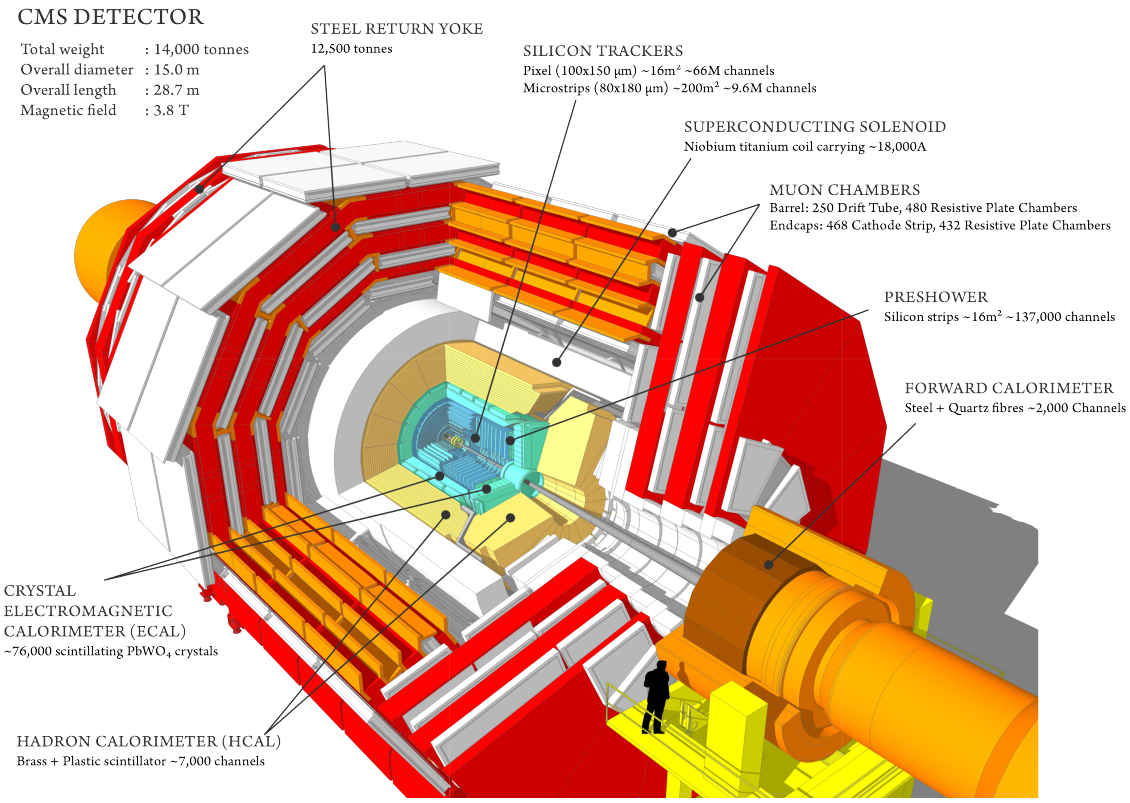


Figure 4.6: Schematic overview of the CMS detector.

4.2.2 CMS coordinate system

The convention used in CMS is to define a right-handed Cartesian coordinate system with the z-axis following the beam axis and the x-axis pointing perpendicular to the beam axis towards the center of the LHC ring. The azimuthal angle ϕ is then defined as in a usual polar coordinate system as $\phi = \tan(\frac{y}{x})$. Rather than the polar angle θ with respect to the z-axis, it is preferred to define the quantity

$$\eta = -\ln(\tan(\frac{\theta}{2})), \tag{4.1}$$

with η referred to as the "pseudorapidity". The pseudorapidity is preferred over the polar angle θ because the production of low-momentum hadrons in proton collisions is roughly constant as a function of η. Furthermore, the difference in η between two

particles is Lorentz invariant under the approximation that the masses of the particles are much less than the particle energies.

A common metric of angular distance between two particle trajectories is ΔR, defined as

$$\Delta R = \sqrt{\Delta\phi^2 - \Delta\eta^2}, \tag{4.2}$$

which for a given ΔR defines a fixed-size cone with the axis aligned with the particle momentum direction.

Because the LHC collides protons, which are not fundamental particles but rather bound states of quarks, it is not possible to know precisely for a given event the collision momentum in the z direction. The primary momentum observable of interest is therefore the particle momentum in the x-y plane transverse to the beam line, referred to as the transverse momentum p_{T}.

4.2.3 Solenoidal magnet

The CMS superconducting solenoidal magnet is 12.5 m long with a diameter of 6 m, making it the largest superconducting magnet in the world. The magnet coils are made from niobium-titanium that is mechanically reinforced with an aluminum alloy and placed within a cryostat at an operational temperature of 4.5 K, which allows the magnet to maintain the superconductivity necessary to deliver the high magnetic field with no resistive energy loss. The magnet provides a nearly uniform magnetic field of strength 3.8T within its volume, which is crucial in order to resolve with high resolution the momentum of the high energy charged particles which are produced in the collisions. The magnetic field is returned via an iron yoke, which is interleaved with muon detection systems as described in Section 4.2.7.

4.2.4 Silicon tracker

The CMS tracking system is the largest silicon detector ever built. It is the part of the CMS detector closest to the beamline, with the innermost layer 2.9 cm from the interaction point. The outer tracker uses p-n type silicon strip sensors with a total of 207 m^2 of active silicon and 9.6 million channels, with a length ranging from 10 cm to 20 cm in the outermost layers. The granularity of the strip sensors ranges from 20 μm to 50 μm in the radial direction and from 200 to 500 μm in the longitudinal direction, with the spatial resolution reducing with increasing r. A slight angle (pitch) is set between strip sensors, which allows for an improvement in the ability to resolve ambiguities in the incident particle hit positions. The pitch of the strip sensors ranges from 80 to 205 μm at the outermost layers.

The innermost tracker layers contain 66 million pixel sensors covering a surface area of 1 m^2. The spatial resolution of each pixel sensor is 10 μm in the radial direction and 20 μm in the z direction. The much higher granularity of the pixel sensors in the azimuthal direction leads to some improvement in the track resolution, however the main advantage in using the highly granular pixels close to the beam line is a large improvement in reducing hit ambiguities in the highly track dense LHC collisions environment.

Figure 4.7 shows a schematic view of the CMS tracker as it was initially installed. The pixel tracker consists of three barrel layers and four endcaps at both $+z$ and $-z$. The strip tracker is partitioned into four regions, namely the tracker inner barrel (TIB), tracker outer barrel (TOB), tracker inner detector (TID), and tracker endcaps (TEC+/TEC-). Because the particle occupancy decreases as $1/r^2$, the detector granularity tends to decrease with increasing r. The tracker geometry requires track $|\eta| < 2.5$.

The CMS tracker is subjected to very high radiation doses, especially in the pixel detector layers, due to its very close proximity to the interaction point. The large

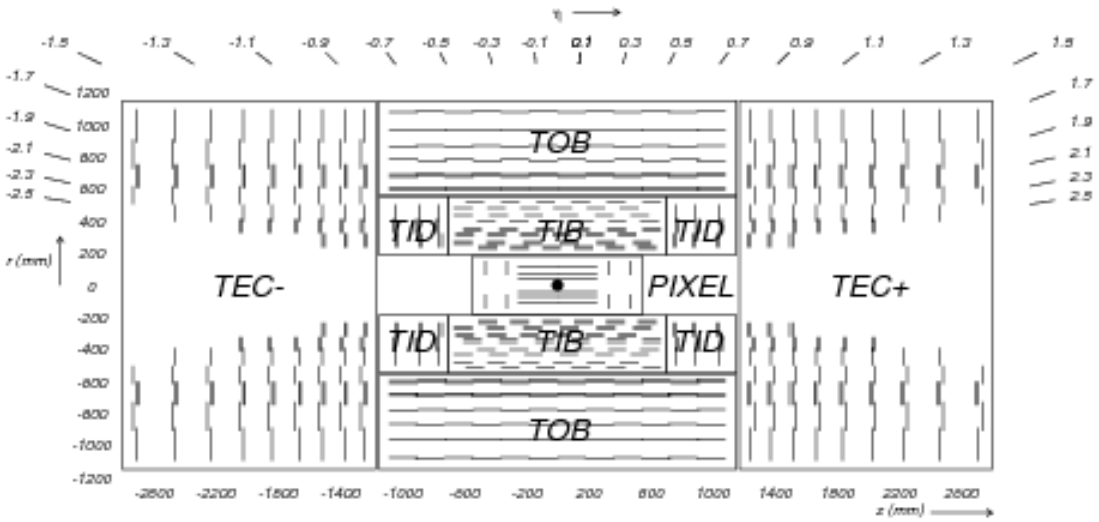


Figure 4.7: Schematic view of the CMS tracker in the r-z plane [5].

influx of massive particles traversing the silicon over time leads to degradation in the signal to noise of the tracker sensors as well as an increase in hit reconstruction inefficiencies. The pixel detector, which had already received significant radiation damage after collecting collision data since 2010, was fully replaced at the end of 2016. Figure 4.8 shows a view of the upgraded pixel detector in the r-z plane, compared with the original pixel detector. The upgraded pixel detector in particular contains four barrel layers as opposed to three, with the innermost layer about 1 cm closer to the beamline than before.

4.2.5 Electromagnetic calorimeter

The CMS electromagnetic calorimeter (ECAL) is a homogeneous calorimeter made of scintillating lead tungstate ($PbWO_4$) crystals. The ECAL is designed to induce electromagnetic showers from electrons and photons. The particles from the shower produce scintillation light proportional to the incident particle energy. By precisely measuring the overall scintillation light produced by the shower, the ECAL measures

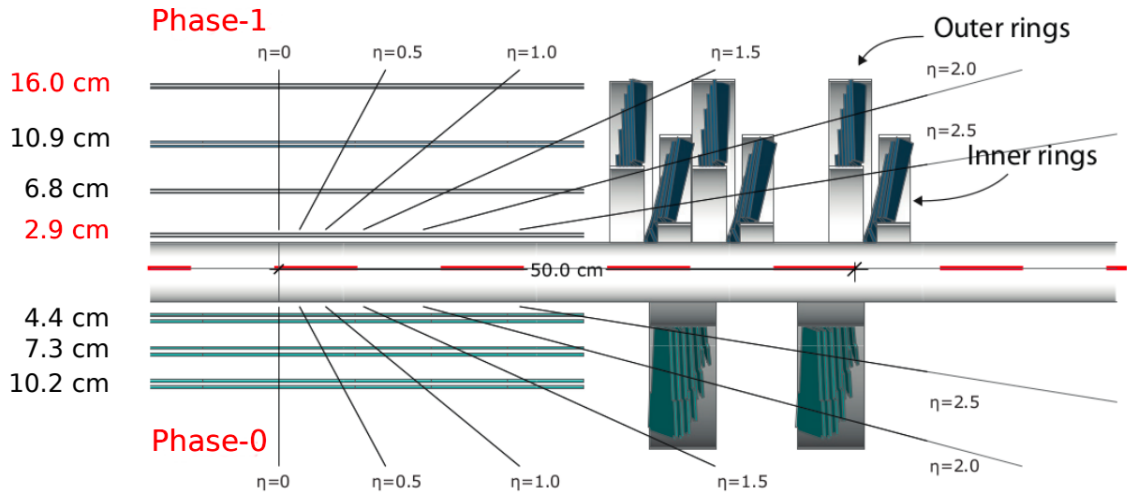


Figure 4.8: Schematic view of the upgraded CMS pixel detector (top) compared with the original pixel detector (bottom) [6].

the energy of the incident electron or photon. The choice of $PBWO_4$ was motivated by its high density, relatively fast light yield, and radiation hardness. The lead tungstate crystals have a radiation length of $X_0 = 8.9\,\text{mm}$ and a Molière radius of 22 mm. This ensures that the electromagnetic shower is typically fully contained within the ECAL crystals, which are 25 X_0 in length. 80% of the light emitted by an electromagnetic shower in the ECAL is emitted within 25 ns, the amount of time between LHC proton bunch crossings. The $PBWO_4$ serves as both an absorbing and a scintillation material, which enables excellent energy resolution.

Figure 4.9 shows a view of the ECAL in the y-z plane. The ECAL consists of two primary detectors, the ECAL barrel (EB), which covers the range $|\eta| < 1.479$, and the ECAL endcap (EE), which covers $1.653 < |\eta| < 3.0$. The EB consists of 61,200 crystals 23 cm in length with a transversal size of 22 mm × 22 mm. Each piece of the EE ($+z$ and $-z$) consists of 7,324 crystals 22 cm in length with a transversal size of 28.62 mm × 28.62 mm. The crystals in the EB are organized into 36 "supermodules" which each cover 20 degrees in ϕ, whereas crystals in the EE are grouped into two semicircular "dees". In both the EB and the EE, the crystals are oriented at an angle

of 3 degrees relative to the collision point in order to ensure that particles do not escape detection by traversing through the small gaps between crystals.

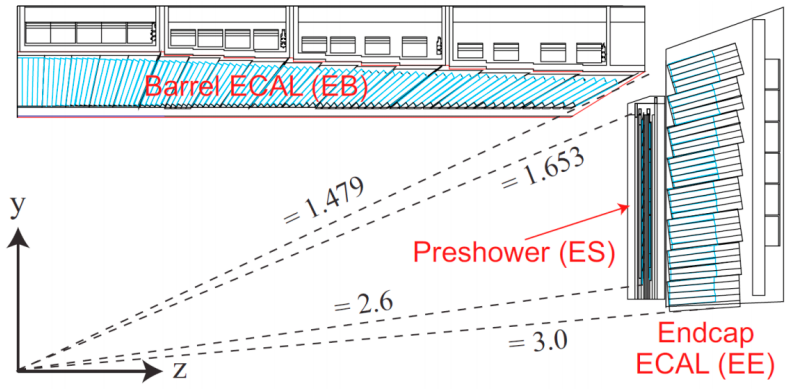


Figure 4.9: View of the CMS electromagnetic calorimeter in the y-z plane.

An additional sampling preshower (ES) detector is installed just before the EB and covering the range $1.653 < |\eta| < 2.6$. The ES consists of two layers of lead absorber which initiates the elecromagnetic shower, followed by 2 mm-wide silicon strips which measure the deposited energy and transverse profile of the shower. The ES enables improved differentiation between photons from the hard interaction and photons from neutral pion decays ($\pi_0 \to \gamma\gamma$).

The scintillation light yield of $PBWO_4$ is relatively low and strongly temperature dependent, meaning that effective operation of the ECAL requires very precise temperature control and careful calibration. The scintillation light is amplified by silicon avalanche photodiodes (APDs) in the EB and by vacuum phototriodes in the EE. The ECAL temperature is maintained at 18°C with a precision of 0.05°C in the EB and 0.1°C in the EE.

The energy resolution of calorimeters is generally given in the form

$$(\frac{\sigma}{E})^2 = (\frac{S}{\sqrt{E}})^2 + (\frac{N}{E})^2 + C^2, \tag{4.3}$$

where S, N, and C denote the stochastic, noise, and constant terms, respectively. The stochastic term is related to the number of scintillation photons n, which is proportional to E. The stochastic sampling resolution scales with $\sqrt{n}$ and therefore with $\sqrt{E}$. N is related to the detector noise, and C arises from detector inhomogeneities. These constants have been measured for the ECAL in test beam studies with incident electrons [28], giving the values S = 2.8%, N = 12%, and C = 0.3%.

4.2.6 Hadronic calorimeter

The CMS hadronic calorimeter (HCAL) is a sampling calorimeter with alternating layers of brass absorber and plastic scintillation material. The primary hadrons that have sufficiently long lifetimes to traverse the CMS calorimetry are pions, kaons, protons, and neutrons. These hadrons traverse the ECAL quite transparently, but form complex hadronic showers in the brass absorber. Whereas the electromagnetic showers are a comparatively simple cascade of photon conversions $\gamma \rightarrow e^+e^-$ and Bremsstrahlung radiation $e \rightarrow e + \gamma$, hadronic showers proceed through an increasing number of primarily strong interactions with many particle types including electromagnetic components via neutral pion decays $\pi_0 \rightarrow \gamma\gamma$. The fraction of the hadronic shower energy transferred to an electromagnetic cascade depends on the shower energy, with the fraction about 50% for a 100 GeV shower and 70% for a 1 TeV shower. The energy reconstruction efficiency is generally different for the hadronic and electromagnetic shower components, in particular because the timescale of the electromagnetic shower is much longer, up to 1 μs, and therefore not possible to fully reconstruct at the LHC collision rate of 40 MHz. It is therefore very important to precisely control the energy response with respect to the incident hadron p_{T} in order to maintain good precision.

Figure 4.10 shows a view of the HCAL in the y-z plane. The HCAL is composed of a barrel (HB) and endcap (HE) component, which are both contained inside the

solenoidal magnet, and the outer (HO) and forward (HF), which are located outside the solenoid. The HB covers the range $|\eta| < 1.3$ while the HE covers $1.3 < |\eta| < 3.0$. Due to the limited space available for the HCAL within the solenoid, the HO is included outside the solenoid in order to increase the total interaction length. The HO extends the total interaction depth to about eleven times the average interaction length of hadrons in the calorimeter. The leading contribution to the HCAL energy resolution is however still due to effects from not fully containing the hadronic shower, with a stochastic noise term S of 110% and a constant term of 9%, following the formula in Equation 4.3. Note that the noise term is more complicated for the HCAL due to the varying fraction of the electromagnetic shower as a function of the total shower energy.

The HCAL is critical in order to measure precisely the energy of the particle constituents of jets, which will be described in Section 5.5. In particular, the energy of the b jets that arise from the fragmentation of the b partons from the $H \to b\bar{b}$ decay must be measured with sufficiently high precision to resolve $H \to b\bar{b}$ decays from $Z \to b\bar{b}$. Despite the limitations mentioned above, the jet energy resolution is typically about ten percent. This is achieved by combining information from the individual CMS subsystems, as will be described in Section 5.4.

4.2.7 Muon systems

Muons leave track signatures but then pass through the ECAL and HCAL without depositing significant energy. The 2.0 T return field in the iron yoke surrounding the solenoidal magnet is opposite in direction to the 3.8 T field within the magnet volume. The muons therefore bend in the opposite direction when traversing the iron yoke. Muon detection systems are interleaved within the iron yoke with varying type depending on the muon η coverage. The high magnetic field as well as the additional track lever arm from the trajectory through the return yoke enable particularly ex-

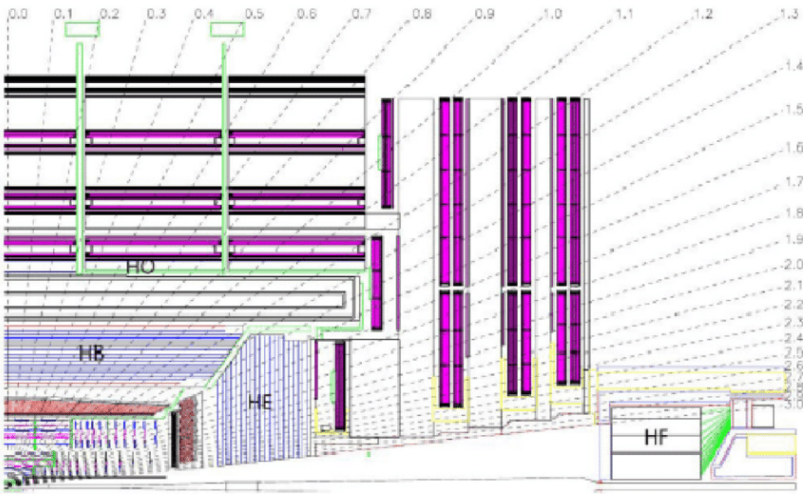


Figure 4.10: View of the CMS hadronic calorimeter in the y-z plane.

cellent muon momentum resolution with CMS, at the level of roughly one percent depending on muon p_{T}.

Figure 4.11 shows a view of one quadrant of the CMS muon systems. In the barrel region, 250 drift tube (DT) detectors are uniformly distributed in five wheels. Each wheel consists of four concentric rings of twelve sectors each. Each DT is a rectangular cell 4.2 cm $\times$ 4.2 cm in the transverse plane containing an anode wire and a mixture of Ar and CO_2 gas. Electrodes placed on the top and the bottom of the cell ensure a constant field and a uniform drift velocity of 55 μm/s. A muon traversing the DT ionizes the gas, such that the free electrons then drift to the anode wire. The position and angle of the incident muon can be inferred from the time it takes for the electrons to drift. Each DT consists of three elements, two of which measure the muon position in the (r,ϕ) plane, and one which measures the z position. Each DT cell has a spatial resolution of about 200 μm, resulting in a resolution of 80 μm to 120 μm for the global DT position measurement.

The cathode strip chamber (CSC) detectors instrument the endcap region ($0.9 < |\eta| < 2.5$). The CSC detectors are designed to handle the stronger magnetic field and higher background rates in this region. The CSCs are trapezoidal in shape and

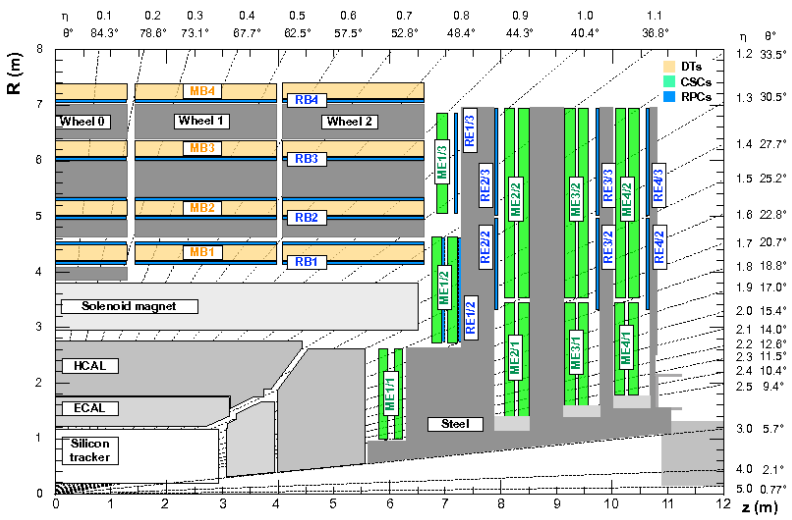


Figure 4.11: View of one quarter of the CMS muon systems in the r-z plane. The DT, CSC, and RPC muon subsystems are shown in orange, green, and blue, respectively.

composed of six layers of anode wires situated between seven segmented cathode plates. The CSCs contain a mixture of Ar, CO_2, and CF_4 gases, which ionize when traversed by a muon. The ionized electrons drift to the anode wires, allowing for position measurements in the (r,ϕ) plane and z direction. The CSC detectors achieve a spatial resolution of 40 μm to 150 μm.

Resistive plate chambers (RPCs) are installed in both the barrel and endcap regions, covering $|\eta| < 1.6$. The RPCs consist of two gaps of 2 mm resistive Bakelite layers separated by a 2 mm volume of a $C_2H_2F_4$/i - C_4H_{10}/SF_6 gas mixture. When a muon traverses the detector, an avalanche is generated by the high electric field inside the gas volume and read out by strips located on the outer surface of the gap. The RPCs achieve a spatial resolution of 0.8 cm to 1.2 cm, much less precise than the DT and CSC systems. The RPCs, however, benefit from excellent timing resolution at the nanosecond level, allowing for the determination of the proton bunch crossing and for use in the Level 1 trigger system, described in the next section.

4.2.8 Data acquisition and trigger

During Run-2 of the LHC, proton bunches collided at a rate of 40 MHz. The full information stored for a collision event saved by CMS is about 1 MB. It is not technically feasible to either store this quantity of data (40 TB/s), or to process the data at 40 MHz. It is therefore necessary to fully process and subsequently write out to disk only a small fraction of the proton-proton collision events in CMS. The vast majority of proton-proton collisions at the LHC have relatively low momentum exchanged between the two protons. These collisions typically produce O(10) charged particles which leave track signatures in the detector. The momentum of these charged particles is typically quite low, such that less than 1% of the charged particles have $p_{\mathrm{T}} > 3$ GeV. The collisions of interest, however, typically concern at least the much larger electroweak energy scale. Leptons from W($\ell\nu$) decays, for example, typically have a p_{T} of at least roughly 30 GeV since $m_W \approx 80$ GeV. A similar argument applies to events where Higgs bosons are produced. The cross sections for these processes are many orders of magnitude smaller than the total proton-proton cross section, such that a reasonable trigger efficiency can be preserved for these events of interest while keeping the overall bandwidth manageable. A set of progressively higher-level decisions are made to decide whether to save a proton bunch crossing event. At each decision stage, a significant fraction of the total event rate is rejected, allowing for a longer computation time budget in the next decision steps.

The CMS Level 1 (L1) trigger is tasked with reducing the event rate from 40 MHz to roughly 100 KHz. The decision of whether to reject an event or not must be made within 3 μs in order to avoid quickly overflowing the buffers and saturating the trigger. With this very stringent time constraint it is necessary to rely primarily on simple pattern recognition of calorimeter clusters and standalone muon tracks built exclusively from hits in the muon detectors. The readout from the tracking detector is not used by the L1 trigger due to these constraints. It is possible at L1 to

consider correlated information, for example by making the trigger decision based on a threshold on both an electromagnetic cluster and a standalone muon track or by inferring a large amount of missing transverse energy in the event.

The High Level Trigger (HLT) further reduces the event rate from 100 KHz to about 1 KHz. The decision of whether to reject an event must be made within several ms in order to avoid saturation. The HLT uses software similar to what is used in the offline reconstruction, but streamlined to run within the HLT time budget. The HLT is able to make much more sophisticated trigger decisions than at L1 based on for example multivariate b-tagging discriminators and reconstructed jets. Events that pass the HLT trigger are saved to disk for offline processing of events, which typically takes several seconds per event and performs a much more computationally intensive event reconstruction which is then used for data analysis.

Chapter 5

Physics object reconstruction

5.1 Primary vertex selection and pileup treatment

Tracks passing minimal quality criteria are clustered using a Deterministic Annealing algorithm [29], which closely parallels the minimization of the free energy in statistical mechanics. The "energy" in this case is defined with respect to the z position of the point of closest approach for a given track z_i^T with uncertainty σ_i^Z compared to the potential vertex z position z_k^V as

$$E_{\text{eff}} \equiv \frac{(z_i^T - z_k^V)^2}{(\sigma_i^Z)^2}. \tag{5.1}$$

This algorithm is significantly more robust with respect to the number of simultaneous pp collisions with respect to the original simple gap clustering algorithm.

The resulting clusters of tracks are then fitted using an adaptive vertex fitter to compute the best estimate of the vertex parameters. Reconstructed primary vertices are required to have a z position within 24 cm of the nominal detector center, a radial

position within 2 cm of the beamspot axis, and a vertex fit exceeding four degrees of freedom, where

$$n_{\text{dof}} = -3 + 2 \sum_{\text{i = 1}}^{\text{\# tracks}} w_i \tag{5.2}$$

and w_i is the probability that a given track corresponds to the given vertex. The signal vertex is chosen as the vertex with the largest $\sum p_T^2$ of associated particle candidates.

Given the high instantaneous luminosity of proton-proton collisions at the LHC, the data sample contains a significant number of additional interactions per bunch crossing, referred to as pileup (PU). The number of reconstructed primary vertices is related to the number of PU interactions in each triggered event, with a primary vertex reconstruction efficiency that is around 70%. The number of PU interactions per event decreases as a function of time during each LHC fill as the instantaneous luminosity decreases, and varies throughout the year due to differences in the LHC collisions settings.

The presence of PU interactions affects the resolution of the reconstructed physics objects described in this chapter. Charged particles from PU interactions are removed from the event of interest by imposing tight requirements on the track position with respect to the primary vertex. The neutral particle PU contribution is removed based on estimates of the PU energy density per unit area. After this removal there is, however, residual resolution degradation from PU that cannot be entirely avoided.

5.2 Electron reconstruction

A significant fraction of the electron energy is emitted via Bremsstrahlung radiation before showering in the ECAL. When the intervening material is minimal ($\eta \approx 0$), on average 33% of the electron energy is radiated. When the amount of intervening

material is largest ($\eta \approx 1.4$), the average percentage of electron energy radiated is about 86% [30]. It is therefore crucial to take into account Brehmsstrahlung radiation when fitting the electron trajectory and measuring the electron energy.

Electron tracks are reconstructed with the Gaussian Sum Filter (GSF) algorithm [30], which uses a modified version of the Kalman Fitter which takes into account the electron energy loss with the Bethe-Heitler function when projecting the track candidate to potential hits on the next tracker layer. ECAL superclusters (SC) group ECAL crystals with an algorithm which attempts to recover the radiated photons, which tend to be spread along the ϕ direction from the electron with similar η due to the bending direction of the magnetic field. GSF electrons are preselected by requiring $p_{\mathrm{T}} > 7$ GeV, $|\eta| < 2.4$, $d_{xy} < 0.05\,\mathrm{cm}$, $d_z < 0.2\,\mathrm{cm}$, where both distances are taken with respect to the primary vertex. A tighter identification is then applied using a multivariate approach. A general purpose multivariate discriminator is trained for electrons that pass a set of relatively loose selection critera. A set of offline cuts on ECAL-based electron quantities is applied on top of the multivariate discriminator to reproduce the conditions of the training sample. Two cuts on the multivariate discriminator are applied, defining two working points based on the expected selection efficiency of either 90% (loose, WP90) or 80% (tight, WP80).

5.3 Muon reconstruction

CMS is particularly well adapted to measure muons with high momentum resolution due to the high magnetic field and the extended lever arm as muons traverse the magnetic field return yoke. Muon tracks are first reconstructed separately in the tracker and from the muon chamber information [31, 32]. A full description of how this information is combined to form the muon candidate is given in Section 5.4.

Muons are preselected by requiring $p_{\mathrm{T}} > 5$ GeV, $|\eta| < 2.4$, $d_{xy} < 0.5\,\mathrm{cm}$, and $d_z < 1.0\,\mathrm{cm}$. The muon candidates are then required to pass either a loose or tight identification working point, depending on the required fake muon candidate rejection. These working points consist of a set of selections on detector-level quantities corresponding to the muon candidate.

5.4 Particle flow

The raw detector readout information is interpreted in terms of final state physics objects (muons, electrons, photons, and charged/neutral hadrons) using the Particle Flow (PF) algorithm [33]. PF takes advantage of the hermeticity of the CMS detector and redundancy of energy measurements to achieve resolutions typically better than is possible with the individual detector subsystems alone. The PF algorithm begins with a set of particle trajectories from the tracker and calorimeter clusters built from deposits in the ECAL and HCAL. Note that in addition to the tracks reconstructed in the tracker, an additional set of standalone muon tracks through the magnet return yoke are obtained by fitting the hit positions in the muon detectors. Each of these objects are "linked" with other objects when they can be associated with a specified distance parameter less than a certain threshold value. The three possible forms of links are:

- **track - calorimeter cluster**: the track trajectory is extrapolated to the calorimetry with a distance parameter based on the η,ϕ separation between the extrapolated track position and the center of the calorimeter cluster.
- **calorimeter cluster - calorimeter cluster**: individual clusters are linked when satisfying sufficiently small η,ϕ separation. This linking includes the possibility of combining clusters in the ECAL and HCAL.

- **track - standalone muon track**: a global fit is performed on the hits corresponding to combinations of tracks and standalone muon tracks. Links are formed when the global track fit χ^2 is below a given threshold.

The linked detector-level objects form "blocks", which are then classified by particle type. The classification procedure is sequential, starting from the particles that are most efficiently identified. Muons are first classified from blocks with links between tracks and standalone muon tracks, where the momentum of the muon track is consistent with the individual track momentum within three standard deviations. Next, electrons are identified from blocks with links between a track and a calorimeter cluster in the ECAL. In addition, the track is refit following the GSF procedure described in Section 5.2 to take into account Brehmsstrahlung radiation by the electron. The refitted GSF track must be of reasonable fit quality and still consistent with the ECAL cluster.

Blocks not yet classified as muons or electrons with links between one or more tracks and a calorimeter cluster are then considered. The total track momentum is compared with the estimated calorimeter energy. If the track momentum is larger than the calorimeter energy, additional track filtering requirements are applied in order to reduce fake tracks and identify muons with looser criteria. The block with filtered tracks is then classified as a charged hadron. If instead the track momentum is less than the calorimeter energy, the excess energy is interpreted as a photon. If the excess energy is more than the total energy deposited in the ECAL, a neutral hadron candidate is also identified with the remaining excess energy. The remaining track to calorimeter cluster link, with the neutral particle energy removed, is identified as a charged hadron. Blocks formed from links between calorimeter clusters without linked tracks are subjected to tighter linking requirements, then identified as photons (from ECAL clusters) or neutral hadrons (from HCAL clusters).

5.5 Jet reconstruction

Quarks and gluons from the hard scattering process "shower", radiating away energy by emitting gluons which subsequently decay to quark-antiquark pairs. The parton shower is described with perturbative QCD under the collinear and soft radiation approximation, which allows for the otherwise intractable full matrix element calculation to be factorized into a time-ordered sequence of probabilistic radiative processes. Once the energy scale of the particles reaches roughly 1 GeV, however, the strong coupling constant α_S approaches unity and perturbative QCD is no longer valid. QCD does not allow for free quarks but rather requires quarks to form color singlet bound states (hadrons). At the 1 GeV energy scale, color confinement forces become dominant and hadrons form. This process is referred to as "hadronization", and is described by multiple models which attempt to mimic known effects of non-perturbative QCD from lattice QCD calculations. The parton shower and hadronization happen on a timescale much shorter than can be resolved by the detector, meaning that quarks and gluons produced in the hard scattering process leave detector signatures of collimated sprays of many particles known as jets.

The observable of interest is typically the kinematics of the parton from the hard interaction rather than the individual jet constituents. It is therefore necessary to correctly group the collection of final state hadrons to each original parton such that the parton four-vector can be reconstructed from the sum of all particles within the jet. It is important that the jet reconstruction algorithm used experimentally is also well defined theoretically so that the experimental data can be meaningfully compared with theoretical predictions. Simulation of the parton shower is truncated at a fixed order in QCD perturbation theory, which requires a balancing between real and virtual corrections in order to cancel collinear and soft radiation divergences. If the experimental observables are sensitive to very soft or collinear radiation effects, the fixed order simulation is no longer a valid description of the data.

The jet clustering algorithms used in CMS are sequential algorithms with the following distance measure used to choose the next particle pairing

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \tag{5.3}$$

where k_{ti} denotes the transverse momentum of a given particle i. Jets are reconstructed from PF candidates using the anti-k_T clustering algorithm, which corresponds to p = -1, with distance parameter $R = 0.4$ [34, 35]. The anti-k_T algorithm is favored because it is infrared and collinear safe, and because it generally yields conical jet shapes of radius R even though the algorithm is sequential.

Reconstructed jets require a small additional energy correction, mostly due to thresholds on reconstructed tracks and clusters in the PF algorithm as well as various reconstruction inefficiencies [36]. Jet identification criteria are applied to reject misreconstructed jets resulting from detector noise, as well as jets primarily reconstructed with particles from pileup interactions [37].

As described in Section 4.2.4, the tracker geometry requires track $|\eta| < 2.5$. The CMS calorimetry extends to larger η, making it possible to reconstruct forward jets with $|\eta| > 2.5$, reconstructed from PF candidates built without tracks. Forward jet reconstruction is essential for measuring VBF signatures (Sec. 4.1.2), where typically one or both of the VBF quark jets has $|\eta| > 2.5$ in the kinematic region where it is possible to distinguish signal from background. Forward jets generally have much larger energy scale uncertainties than central ($|\eta| < 2.5$) jets because it is much more difficult to reject pileup without the jet constituent track information. For VH production, the Higgs boson is expected to be produced primarily with $|\eta| < 2.5$, therefore in this analysis only jets with $|\eta| < 2.5$ are considered.

The jet energy resolution, defined with respect to the generator-level parton p_{T}, is generally underestimated by the simulation. The jet energy resolution in simulation is therefore corrected to match the data via calibration studies in a Z + 1-jet control

sample. The correction is derived as a function of jet p_{T} and η, with an average correction of about 10%.

5.6 Identification of b jets

Hadrons containing b quarks (B hadrons) have lifetimes on the order of 1 ps, such that B hadrons travel distances ranging from several mm to several cm in the lab frame, depending on momentum. This is significantly longer than other non-stable hadrons which decay within the detector, such that jets formed from b partons contain a characteristic experimental signature of secondary displaced vertices within the jet from the B hadron decays.

The identification of jets that originate from b quarks is performed with a deep neural network (DNN) multivariate classifier, the Deep Combined Secondary Vertex (DeepCSV) algorithm [38]. The most important DeepCSV training inputs consider the potential for a significantly displaced secondary vertex, but additional training inputs involving the individual jet constituent kinematics also improve the DeepCSV performance. These additional inputs increase performance because b jets tend to have a larger jet mass and harder fragmentation than jets originating from light quarks. DeepCSV is a multiclassifier that returns separate probabilities that a jet corresponds to each of the following original parton hypotheses: b, bb, c, cc, and light quarks. The sum of all probabilities for a given event is always unity. The hypotheses bb and cc refer to the case that the two partons are sufficiently close together such that the jets overlap and are reconstructed as a single jet. This analysis uses P(b)+P(bb) to identify b jets. Table 5.1 summarizes the selection efficiencies for b jets, c jets, and light quark (udsg) jets for each of the three working points used in this analysis to select b jets.

Table 5.1: The overall selection efficiency for b jets, c jets, and light quark (udsg) jets for each of the three DeepCSV P(b) + P(bb) working points used in this analysis.

	working point	$\epsilon_{\mathrm{b}}(\%)$	$\epsilon_{\mathrm{c}}(\%)$	$\epsilon_{\mathrm{udsg}}(\%)$
DeepCSV P(b) + P(bb)	loose	84	41	11
	medium	68	12	1.1
	tight	50	2.4	0.1

5.7 Lepton isolation

In addition to muons and electrons from the electroweak decay of massive particles such as Z or W bosons (prompt leptons), a large number of leptons are generally produced in jets via decays of heavy flavor hadrons or the decay in flight of charged pions or kaons. The leptons from jets are generally in close proximity to other PF candidates within the jet. Prompt leptons can therefore be identified by vetoing the presence of significant additional activity in close proximity to the lepton candidate.

The lepton isolation is quantified by estimating the total p_{T} of particles within a given $\Delta\mathrm{R}$ of the lepton [39, 40].

$$\mathrm{I_{PF}} \equiv \frac{1}{\mathrm{p_T^{\ell}}}\Big(\sum \mathrm{p_T^{charged}} + \max\Big[0, \sum \mathrm{p_T^{neutral}} + \sum \mathrm{p_T^{\gamma}} - \mathrm{p_T^{PU}}(\ell)\Big]\Big). \tag{5.4}$$

The term $\sum \mathrm{p_T^{charged}}$ denotes the scalar sum of the transverse momenta of PF charged hadrons with tracks matched to the primary vertex, while the terms $\sum \mathrm{p_T^{neutral}}$ and $\sum \mathrm{p_T^{\gamma}}$ denote the scalar sums of the transverse momenta for PF neutral hadrons and PF photons, respectively. Note that due to the lack of an associated track for the neutral particle candidates, it is much more difficult to reject pileup energy than for charged hadrons. Because $\mathrm{I_{PF}}$ is particularly sensitive to energy deposits from pileup interactions, the estimated PU contribution $\mathrm{p_T^{PU}}(\ell)$ is subtracted, using two different techniques. For muons, the definition $\mathrm{p_T^{PU}}(\mu) \equiv 0.5 \times \sum_i \mathrm{p_T^{PU,}}^i$ is used, where i runs over the momenta of the charged hadron PF candidates not originating from the

primary vertex, and the factor of 0.5 corrects for the expected 2:1 fraction of charged to neutral particles from hadronic decays. For electrons, the FASTJET technique [41] is used, in which $\mathrm{p_T^{PU}}(e) \equiv \rho \times A_{\mathrm{eff}}$, where the effective area A_{eff} is the geometric area of the isolation cone scaled by a factor that accounts for the residual dependence of the average pileup deposition on the η of the electron, and ρ is the median of the $\mathrm{p_T}$ density distribution of neutral particles within the area of any jet in the event.

5.8 Missing transverse energy reconstruction

Neutrinos produced in the collision, either from the hard scattering process or from hadron decays, traverse the CMS detector without leaving any experimental signature. The neutrino energy can therefore only be inferred from a momentum imbalance in the observed particles from the collision. As described in Section 4.2.2, the momentum measurement of interest in hadron collisions is in the x-y plane transverse to the beam line. The missing energy in the transverse plane is similarly the relevant quantity when reconstructing the neutrino, which should be null in the absence of undetected particles by momentum conservation.

The vector $\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the negative of the vectorial sum of transverse momenta of all PF candidates in the event. The scalar quantity $\mathrm{E_T^{miss}}$ is defined as $\mathrm{E_T^{miss}} = |\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}|$. A set of filters is applied to remove known issues of instrumental noise and problematic events. An additional quantity of interest is the $\mathrm{E_T^{miss}}$ significance, defined as

$$\mathrm{E_T^{miss}\ significance} \equiv \frac{\mathrm{E_T^{miss}}}{\sqrt{\sum_i |\vec{p_{Ti}}|}}, \tag{5.5}$$

where $\sum_i |\vec{p_{Ti}}|$ considers all PF candidates in the event. The $\mathrm{H_T^{miss}}$ is also considered, defined similarly to $\mathrm{E_T^{miss}}$ except that only jets with $p_T > 30$ and $|\eta| < 2.4$ are considered rather than all PF candidates. The $\mathrm{H_T^{miss}}$ is less sensitive to pileup than

the $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$, making it a useful complementary identifier of events with high-momentum neutrinos.

5.9 Additional "soft" hadronic activity

Jets built from PF candidates, as described in Section 5.5, are typically only considered when the jet p_{T} exceeds 15 GeV. Below 15 GeV, the contamination from PU energy is very large, such that the reconstructed jets often primarily consist of energy from PU rather than the primary interaction. Hadronic activity at much lower p_{T} can be probed, however, by using as jet constituents only tracks matched to the primary vertex.

A collection is built of tracks in the event with $p_{\mathrm{T}} > 300$ MeV and with the smallest distance in the z direction between the track and the primary vertex less than 2 mm. This track collection is much less contaminated by PU than the PF candidates, particularly for low-p_{T} particle candidates. Tracks overlapping with either of the Higgs boson candidate b jets (Sec. 7.3) are excluded. In addition, tracks in the region between the two b jets are removed by defining an ellipse in the (η, ϕ) plane around the two b jets with axes $(a; b) = (\Delta R(\mathrm{bb}) + 1; 1)$, and excluding all tracks pointing within the ellipse. This final requirement excludes charged particles likely arising from color exchange between the b jets.

The "soft" track-jets are then clustered from this track collection using the anti-k_T algorithm with distance parameter $R = 0.4$, the same algorithm as used for jets clustered from PF candidates (Sec. 5.5). The number of soft track-jets in the event with $p_{\mathrm{T}} > 5$ GeV (N_5^{soft}), is useful in the discrimination between signal and background, as will be described in Chapter 8.

Chapter 6

Event simulation

6.1 Monte Carlo event generators

In order to interpret the observed data, it is necessary to accurately simulate the expected contributions in the selected analysis regions of all potential physics processes. The simulation must accurately predict not only the process yield in the region of interest but also the correct distribution of many analysis observables simultaneously. The event simulations use Monte Carlo (MC) random sampling techniques to generate events for each process of interest which cover the full kinematic range of the data provided sufficiently many generated events.

This challenging task begins with a simulation of the hard scattering process, which calculates the matrix element for the given scattering process at fixed order. This analysis uses the MADGRAPH5 amc@NLO v2.4.2 [42] and POWHEG v2 [43, 44, 45] event generators at both leading order (LO) and next-to-leading order (NLO) accuracy in QCD, depending on the process.

As described in detail in Section 5.5, quarks and gluons from the hard scattering process shower and hadronize. This analysis interfaces the generated hard scattering events with PYTHIA v8.230 [46] to simulate these effects as well as the contribution

from the underlying event and multiple parton interactions. Events are then processed with GEANT4 [47] to simulate the detector response and subsequently reconstructed using the same algorithms that are applied to the data.

The quark-induced ZH and WH signal processes are generated at NLO QCD accuracy using POWHEG v2 extended with the MiNLO procedure [48, 49], while the gluon-induced ZH process is generated at LO accuracy with POWHEG v2. The Higgs boson mass is set to 125 GeV for all signal samples. Diboson background events are generated with MADGRAPH5 aMC@NLO v2.4.2 at NLO with the FxFx merging scheme [50] and up to two additional partons. The same generator is used at LO accuracy with the MLM matching scheme [51] to generate V+jets events in inclusive and b-quark enriched configurations with up to four additional partons, and to generate a sample of QCD multijet events. The $t\bar{t}$ [52] and single top production processes in the tW [53] and t [54] channels are generated to NLO accuracy with POWHEG v2, while the s channel [55] single top process is generated with MADGRAPH5 aMC@NLO v2.4.2. The parton distribution functions used to produce all samples are the next-to-next-to-leading order (NNLO) NNPDF3.1 set [56]. For all samples, simulated additional pp interactions (pileup) are added to the hard-scattering process with the multiplicity distribution matched to the 2017 data.

6.2 Additional corrections

It is necessary to apply residual corrections to the MC in order to match the data for the important m_{jj} and $p_{\mathrm{T}}(\mathrm{V})$ distributions. Some of these residual corrections account for differences between the leading order and NLO predictions for the V+jets simulation. It is difficult to generate NLO V+jets MC with sufficiently high statistics to model the V+jets distributions for this analysis. This is due to the large computation time needed to generate NLO events and the significant fraction of generated NLO

events with negative weights, which highly reduces the effective statistical power of the sample. This analysis therefore uses the LO prediction for the V+jets simulation and uses the NLO V+jets MC to derive an effective reweighting to apply to the LO V+jets MC. There is additionally a per-event correction applied as a function of $p_{\mathrm{T}}(\mathrm{V})$ to the signal and V+jets simulation which takes into account electroweak NLO corrections.

6.2.1 Differential electroweak NLO corrections in $p_{\mathrm{T}}(\mathrm{V})$

As described in Section 6.1, the predominant quark-induced signal is simulated at NLO accuracy in QCD. The total signal cross section at QCD NNLO and electroweak NLO accuracy σ^{VH} [57] is given by:

$$\sigma^{VH} = \sigma^{VH,DY}(1 + \delta_{EW}) + \sigma_{\text{t-loop}} + \sigma_{\gamma}, \tag{6.1}$$

where $\sigma_{\text{t-loop}}$ corresponds to NNLO diagrams including closed fermion loops, σ_{γ} includes virtual and real gluon or quark radiation effects, and $\sigma^{VH,DY}$ is the dominant contribution arising from diagrams similar to single vector-boson production diagrams. The term δ_{EW} considers the reduction in total cross section from NLO electroweak effects, ranging from -4% to -7% depending on the process.

The NLO electroweak corrections affect not only the total cross section but also the $p_{\mathrm{T}}(\mathrm{V})$ distribution, one of the most important analysis observables. Because the electroweak corrections factorize up to NLO, they can be applied differentially in $p_{\mathrm{T}}(\mathrm{V})$ as a multiplicative weight $(1 + \delta_{EW})$ to the signal. Figure 6.1 shows the shape of these corrections for two of the channels.

Similar differential NLO electroweak corrections to the $p_{\mathrm{T}}(\mathrm{V})$ are applied to the V+jets simulation in all channels [58]. Fig. 6.2 shows the differential correction applied to the V+jets simulation.

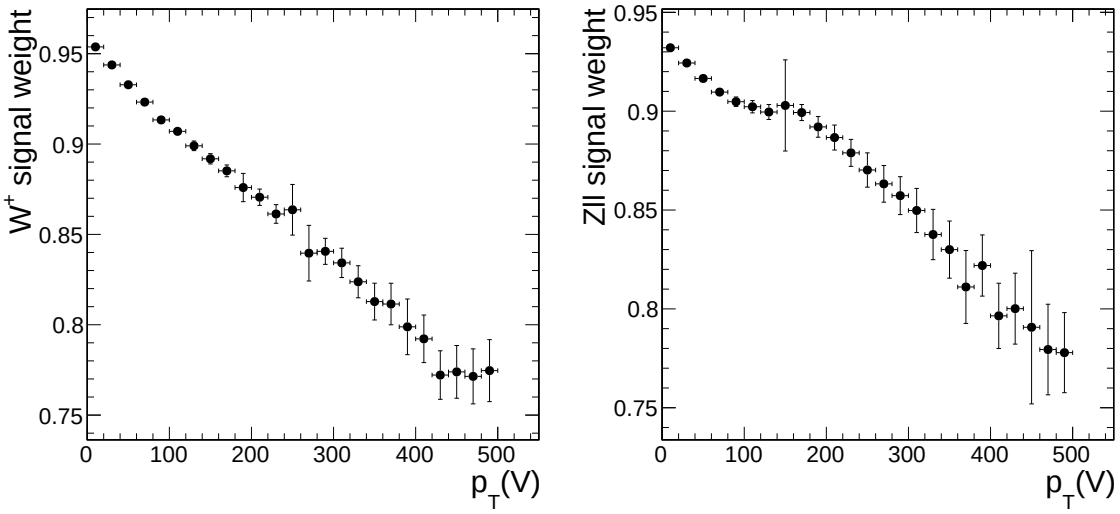


Figure 6.1: Differential NLO electroweak correction for the W^+H (left) and ZH (right) processes as a function of $p_{\mathrm{T}}(\mathrm{V})$ [7, 8].

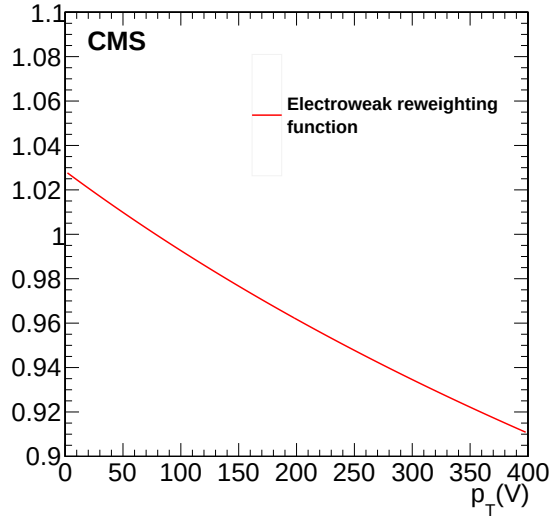


Figure 6.2: NLO electroweak correction as a function of $p_{\mathrm{T}}(\mathrm{V})$ for the V+jets processes.

6.2.2 Dijet pseudorapidity difference reweighting for leading order V+jets simulation

Reconstruction-level comparisons of the $\Delta\eta$(jj) distributions for LO and NLO V+jets simulated events are used to derive an event reweighting. The NLO/LO ratio as calculated in an inclusive phase space is shown in Fig. 6.3, considering separately the cases $V + 0b$, $V + 1b$, and $V + 2b$. The ratio is not changed significantly if generator-level quantities are considered, or if the selections are varied. The correction is applied to both Z+jets and W+jets simulation. After the reweighting is applied, the LO MC prediction for the invariant mass distribution matches well the data. Other distributions are unaffected, apart from a slight improvement in the jet p_{T} modeling. The full correction is considered as a systematic uncertainty in the signal extraction fits (Sec. 10.2).

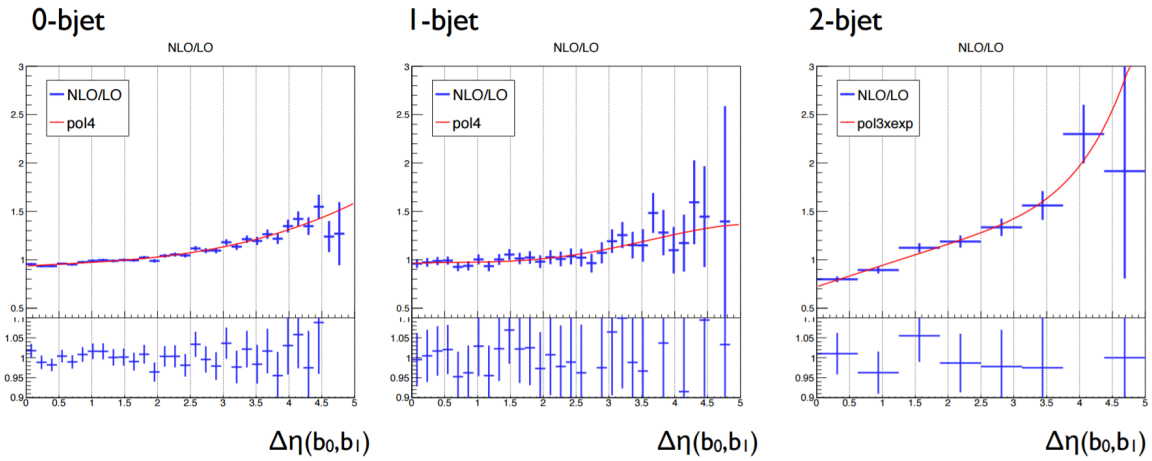


Figure 6.3: Ratio of the NLO to LO DY+jets MC prediction as a function of the reconstruction-level $\Delta\eta$(jj) for $Z + 0b$ (left), $Z + 1b$ (center), and $Z + 2b$ (right) events.

Chapter 7

Analysis strategy

7.1 VH, H $\rightarrow b\bar{b}$ overview

The production of bottom quarks via strong interactions, hereby denoted as the QCD multijet background or simply QCD, has a production rate at the LHC seven to nine orders of magnitude larger than the expected rates for Higgs boson production. This makes an inclusive search for H $\rightarrow b\bar{b}$ at the LHC extremely experimentally challenging. As described in Section 4.1.2, the dominant production mode for Higgs bosons at the LHC is the gluon fusion production mode. Due to the overwhelming QCD background, however, it is not possible to search for H $\rightarrow b\bar{b}$ in this production mode without imposing stringent additional kinematic requirements. CMS was recently the first experiment to attempt a measurement of H $\rightarrow b\bar{b}$ produced via gluon fusion by considering events with an additional high-p_{T} (greater than 450 GeV) jet radiated by one of the initial-state partons [59]. The most effective H $\rightarrow b\bar{b}$ search strategy at the LHC is however to consider VH production. Requiring a high-p_{T} leptonically decaying W or Z boson in the event reduces the QCD contribution by many orders of magnitude, allowing for a signal-to-background ratio sufficiently high to make the H $\rightarrow b\bar{b}$

analysis feasible. After this requirement, the QCD background contamination to the signal region is a small fraction of the overall remaining background.

7.1.1 Signal topology

VH, H $\to$ b$\bar{\text{b}}$ signal events are are characterized by the presence of a high-p_{T} vector boson recoiling against two b jets with an invariant mass consistent with $m_H = 125$ GeV within the experimental mass resolution. The vector boson and dijet system are expected to be central ($|\eta| < 2.4$) and back-to-back in the transverse plane. Furthermore, the Higgs boson p_{T} spectrum for VH production is expected to be significantly higher than for SM backgrounds. The distinct topology of a high-p_{T} vector boson recoiling against a high-p_{T} H $\to$ b$\bar{\text{b}}$ candidate is exploited to reduce the remaining backgrounds, as will be described in Chapter 8. Furthermore, vector boson decays to leptons, which yield high-p_{T} charged electrons and muons in the final state or large missing transverse energy, provide a simple and efficient strategy to identify signal events. This is critical in order to save signal events under the very tight bandwidth and latency requirements at the LHC, as will be described in Section 7.4.1. Reconstructing the Higgs boson and vector boson candidates with high resolution is essential to fully exploit the kinematic differences between signal and background events.

7.2 Vector boson reconstruction

Reconstruction of W and Z bosons begins with the identification and selection of charged leptons and $\text{E}_{\text{T}}^{\text{miss}}$, as described in Chapter 5. Three leptonic vector boson decay modes are considered: Z $\to \ell^+\ell^-$, denoted Z($\ell\ell$), W $\to \ell^-\bar{\nu}$ and W $\to \ell^+\nu$, denoted W($\ell\nu$), and Z $\to \nu\bar{\nu}$, denoted Z($\nu\bar{\nu}$). The symbol ℓ refers to both electrons and muons. Candidate Z($\ell\ell$) decays are reconstructed from opposite-sign electron

or muon pairs with $p_{\mathrm{t},\ell} > 20$ GeV and dilepton invariant mass consistent with a Z boson: $75 < m_{\ell\ell} < 105$ GeV. The electrons must pass WP90 (Sec. 5.2) and $\mathrm{I_{PF}} < 0.15$ (Sec. 5.7). Muons must pass loose identification criteria (Sec. 5.3) and $\mathrm{I_{PF}} < 0.25$. Z($\ell\ell$)H candidate events are split into two categories depending on $p_{\mathrm{T}}(\mathrm{Z})$. The low $p_{\mathrm{T}}(\mathrm{Z})$ category considers $50 < p_{\mathrm{T}}(\mathrm{Z}) < 150$ GeV events, while the high $p_{\mathrm{T}}(\mathrm{Z})$ category requires $p_{\mathrm{T}}(\mathrm{Z}) > 150$ GeV.

Candidate W($\ell\nu$) events are identified by the topology of a single isolated electron passing WP80, $\mathrm{I_{PF}} < 0.06$, and $p_{\mathrm{T}}(e) > 30$ GeV, or a single isolated muon with $\mathrm{I_{PF}} < 0.06$ and $p_{\mathrm{T}}(\mu) > 25$ GeV. The transverse momentum $p_{\mathrm{T}}(\mathrm{W})$ and mass $M_{\mathrm{T}}(W)$ of the W candidate are computed as

$$p_{\mathrm{T}}(\mathrm{W}) = \sqrt{(\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}} \cdot \hat{x} + p_x^{\ell})^2 + (\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}} \cdot \hat{y} + p_y^{\ell})^2}, \quad \texttt{and} \tag{7.1}$$

$$M_{\mathrm{T}}(W) = \sqrt{2 p_{\mathrm{T}}^{\ell} \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}} (1 - \cos\theta)}, \tag{7.2}$$

where θ is the angle in the transverse plane between the lepton and $\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}$.

The W($\ell\nu$)H analysis is performed in one category with $p_{\mathrm{T}}(W) > 150$ GeV, which is sufficiently stringent to eliminate any residual QCD background contribution while preserving the high-S/B kinematic region where the signal is measured.

Candidate Z($\nu\bar{\nu}$) decays are preselected by requiring $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV. The transverse momentum of the Z boson candidate is then defined as $p_{\mathrm{T}}(\mathrm{Z}) \equiv \min(\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{H}_{\mathrm{T}}^{\mathrm{miss}})$. The Z($\nu\nu$)H analysis is performed in one category with $p_{\mathrm{T}}(\mathrm{Z}) > 170$ GeV. This selection is driven primarily by trigger requirements, as will be discussed in Section 7.4.1.

7.3 Higgs boson reconstruction

It is critical to achieve the best possible dijet mass resolution for the Higgs boson candidate and to maximize the signal efficiency when selecting the $\mathrm{H} \to \mathrm{b\bar{b}}$ candidate

b jets, particularly when more than two jets are present in the event. The energy resolution for jets in the p_{T} range $\sim [50, 150]$ GeV is substantially better than for jets with either p_{T} near the kinematic reconstruction threshold ($\sim$20 GeV) or very large p_{T}. For low-p_{T} jets this is due to a much higher contamination from PU. For very high-p_{T} jets, the jets are highly collimated in the lab frame with a dense jet core of many relatively straight and potentially overlapping high-p_{T} tracks. Requiring $p_{\mathrm{T}}(\mathrm{H}) \gtrsim 150$ GeV therefore not only enhances the inclusive S/B but selects a region with relatively high dijet mass resolution.

Selecting $\mathrm{H} \to \mathrm{b\bar{b}}$ candidates by identifying the dijet combination with the largest associated b-tagging discriminate (DeepCSV, Sec. 5.6) values ensures high signal efficiency. The signal efficiency of this selection is around 80% for $p_{\mathrm{T}}(\mathrm{V}) > 250$ GeV. The invariant mass resolution of the selected Higgs boson candidate is further improved with a novel b-jet energy regression, a kinematic fit to exploit the absence of real missing transverse energy for signal events in the $\mathrm{Z}(\ell\ell)\mathrm{H}$ channel, and a procedure to recover final state radiation. Each of these improvements will be described in this section.

7.3.1 b-jet energy regression

As mentioned in Section 5.6, b jets have distinct kinematic properties which can be exploited to better estimate the energy of the b quark from the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay. In particular, semileptonic B hadron decays within the b jet yield neutrinos which are not detected, leading to an underestimation of the b-quark p_{T} by the reconstructed b jet. The use of multivariate regression techniques to improve the resolution of the dijet invariant mass was first introduced at the Tevatron [60] and further optimized at the LHC [61]. A multivariate regression is trained to estimate the generator-level b-jet energy for a given reconstruction-level b jet based on the reconstructed b-jet kinematic observables. The per-jet resolution and energy scale bias are subsequently improved.

The primary improvement arises from correcting for the estimated missing neutrino energy in the case of semileptonic B hadron decays. There is however an additional secondary improvement achieved from performing a dedicated energy calibration for b jets. New for this analysis is the use of a deep neural network to perform the regression.

A neural network with 6 hidden layers is trained, with the ratio between the generator-level p_{T} of the jet including neutrinos and the reconstructed jet p_{T} as the training target. The regression is trained on simulated $\mathrm{t\bar{t}}$ events in order to avoid biases towards the $\mathrm{H} \to \mathrm{b\bar{b}}$ signal properties. Only b jets matched with a generator-level b quark with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.4$ are considered.

The complete set of observables used in the regression training are:

- Jet kinematics:
  - Jet p_{T}, η, mass, and transverse mass (M_{T})
  - p_{T} of the highest-p_{T} track within the jet
  - Jet momentum dispersion p_T^D – $\sqrt{\frac{\sum_i {p_{\mathrm{T}}}_i^2}{\sum_i {p_{\mathrm{T}}}_i}}$
- Jet vertex information:
  - flight length of the jet secondary vertex (and error)
  - mass and p_{T} of the jet secondary vertex
  - number of tracks associated with the jet secondary vertex
- Jet constituent properties:
  - energy fraction of the neutral constituents detected in the ECAL
  - energy fraction of the neutral constituents detected in the HCAL
  - energy fraction of the charged constituents detected in the ECAL
  - energy fraction of the charged constituents detected in the HCAL

  - ΔR of soft lepton candidate (if present) with respect to the jet axis
  - magnitude of jet momentum transverse to lepton axis
  - magnitude of lepton momentum transverse to jet axis

- Further jet constituent properties (new for this analysis):
  - Lepton flavor (electron, muon, no lepton)
  - Number of jet daughters with $p_{\mathrm{T}} > 0.3$ GeV
  - Energy fractions in rings of increasing radius around the jet axis ([0,0.05], [0.05,0.1], [0.1,0.2], [0.2,0.3], [0.3,0.4]), considered separately for:
    * Electrons and photons
    * Muons
    * Charged hadrons
    * Neutral hadrons

- ρ – average jet energy density (highly correlated with PU).

The agreement between simulation and data has been validated in dedicated control regions for all the regression training input observables. The same regression is used consistently in all analysis categories. The regression improves the per-jet energy resolution by 15% on average and therefore the dijet mass resolution by 20%. Figure 7.1 shows a comparison of the dijet invariant mass for $\mathrm{Z}(\ell\ell)\mathrm{H}(\mathrm{b\bar{b}})$ signal simulation before (blue) and after (red) applying the b-jet energy regression.

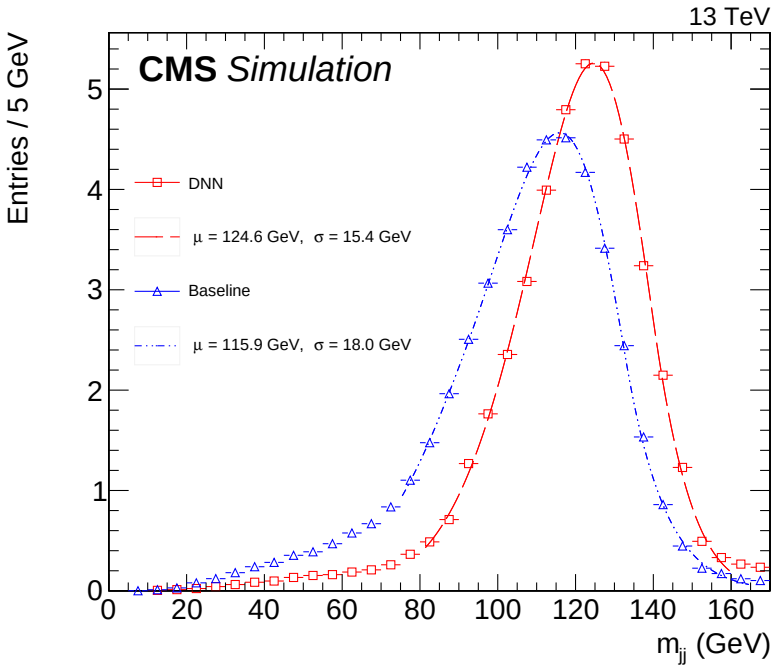


Figure 7.1: Dijet invariant mass distribution for $Z(\ell\ell)H(b\bar{b})$ signal before (blue) and after (red) applying the DNN b-jet energy regression. Each distribution is fit with a Bukin function. The fitted mean and width of the distributions are displayed in the figure.

7.3.2 Final state radiation recovery

Partons in the proton can emit initial state radiation (ISR) before the hard scattering. Colored final state particles can also be emitted as final state radiation (FSR). These emissions can be identified as additional jets in the event, other than the two Higgs candidate b jets. In particular, FSR jet radiation is generally soft or collinear. It is not possible to reconstruct jets with $p_{\mathrm{T}} < 15$ GeV due to large PU contamination. However, approximately collinear FSR jets can be recovered by considering jets with small angular separation from the selected b jets. In order to improve the estimation of the Higgs boson mass, the four-vector of the Higgs boson candidate is corrected by adding the four-vector of any additional jet with $p_T > 20$ GeV, $|\eta| < 3.0$, and within $\Delta R < 0.8$ of either Higgs candidate b jet. The FSR recovery improves the mass resolution of signal events by 2%, without sculpting the background shape.

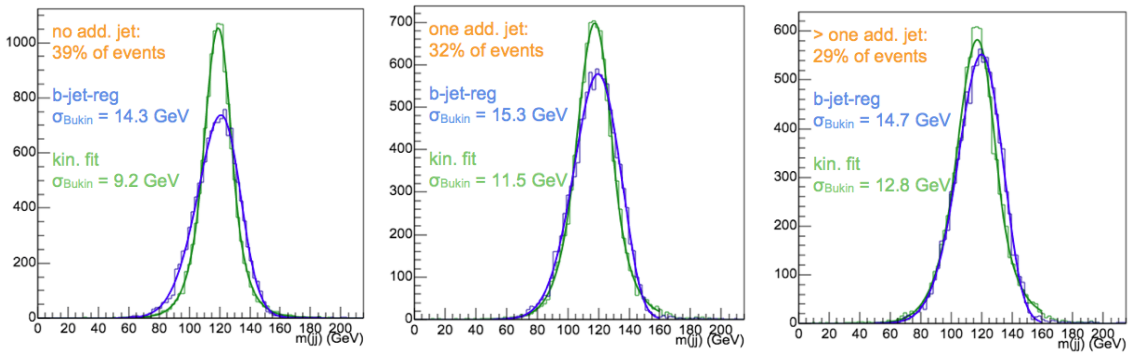
7.3.3 Kinematic fit in the Z($\ell\ell$)H channel

It has been shown that the resolution of the measured objects in the final state of proton collisions can be improved by imposing well-motivated kinematic hypotheses through an event-by-event least square fitting technique [62]. The resulting probability of the chi-square of the fit can be interpreted as the probability of the proposed kinematic hypotheses to be true for the observed event.

As mentioned in earlier chapters, the vector sum of the transverse momenta of all particles in events from pp collisions at the LHC should be null by momentum conservation. In events where high-resolution final state particles such as charged leptons are present and no undetected particles (neutrinos) are produced, this kinematic constraint can be used to improve the energy estimate of other objects otherwise reconstructed with poor resolution, particularly jets. In the Z($\ell\ell$)H channel such techniques are possible because only two leptons, two b jets, and possibly jets from ISR and/or FSR are present in the final state under the signal hypothesis.

The kinematic fit procedure constrains the dilepton system to the Z boson mass and constrains the dilepton-dijet system in the transverse plane to be momentum-balanced. A fit is then performed to the lepton and jet transverse momentum, allowing the values to vary within uncertainties.

In the following, events are selected according to the signal region definition of the Z($\ell\ell$)H high-p_{T} channel, with $p_{\mathrm{T}}(\mathrm{Z}) > 150$ GeV (Sec. 8.2). Furthermore, both Higgs boson candidate b jets are required to be matched to the generator-level b quarks within $\Delta R < 0.2$. The Higgs boson candidate invariant mass distribution is shown in Fig. 7.2 before, (with regression, blue), and after (green) the kinematic fit, for different ISR jet multiplicities, while Table 7.1 details the resolution before and after the kinematic fit in bins of $p_{\mathrm{T}}(\mathrm{V})$ and the number of ISR jets. The kinematic fit improvement is the largest for events without any ISR jets, where there is no additional degree of freedom in the fit from variations on the ISR jet p_{T}. The resolution improvement is roughly inversely proportional to the number of ISR jets, but sizeable even for events with more than 1 ISR jet.



(a) 0 ISR Jets (b) 1 ISR Jet (c) 2 or more ISR Jets

Figure 7.2: m_{jj} distribution for Z($\ell\ell$)H(b$\bar{\mathrm{b}}$) signal simulation before (with regression, blue curve) and after (green curve) the kinematic fit. The percentage of events written in orange (top left of figures) is derived with respect to the number of events where both reconstructed Higgs boson candidate b jets are matched to the generator-level b quarks.

Table 7.1: Higgs boson candidate invariant mass resolution before and after the kinematic fit in bins of p_{T}(V)and ISR jet multiplicity. The resolutions given are in units of GeV.

p_{T}(V)	#ISR Jets	σ_{reg}	σ_{fit}	Improvement (%)
> 150	0	14.3	9.2	36
> 150	1	15.3	11.5	25
> 150	> 1	14.7	12.8	13

7.4 Dataset

41.3 fb^{-1} of proton-proton collision data with center-of-mass energy $\sqrt{s} = 13$ TeV and 25 ns proton bunch spacing collected by CMS in 2017 has been used in this analysis (Sec. 4.1.1).

7.4.1 Trigger requirements

A small fraction of LHC collision events are recorded by CMS (Sec. 4.2.8). Collision events must first pass the L1 trigger, which decides whether to record the event based on simple pattern recognition of calorimeter clusters and muon detector readouts. The remaining events are then processed by the HLT trigger, which performs a streamlined version of the full offline event reconstruction in order to determine whether to save the event to disk. Table 7.2 summarizes the triggers used in this analysis to collect events consistent with the signal hypothesis in each channel.

The chosen triggers constitute the minimal set of selections that must be made in order to reduce the event rate to a technically feasible throughput while ensuring as high a signal efficiency as possible. The W($\mu\nu$)H and W(eν)H channels require single lepton triggers. The Z($\mu\mu$)H and Z(ee)H channels are based on dilepton triggers, which enable a lower lepton p_{T} requirement with respect to the single lepton triggers. Triggering events for the Z($\nu\nu$)H channel is more difficult due to the lack of electrons or muons in the final state. The Z($\nu\nu$)H triggers instead require a large missing

Table 7.2: List of L1 and HLT triggers used for the 2017 dataset, and the channels to which they apply.

Channel	L1 Seeds	HLT Paths
W($\mu\nu$)H	muon with $p_T > 22$ GeV	isolated muon with $p_T > 27$ GeV
Z($\mu\mu$)H	two muons with $p_T > 12(5)$ GeV	two isolated muons with $p_T > 17(8)$ GeV
W(eν)H	electron with $p_T > 38$ GeV OR isolated electron with $p_T > 30$ GeV OR isolated electron with $p_T > 28$ GeV, $\lvert\eta\rvert < 2.1$ OR two electrons with $p_T > 25(12)$ GeV	electron with $p_T > 32$ GeV and passing tight id.
Z(ee)H	electron with $p_T > 30$ GeV OR isolated electron with $p_T > 22$ GeV, $\lvert\eta\rvert < 2.1$ OR isolated electron with $p_T > 24$ GeV OR two electrons with $p_T > 15(10)$ GeV	two electrons with $p_T > 23(12)$ GeV, loose id.
Z($\nu\bar{\nu}$)H	$E_T^{miss} > 110$ GeV OR $H_T^{miss} > 120$ GeV $H_T^{miss} > 110$ GeV AND $H_T > 60$ GeV	($E_T^{miss} > 120$ GeV AND $H_T^{miss} > 120$ GeV) OR ($E_T^{miss} > 120$ GeV AND $H_T^{miss} > 120$ GeV AND $H_T > 60$ GeV)

transerse energy, targeting the high momentum Z boson decaying to two neutrinos. The main trigger used in the Z($\nu\nu$)H channel requires at least 120 GeV of E_T^{miss} or H_T^{miss} (Sec. 5.8) at HLT level and E_T^{miss} at L1 from 100 to 120 GeV, depending on the instantaneous luminosity of the LHC. Note in particular that at L1 the E_T^{miss} and H_T^{miss} must be calculated directly from calorimeter clusters and standalone muon tracks alone. This can lead to significant differences in the value of the missing energy at L1 with respect to the HLT, resulting in significant inefficiencies for events with H_T^{miss} and E_T^{miss} values near the selection thresholds.

The triggers are emulated in the MC, with simulated events required to satisfy the same trigger conditions as those used for data. Differences in trigger efficiencies between data and simulation are corrected for with scale factors derived using a tag-and-probe method with dilepton events from Z boson decays. Tight identification and isolation requirements are imposed for one "tag" lepton, whereas loose requirements are imposed for the second "probe" lepton. After requiring that the dilepton invariant mass is close to the Z boson mass, the sample is very pure in real leptons and the probe lepton can be used to measure the efficiency of the trigger requirements. The trigger efficiency measurement is performed on both data and simulation, and the

resulting differences in the ratio of the data efficiency relative to the efficiency on simulation is applied to the simulation as a function of the lepton p_{T} and η.

The trigger efficiencies are measured with respect to the offline lepton identification and isolation requirements. For the dilepton triggers, the scale factors for each lepton are computed separately because of the different selection requirements. The efficiency correction scale factors are measured to be around 0.97 for single lepton triggers as well as for each lepton of the dilepton triggers.

The overall Z($\nu\nu$)H trigger efficiency is measured in data collected by single electron triggers and additionally requiring the presence of two jets within the tracker acceptance. In order to avoid bias from the L1 $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$, reconstructed from only calorimeter clusters, the lepton is required to not be aligned with the reconstructed $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ in azimuth. The correction applied to simulation for the Z($\nu\nu$)H trigger efficiency is 0.93 at lower values of the offline $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ and roughly unity (no correction) for events with high ($\gtrsim$ 250 GeV) offline $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$.

7.4.2 W boson and $\mathrm{t\bar{t}}$ transverse momentum reweighting

A residual mismodeling of the reconstructed p_{T}(V) is observed in the W($\ell\nu$)H channel for the primary backgrounds, and for $\mathrm{t\bar{t}}$ in all channels, after applying to simulation the corrections described in Chapter 6. This residual mismodeling is expected to be due to higher-order QCD and electroweak corrections not taken into account. Independent linear reweighting functions are derived to correct this effect for $\mathrm{t\bar{t}}$, W + udscg, and the combination of W + $\mathrm{b\bar{b}}$, W + b, and single top via a simultaneous fit to data of the reconstructed p_{T}(W) in the W($\ell\nu$)H background-enriched control regions (Chapter 9). The input PDF for the fit in each control region is a sum of the MC prediction for each process corrected by a linear function of the reconstructed p_{T}(W) with a slope that is allowed to float in the fit. The relative composition of

the fitted processes in each control region is fixed. Table 7.3 lists the fitted slopes for each process as well as the uncertainties from the fit.

Table 7.3: Linear correction factors obtained from a simultaneous fit to the $p_{\mathrm{T}}(\mathrm{W})$ distribution in data in the $\mathrm{W}(\ell\nu)\mathrm{H}$ control regions.

Process	$\mathrm{t\bar{t}}$	$\mathrm{W}+\mathrm{udscg}$	$\mathrm{W}+\mathrm{b\bar{b}}+\text{single top}$
Fitted Slope (/GeV)	0.00061 ± 0.00008	0.00064 ± 0.00004	0.0016 ± 0.0001
Norm.-preserving constant	1.103	1.115	1.337

The $\mathrm{t\bar{t}}$ correction is applied to $\mathrm{t\bar{t}}$ simulation in all channels. It has been verified that the result of the simultaneous fit is not sensitive to changes in the definition of the fitted control regions such as loosening the additional jet multiplicity requirement or adjusting the m_{jj} selection.

The systematic uncertainties on the $p_{\mathrm{T}}(\mathrm{W})$ corrections are taken from the uncertainties on the fitted slopes given by the fit, which take into account statistical uncertainties. This corresponds to a 13% uncertainty on the fitted slope for $\mathrm{t\bar{t}}$ and a 6% uncertainty for both $\mathrm{W}+\mathrm{udscg}$ and the combination of $\mathrm{W}+\mathrm{b\bar{b}}$, $\mathrm{W}+\mathrm{b}$, and single top. These uncertainties are sufficient to cover the residual differences between data and simulation in the $p_{\mathrm{T}}(\mathrm{V})$ distribution after applying the corrections. The overall impact of these uncertainties on the $\mathrm{W}(\ell\nu)\mathrm{H}$ analysis sensitivity is less than 3%, and even smaller for the other channels.

7.5 Validation in data

The effects of the dijet mass resolution improvements described in this chapter are assessed on background events in control regions (Chapter 9) to verify that the jet resolution improves as expected in data. Figure 7.3 shows the ratio of the dijet p_{T} to the dimuon p_{T} in the high-p_{T}, Z+(b)b-enriched control region (defined in Chapter 9) with nominal corrections only (left), after applying the b-jet energy regression (mid-

dle), and after applying both the b-jet energy regression and the kinematic fit (right). The corrections shift the event topology towards improved momentum balance, as expected due to the improved jet momentum resolution. Most importantly, the effect of these improvements on data is consistent with the expectation from simulation.

The dijet invariant mass distributions in all control regions after applying the regression and kinematic fit will be shown in Sec. 9.3. No sculpting of the invariant mass distributions is observed for the backgrounds.

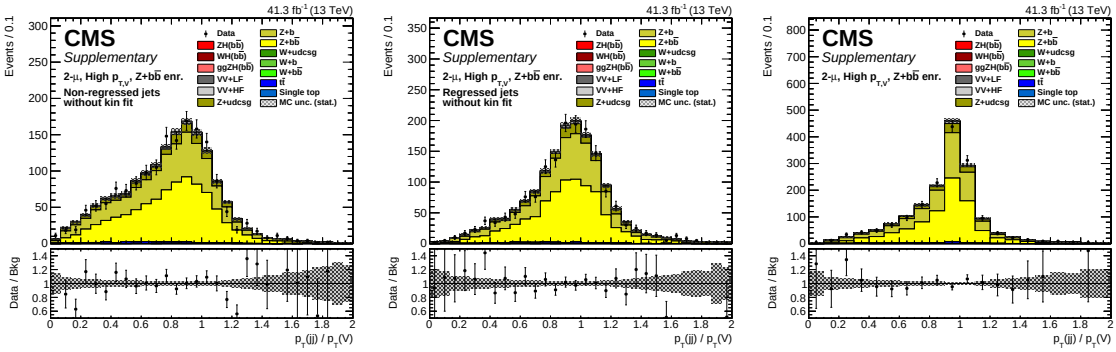


Figure 7.3: The ratio of the of the dijet p_{T} to the dimuon p_{T} in the high-p_{T}, Z+(b)b-enriched control region (Ch. 9) with nominal corrections only (left), after applying the b-jet energy regression (middle), and after applying both the b-jet energy regression and the kinematic fit (right).

Chapter 8

Signal discrimination

Events are loosely pre-selected in each channel, with the selections primarily based on the characteristics of the targeted vector boson decay and also taking into account limitations imposed by the trigger requirements (Sec. 7.4.1). A multivariate classifier is then trained in each channel to distinguish signal from background. This approach ensures high signal efficiency and allows the multivariate classifier to identify the most signal-like events, using not only a set of selections on physical event observables but also exploiting correlations among observables.

8.1 Signal and background characteristics

As mentioned in Section 7.1.1, signal events leave a distinct experimental signature of two high-p_{T} b jets with invariant mass consistent with $m_H = 125$ GeV recoiling against a high-p_{T} vector boson. The vector boson and Higgs boson are expected to be central ($|\eta| < 2.4$) and back-to-back in the transverse plane, such that the distribution of the azimuthal opening angle between the vector boson and Higgs boson candidate peaks at π radians. A large dijet p_{T} is expected more frequently for signal than for the backgrounds, and the transverse momentum distribution for each b jet peaks at roughly $p_{\mathrm{T}} \sim \mathrm{m_H}/2$. Additional jet activity in the event is relatively low for signal

(from ISR and FSR), and isolated leptons not arising from the decay of a W or Z boson are expected to be negligible.

The dominant backgrounds are summarized below. The contribution from each background to the signal regions can be mitigated by exploiting distinctive background features with respect to signal events:

- **V+jets**: production of W and Z bosons in association with one or more jets. This background has a large cross section and has very similar characteristics to the signal, but has a generally a lower p_{T} spectrum and a a sharply falling dijet mass distribution. The contribution from V+udscg events, where the associated jets are not b jets, is much reduced after the application of b-tagging on both Higgs boson candidate jets. In the highest signal purity selected regions of the analysis, residual contributions from $\mathrm{V} + \mathrm{b\bar{b}}$ are nearly indistinguishable from signal and largely dominate the overall background uncertainty.

- **Top quarks:** production of $\mathrm{t\bar{t}}$ pairs, as well as single top quarks in the tW, t-channel, and s-channel processes. These background processes include one or two real W bosons and at least one or two b jets, with intrinsic mass and momentum scales close to the Higgs boson mass scale. The $\mathrm{t\bar{t}}$ background is particularly challenging due to its cross section three orders of magnitude larger than signal and the presence of two real b jets and at least one real $\mathrm{W}(\ell\nu)$ decay. In the $\mathrm{Z}(\ell\ell)\mathrm{H}$ channel the $\mathrm{t\bar{t}}$ background can be largely suppressed by requiring $m_{\ell\ell}$ consistent with the Z boson mass. The $\mathrm{t\bar{t}}$ process is however a leading background contribution in the $\mathrm{W}(\ell\nu)\mathrm{H}$ and $\mathrm{Z}(\nu\nu)\mathrm{H}$ channels. In these channels, the $\mathrm{t\bar{t}}$ background can often be differentiated from signal by its higher additional jet multiplicity, from the hadronic decay of one of the W bosons. The single top processes have a smaller cross section and contain only one b quark at tree level. The single top contribution is therefore highly reduced by requiring b-tagging for both Higgs boson candidate jets.

- **Dibosons (WW, WZ, ZZ)**: The production of vector boson pairs, despite relatively small cross sections compared to other backgrounds, is particularly difficult to distinguish from signal in the case that one Z boson decays to $b\bar{b}$ and the other W or Z boson decays leptonically (VZ, $Z \to b\bar{b}$). The momentum scale for these diboson processes is typically higher than other backgrounds but slightly lower than signal. The most distinctive feature of diboson events is the difference in dijet invariant mass between $H \to b\bar{b}$ and $Z \to b\bar{b}$. This difference in dijet mass is only a factor two to three greater than the experimental dijet mass resolution. Good mass resolution is therefore critical in order to separate the VZ, $Z \to b\bar{b}$ background from signal. The kinematic similarities between this background and signal, including the presence of a resonance in the m_{jj} spectrum at a similar scale to signal, is exploited by performing a measurement of the VZ, $Z \to b\bar{b}$ process with an analysis that very closely parallels the nominal analysis and therefore further validates the analysis procedure (Sec. 10.4.1, App. A.1).

- **QCD multijet**: As previously discussed, the enormous rate of LHC collision events with strong interactions which produce multiple bottom quarks is very strongly suppressed by selecting events with a high-p_T vector boson decaying to leptons. Some QCD multijet events can yield a large reconstructed E_T^{miss} resulting from a large mismeasurement of the jet momentum or by the emission of high-p_T neutrinos from hadron decays. This residual QCD contribution is particularly relevant for the $Z(\nu\bar{\nu})$ channel, which selects the Z boson candidate purely based on the presence of large E_T^{miss} in the event. It is, however, highly suppressed by vetoing events where $\vec{E}_T^{miss}$ is closely aligned with a high-p_T jet, a characteristic feature of events with mismeasured jet momentum.

The following variables have been identified as useful in discriminating signal from background:

- $\boldsymbol{m_{jj}}$: dijet invariant mass; peaks at m_H for VH, $H \to b\bar{b}$ signal and M_Z for VZ, $Z \to b\bar{b}$ diboson events, falls sharply for V+jets, and peaks broadly over the region 100–160 GeV for $t\bar{t}$ events.

- $\boldsymbol{p_T(jj)}$: transverse momentum of the Higgs boson candidate; signal events have a higher p_T spectrum than backgrounds.

- $\boldsymbol{p_T(V)}$: vector boson transverse momentum, as defined in Sec. 7.2; signal events have a higher p_T spectrum than backgrounds.

- $\boldsymbol{\Delta\phi(V, H)}$: azimuthal opening angle between the vector boson and the Higgs boson candidate. For signal, the distribution of $\Delta\phi(V, H)$ peaks at π radians (V and H recoiling and back-to-back), whereas for backgrounds there is typically no preferential direction.

- **b-tagging discriminant**: output of the b-tagging discriminant (DeepCSV, Sec. 5.6) for the Higgs boson candidate jets; considered separately for the jet with the higher value ($\text{DeepCSV}_{\text{max}}$), and the jet with the lower value ($\text{DeepCSV}_{\text{min}}$).

- $\boldsymbol{N_{aj}}$: number of additional jets in the event apart from the Higgs boson candidate b jets. Only central jets with $|\eta| < 2.5$ are considered, with p_T thresholds and multiplicity requirements optimized separately for each channel. As mentioned in the previous section, this variable is highly effective in reducing the large $t\bar{t}$ contributions to the $W(\ell\nu)H$ and $Z(\nu\nu)H$ channels.

- $\boldsymbol{M_t}$: reconstructed top mass for events with a $W(\ell\nu)$ candidate and a nearby b jet (Sec. 8.3.1).

- $\boldsymbol{p_{T_j}}$: transverse momentum of the Higgs boson candidate b jets.

- $\boldsymbol{\Delta\eta(jj)}$: pseudorapidity difference between the two Higgs boson candidate b jets.

- $\boldsymbol{\Delta\varphi(jj)}$: azimuthal opening angle between the two Higgs boson candidate b jets.

- $\boldsymbol{\Delta R(jj)}$: ΔR (Sec. 4.2.2) between the two Higgs boson candidate b jets.

- $\boldsymbol{N_{al}}$: number of additional isolated leptons, apart from those associated with the W or Z boson decay. Only leptons satisfying $p_T > 20$ GeV and $|\eta| < 2.5$ are considered.

- $\mathbf{E_T^{miss}}$: event transverse momentum imbalance, as defined in Sec. 5.8.

- $\boldsymbol{\Delta\phi(\vec{E}_T^{miss}, j)}$: azimuthal opening angle between the $\vec{E}_T^{miss}$ and the closest central jet in azimuth. Only jets satisfying $p_T > 30$GeV and $|\eta| < 2.5$ are considered. As mentioned in the previous section, this variable highly reduces the residual QCD background in the Z($\nu\nu$)H channel, where the E_T^{miss} typically arises from large mismeasurement of the momentum of a single jet.

- $\boldsymbol{\Delta\phi(\vec{E}_T^{miss}, lepton)}$: azimuthal opening angle between the $\vec{E}_T^{miss}$ and the leading-p_T lepton.

- $\boldsymbol{\Delta\phi(\vec{E}_T^{miss}, \text{track-only } \vec{E}_T^{miss})}$: the missing transverse energy direction is calculated considering only tracks rather than all PF candidates (track-only $\vec{E}_T^{miss}$) and compared with the nominal $\vec{E}_T^{miss}$ direction. The consistency of the direction of these two calculations of the $\vec{E}_T^{miss}$ is used in the Z($\nu\nu$)H channel to reduce contamination from events with large E_T^{miss} due to PU.

- $\mathbf{DeepCSV_{max,aj}}$: maximum b-tagging discriminant value for the additional jets in the event. This variable helps reduce the $t\bar{t}$ background in the Z($\nu\nu$)H and W($\ell\nu$)H channels.

- $\boldsymbol{N_5^{\text{soft}}}$: number of additional soft track-jets with $p_{\text{T}} > 5$ GeV, as defined in Sec. 5.9.

8.2 Signal region pre-selection

As mentioned previously in this section, relatively loose selections are applied for each vector boson decay channel in order to reduce backgrounds and exclude regions where the recorded data is limited due to trigger requirements. The kinematic regions with the highest signal purity are then identified with multivariate classifiers. Table 8.1 lists the pre-selection requirements for each channel.

Table 8.1: Signal region pre-selection cuts for each channel. The values listed for kinematic variables are in units of GeV.

Variable	Z($\nu\nu$)H	W($\ell\nu$)H	Z($\ell\ell$)H
p_{T}(V)	> 170	> 150	$[50-150], > 150$
$m_{\ell\ell}$	–	–	$[75-105]$
p_{T}^{ℓ}	–	$(> 25, > 30)$	> 20
$p_{\text{T}}(j_1)$	> 60	> 25	> 20
$p_{\text{T}}(j_2)$	> 35	> 25	> 20
p_{T}(jj)	> 120	> 100	–
m_{jj}	$[60-160]$	$[90-150]$	$[90-150]$
$\text{DeepCSV}_{\text{max}}$	>Tight	>Tight	>Loose
$\text{DeepCSV}_{\text{min}}$	>Loose	>Loose	>Loose
N_{aj}	–	< 2	–
N_{al}	$= 0$	$= 0$	–
$\text{E}_{\text{T}}^{\text{miss}}$	> 170	–	–
$\Delta\phi$(V, H)	> 2.0	> 2.5	> 2.5
$\Delta\phi(\vec{\text{E}}_{\text{T}}^{\text{miss}}$, track-only $\vec{\text{E}}_{\text{T}}^{\text{miss}})$	< 0.5	–	–
$\Delta\phi(\vec{\text{E}}_{\text{T}}^{\text{miss}}$, lepton)	–	< 2.0	–

8.3 Multivariate discriminator

New for this analysis is the use of a deep neural network to distinguish the VH, $\text{H} \to \text{b}\bar{\text{b}}$ signal events from SM backgrounds. Deep neural networks are particularly

well suited to take advantage of large training datasets and high-level correlations between observables.

8.3.1 Input variables

The set of input variables used to train the DNN is optimized separately in each channel, and summarized in Table 8.2. The Higgs boson candidate observables are considered after applying the b-jet energy regression and FSR recovery for all analysis categories, and after the kinematic fit for the Z($\ell\ell$) category.

Table 8.2: List of input variables used in the training of the multivariate discriminators for each channel.

Variable	Description	Z($\nu\nu$)H	W($\ell\nu$)H	Z($\ell\ell$)H
m_{jj}	dijet invariant mass	✓	✓	✓
$p_{\mathrm{T}}(\mathrm{jj})$	dijet transverse momentum	✓	✓	✓
$p_{\mathrm{T}}(\mathrm{j}_1)$, $p_{\mathrm{T}}(\mathrm{j}_2)$	transverse momentum of each jet	✓		✓
$\Delta R(\mathrm{jj})$	distance in η–ϕ between jets			✓
$\Delta\eta(\mathrm{jj})$	difference in η between jets	✓		✓
$\Delta\varphi(\mathrm{jj})$	azimuthal angle between jets	✓		
$p_{\mathrm{T}}(\mathrm{V})$	vector boson transverse momentum		✓	✓
$\Delta\phi(\mathrm{V},\mathrm{H})$	azimuthal angle between vector boson and dijet directions	✓	✓	✓
$p_{\mathrm{T}}(\mathrm{jj})/p_{\mathrm{T}}(\mathrm{V})$	p_{T} ratio between dijet and vector boson			✓
M_{Z}	reconstructed Z boson mass			✓
$\mathrm{DeepCSV}_{\mathrm{max}}$	value of the b-tagging discriminant (DeepCSV) for the jet with highest score	✓		✓
$\mathrm{DeepCSV}_{\mathrm{min}}$	value of the b-tagging discriminant (DeepCSV) for the jet with second highest score	✓	✓	✓
$\mathrm{DeepCSV}_{\mathrm{max,aj}}$	value of b-tagging discriminant for the additional jet with highest value	✓		
$\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$	missing transverse momentum	✓	✓	✓
$\Delta\phi(\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}},\mathrm{j})$	azimuthal angle between $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ and closest jet with $p_{\mathrm{T}} > 30$GeV	✓		
$\Delta\phi(\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}},\ell)$	azimuthal angle between $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ and lepton		✓	
$M_{\mathrm{T}}(W)$	W boson candidate transverse mass		✓	
M_{t}	reconstructed top quark mass		✓	
N_{aj}	number of additional jets		✓	✓
$p_{\mathrm{T}}(\mathrm{aj})$	transverse momentum of leading additional jet	✓		
N_5^{soft}	number of soft-track jets with $p_{\mathrm{T}} > 5$GeV	✓	✓	✓

Reconstructed top mass in the W($\ell\nu$)H channel

In the W($\ell\nu$) channel, signal events are characterized by the presence of a well-isolated lepton, $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ and two b jets. The same signature arises in semileptonic $\mathrm{t\bar{t}}$ events, where one W boson decays leptonically and the other W boson decays hadronically. Vetos

on additional jet activity reduce the $t\bar{t}$ contribution in the signal region, however in the W($\ell\nu$)H channel the remaining $t\bar{t}$ contribution is significant.

Several variables were analyzed to further discriminate against $t\bar{t}$, and the reconstructed top mass was found to be the most powerful. The unknown neutrino longitudinal momentum (p_Z) can be analytically solved for using the relation

$$M_W^2 = (E_\nu + E_\ell)^2 - (\overrightarrow{p_\nu} + \overrightarrow{p_\ell})^2, \tag{8.1}$$

constraining the W boson candidate mass to the world-average value [63] and assuming that the neutrino p_{T} is equal to $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$. There are in general two solutions to this equation. When both are real, the solution with the smaller p_Z is selected. When the solutions are complex numbers, the neutrino transverse momentum vector is minimally adjusted within uncertainties to give a single physical (real) solution.

The resulting energy-momentum four-vectors of the neutrino, the lepton, and the closest b jet are added to form a four-vector for the top quark candidate, from which the mass is computed. This method including the neutrino p_Z improves the resolution on the reconstructed top mass for $t\bar{t}$ events without biasing the distribution for signal.

8.3.2 Training

A DNN is trained in each channel to distinguish signal events from the sum of SM backgrounds (V+jets, VV, $t\bar{t}$, and single top). In the Z($\ell\ell$)H channel, the training is performed inclusively on electron and muon events. It is important to maintain statistically independent samples for the training of the DNN and for the comparison of the DNN score with data to avoid possible biases and to validate the DNN performance on an independent sample. The simulated events are therefore split randomly into two equally sized samples, one for the DNN training and the other for the DNN evaluation and performance tests.

Z($\ell\ell$)H DNN input variables

Figure 8.1 shows the distributions of the DNN input variables for signal (blue) and the sum of SM backgrounds (red) in the Z($\ell\ell$)H high-p_{T} category. The same set of variables are used to train a DNN classifier in the Z($\ell\ell$)H low-p_{T} category.

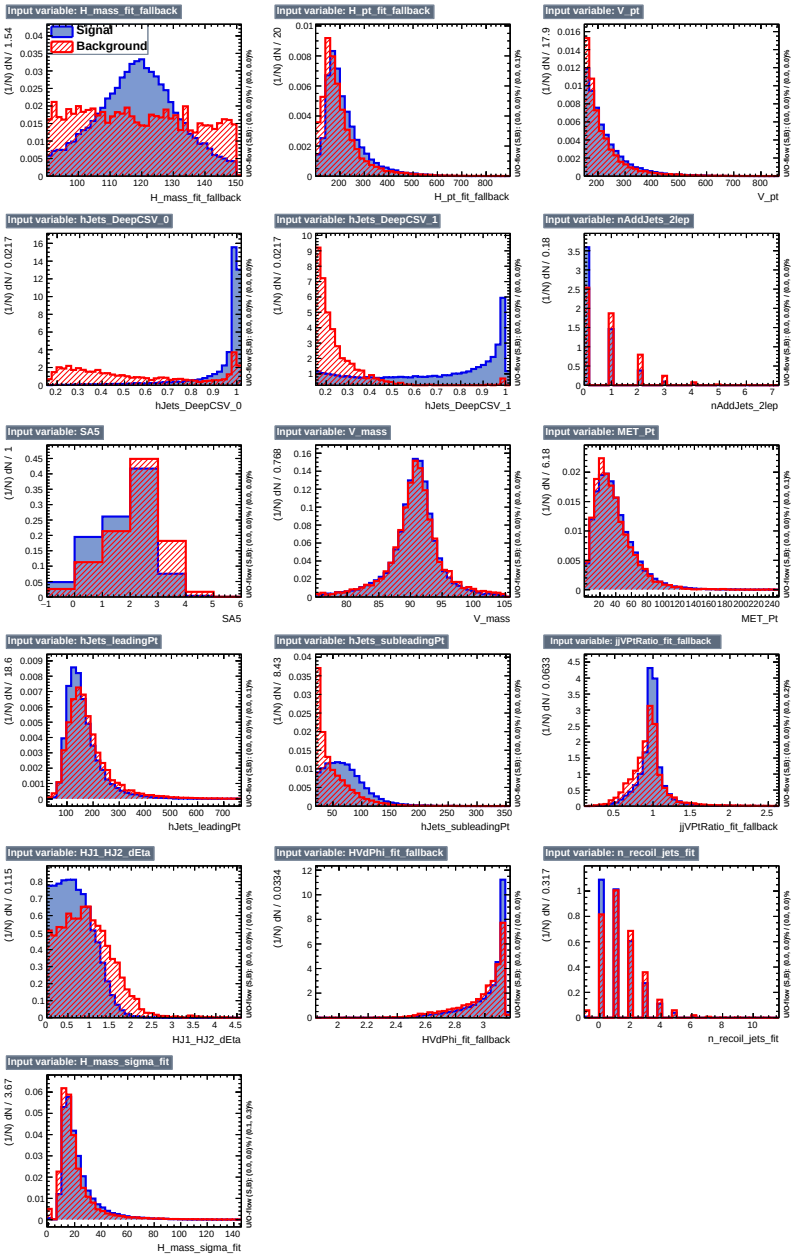


Figure 8.1: Input variables for the Z($\ell\ell$)H high p_{T}(V) DNN training

W($\ell\nu$)H DNN input variables

The distributions of the training input variables for the DNN classifier in the W($\ell\nu$) channel are shown in Fig. 8.2 for signal (blue) and the sum of SM backgrounds (red).

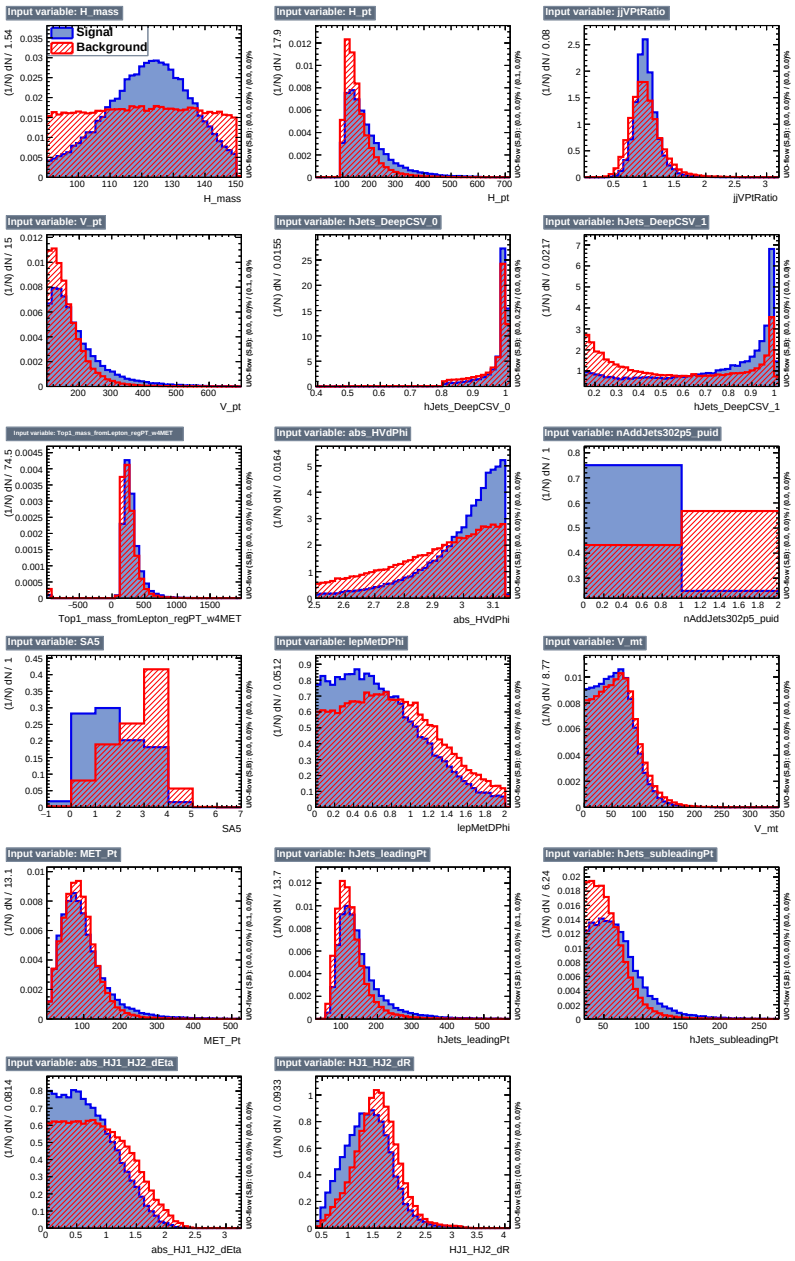


Figure 8.2: Input variables for the W($\ell\nu$)H DNN training

Z($\nu\nu$)H DNN input variables

The distributions of the training input variables for the DNN classifier in the Z($\nu\nu$)H channel are shown in Fig. 8.3 for signal (blue) and the sum of SM backgrounds (red).

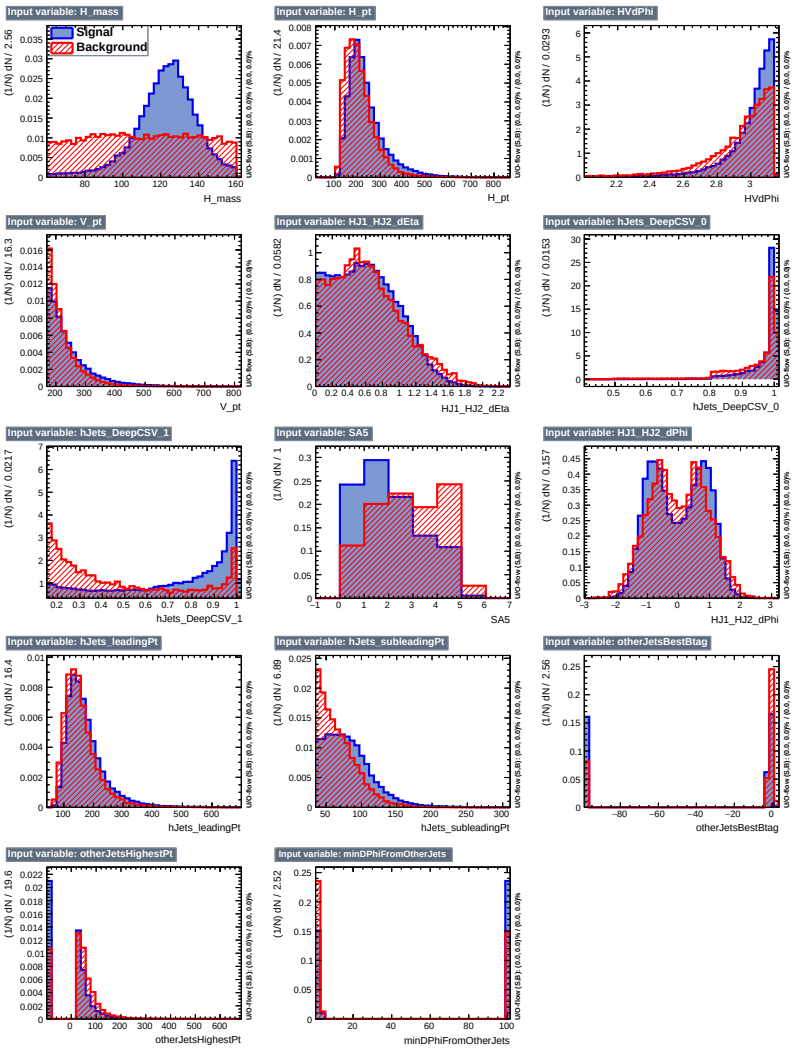


Figure 8.3: Set of input variables for the Z($\nu\nu$)H DNN training

8.3.3 DNN architecture and hyper-parameter optimization

Deep neural networks are implemented with the python library Keras, which is interfaced with TensorFlow to perform matrix calculations. The number of simulated events used in each DNN training is about 200k, evenly divided between signal and background. The implementation of the network has been tested in both Keras and pure TensorFlow in order to customize the loss function. Graphics processing units (GPUs) are used for training (Tesla P100 at CSCS T2, GTX 970), requiring 5-10 minutes for each training. The training with GPUs is faster than with CPUs by a factor of roughly 25-30.

The chosen architecture consists of 5 hidden layers (configurations with 2 to 5 hidden layers were tested). The number of nodes for each layer is 32 (8- to 512-node configurations were tested). A sketch of the architecture is shown in Fig. 8.4. At each

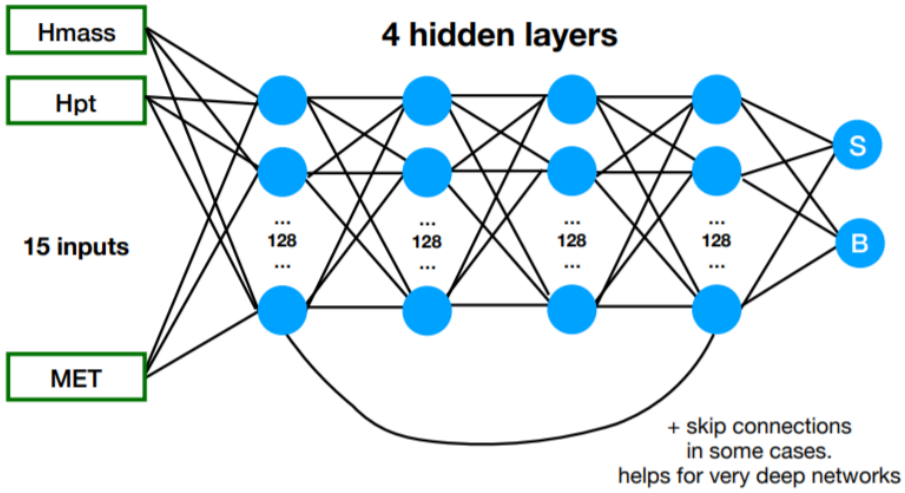


Figure 8.4: Sketch of the DNN architecture

layer, dropout (i.e. during the learning phase the weights of random nodes are set to 0) is used to ensure regularization and make the network more robust. The leaky ReLU activation function is used, although no significant difference in performance

was observed when trying several different activation functions. Two loss functions were compared: cross-entropy and median significance.

The cross-entropy loss is defined as

$$L_{\text{cross-entropy}} \equiv -(y \log(p) + (1-y) \log(1-p)), \tag{8.2}$$

where y is the true label (0 for background and 1 for signal) and p is the predicted signal-like probability (the DNN output score). The cross-entropy loss is equivalent to the log-likelihood for the data y under model p assuming a binomial distribution. The cross-entropy loss assures fast convergence but is not ideal for signal extraction in the case of a low signal to background ratio ($S << B$), where careful rebinning of the DNN output is required.

The counting experiment median significance, otherwise known as the "Asimov-like" or "Bin-aware" loss function, is defined as

$$L_{\text{Azimov-like}} \equiv \text{med}[\text{Z}_0|1] = \sqrt{\text{q}_{0,\text{A}}} = \sqrt{2((\text{s}+\text{b})\ln(1+\text{s}/\text{b})-\text{s})} \tag{8.3}$$

and uses fixed binning in the training. The bins are approximated by smooth kernel functions in order to obtain a fully differentiable loss function. For both loss function choices, the loss function is minimized with the *Adam* algorithm [64].

To estimate the performance of each DNN training, a simple figure of merit is used: $S/\sqrt{S+B}$, summed in quadrature over 15 bins of the discriminator output. The difference in performance between the training and the test set provides a metric to compare DNN performance variance and stability. To further refine the hyper-parameters, checks are performed choosing different dropout factors. It has been observed that using dropout reduces the performance difference between the training and testing sets thereby providing more robust results.

The performance as a function of learning rate has been tested and found to give no significant improvement with respect to the default value of 0.001. A thorough comparison of the performance of the two loss functions requires an optimization of the hyper-parameters in the two cases, as well as a careful choice of the binning. Several studies show that the two loss functions achieve similar performance. For simplicity, the cross-entropy loss is chosen for the final discriminator. This choice of loss function requires a rebinning of the discriminator output, which will be described in Section 8.3.5.

8.3.4 Validation

A comparison of the DNN performance on the training and testing datasets as a function of the training epoch is shown in Fig. 8.5 for the Z($\nu\bar{\nu}$) channel. The performance plateaus at a high number of training epochs, as expected, and the DNN performs similarly on the testing and training datasets.

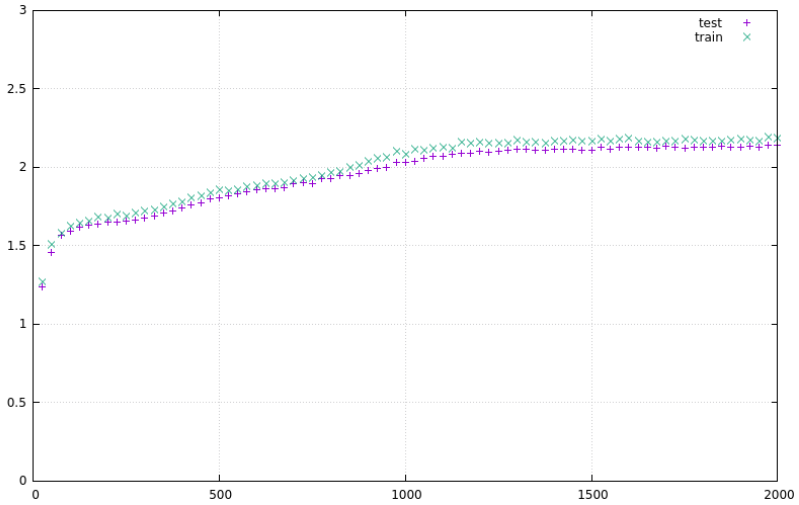


Figure 8.5: A comparison of the training and testing performance in the Z($\nu\nu$)H channel as a function of the training epoch.

The signal and background distributions for the DNN classifier in each channel are shown in Fig. 8.6.

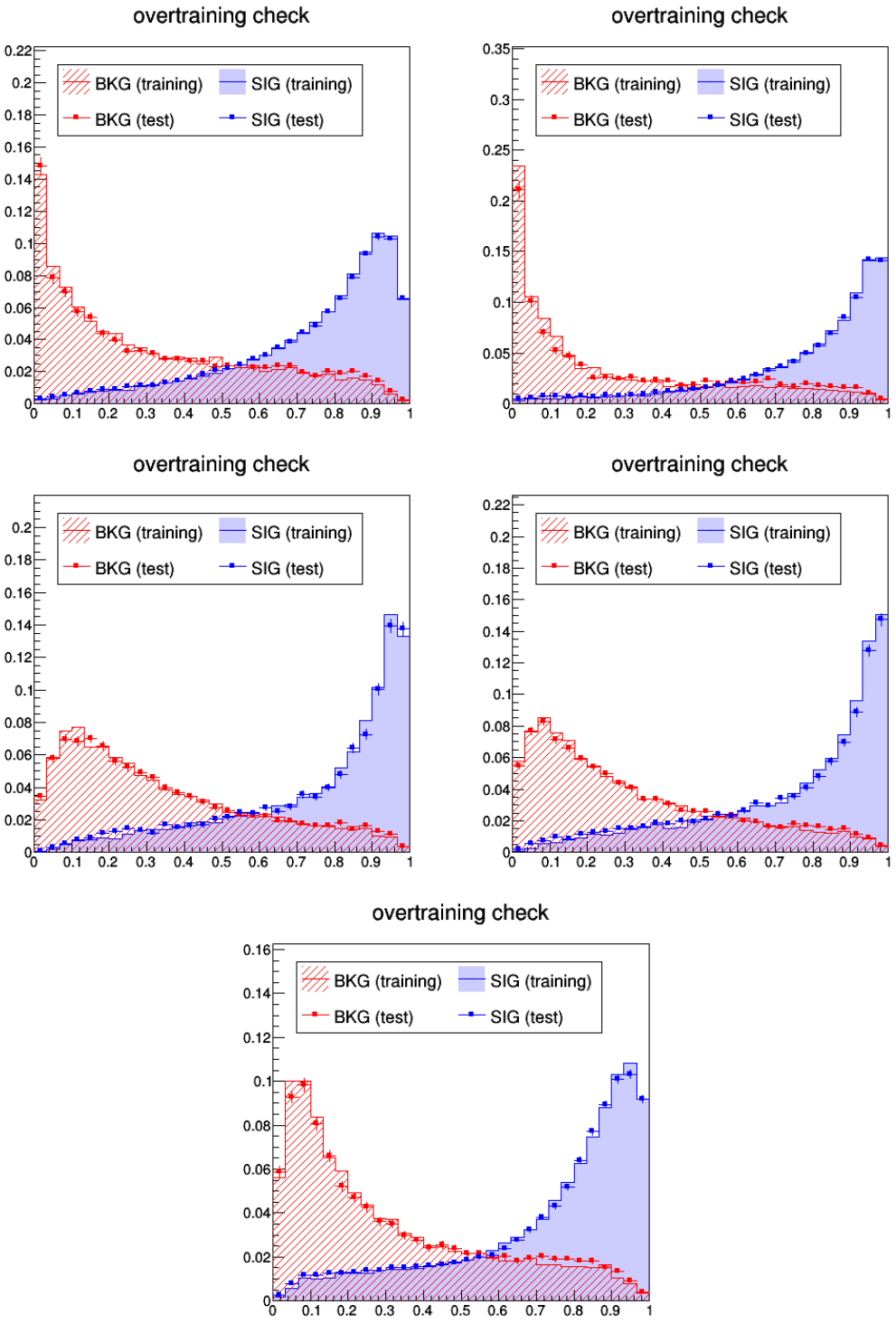


Figure 8.6: DNN output for signal (blue) and background (red) simulation. Top row: Z($\ell\ell$)H low-p_{T} (left) and high-p_{T} (right). Middle row: W($\ell\nu$)H electron channel (left) and muon channel (right). Bottom row: Z($\nu\nu$)H channel.

8.3.5 DNN reshaping

When considering 15 equidistant bins between 0 and 1 for the raw DNN score, most of the events are distributed in the first bin. A rebinning is applied to distribute signal and background over multiple bins with different S/B. This rebinning improves the ability to resolve the signal contribution from background with finite binning. The raw DNN score is transformed with the function $\frac{\sqrt{x}+x^{12}}{2}$ (Fig. 8.7, left) and rebinned in 15 equidistant bins of the transformed DNN score. This is equivalent to rebinning the raw DNN score into 15 variable-sized bins, as shown in Fig. 8.7 (right). The distribution of the transformed DNN score over 15 equidistant bins is used to extract the $H \to b\bar{b}$ signal, as will be described in Chapter 10. The same rebinning function is used in all channels.

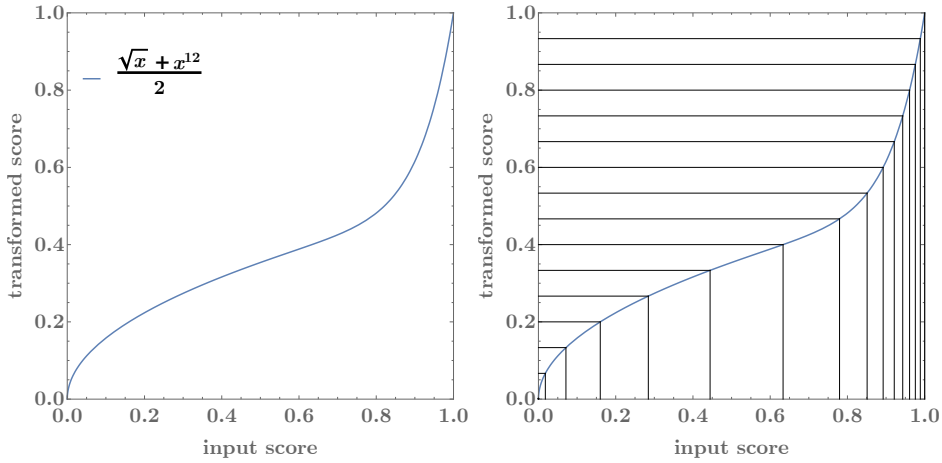


Figure 8.7: Derivation of the DNN rebinning. The left plot shows the transformation function and the right plot shows the obtained binning.

Chapter 9

Background normalization and validation

Control regions (CR) are used to validate in data the modeling of the analysis observables. The control regions are also used to adjust the normalization predictions from simulation for the production of a W or Z boson in association with jets as well as $t\bar{t}$. The CR are designed to be mutually exclusive and orthogonal to the signal regions (SR), while also minimizing extrapolation uncertainties from the CR to SR.

The following sections show comparisons between data and simulation for the analysis observables in the control regions in each channel. The agreement between the simulation and the data in these distributions gives confidence in the background modeling, such that the background predictions can be used as templates in the signal extraction fits, as will be described in Section 10.1. Note that for brevity this section does not show control region distributions for the $Z(\ell\ell)H$ low-p_T category or separate plots for the electron and muon channels. The distributions in these regions have however also been confirmed to match well the data.

9.1 Z($\nu\nu$)H control regions

The Z($\nu\nu$)H channel is characterized by large missing transverse energy, $\mathrm{E_T^{miss}} > 170$ GeV. A set of control regions are designed to target maximal purity in each of the primary backgrounds: $\mathrm{Z} + \mathrm{udscg}$, $\mathrm{Z} + \mathrm{b\bar{b}}$, and $\mathrm{t\bar{t}}$. The selections used to define the control regions are reported in Table 9.1. For comparison, the signal region selection is reported in Table 8.1.

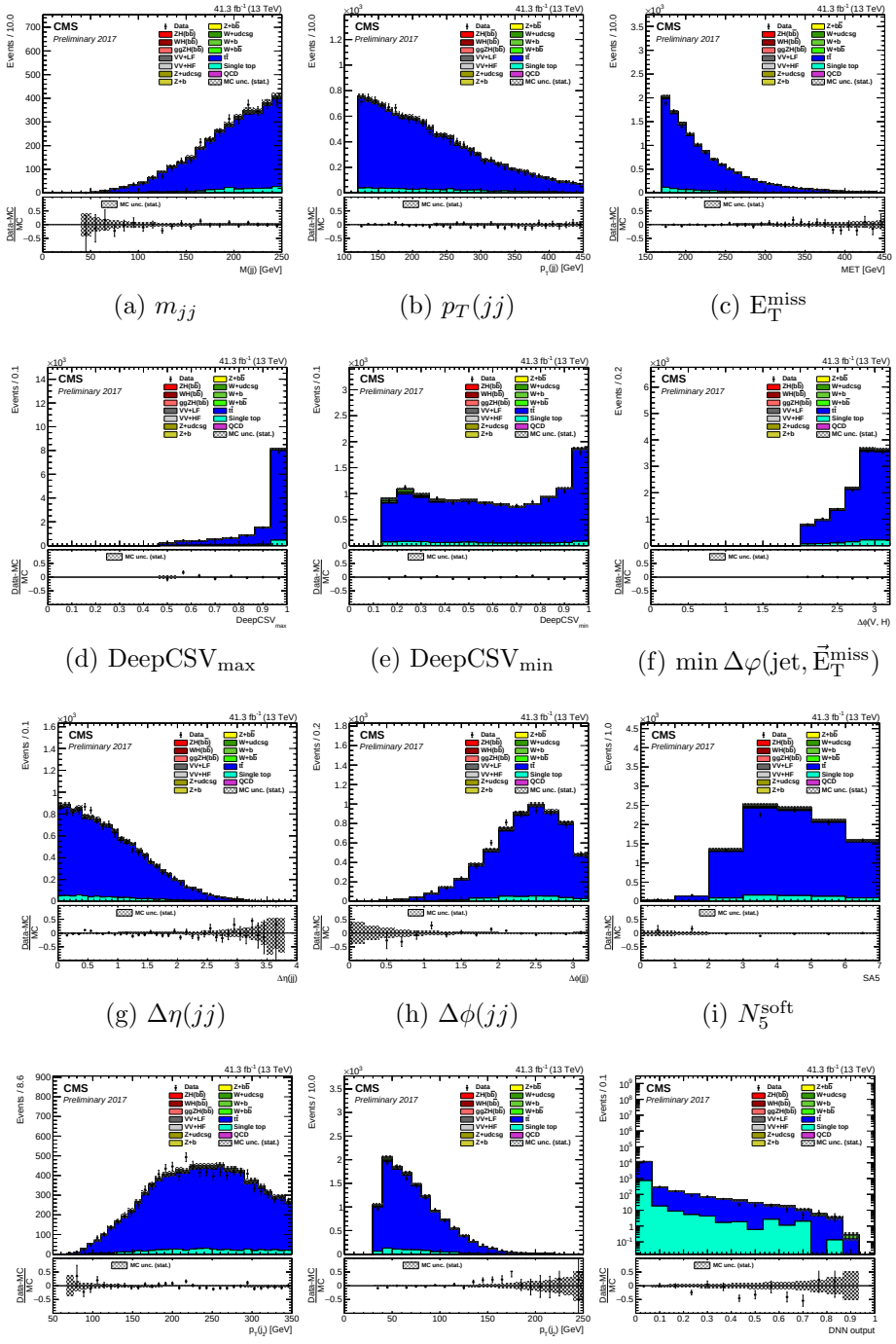
- The $\mathrm{Z} + \mathrm{b\bar{b}}$-enriched control region is the most similar to the signal region, except that the m_{jj} selection is inverted to veto events consistent with m_H.
- The Z + light-jets control region is defined by inverting the b-tagging requirements on the Higgs boson candidate jets and removing the m_{jj} selection. The remaining cuts are identical to the $\mathrm{Z} + \mathrm{b\bar{b}}$-enriched control region.
- The $t\bar{t}$ control region is defined by requiring at least two additional jets (other than the two Higgs boson candidate b jets) with $p_\mathrm{T} > 30$ GeV, at least one b-tagged jet, and at least one isolated lepton.

The data is compared with simulation for the Z($\nu\nu$)H analysis observables in Figures 9.1–9.3.

Table 9.1: Definition of the control regions for the Z($\nu\nu$)H channel. The values listed for kinematic variables are in units of GeV.

Variable	$\mathrm{t\bar{t}}$	Z + udscg	$\mathrm{Z + b\bar{b}}$
$p_{\mathrm{T}}(j_1)$	> 60	> 60	> 60
$p_{\mathrm{T}}(j_2)$	> 35	> 35	> 35
$p_{\mathrm{T}}(\mathrm{jj})$	> 120	> 120	> 120
$\mathrm{E_T^{miss}}$	> 170	> 170	> 170
$\Delta\phi(\mathrm{V,H})$	> 2	> 2	> 2
N_{al}	≥ 1	$= 0$	$= 0$
N_{aj}	≥ 2	≤ 1	≤ 1
$M(\mathrm{jj})$	–	–	$\notin [60 - 160]$
$\mathrm{DeepCSV_{max}}$	>Medium	<Medium	>Tight
$\mathrm{DeepCSV_{min}}$	>Loose	Loose	>Loose
$\Delta\varphi(\mathrm{j,E_T^{miss}})$	–	> 0.5	> 0.5
$\Delta\varphi(\mathrm{tkMET,E_T^{miss}})$	–	< 0.5	< 0.5
$\min\Delta\varphi(\mathrm{j_{1/2},E_T^{miss}})$	$< \pi/2$	–	–

Comparisons of the data with the prediction from simulation are shown in Fig. 9.1 for the $t\bar{t}$-enriched control region in the Z($\nu\nu$)H channel.



(a) m_{jj} (b) $p_T(jj)$ (c) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$

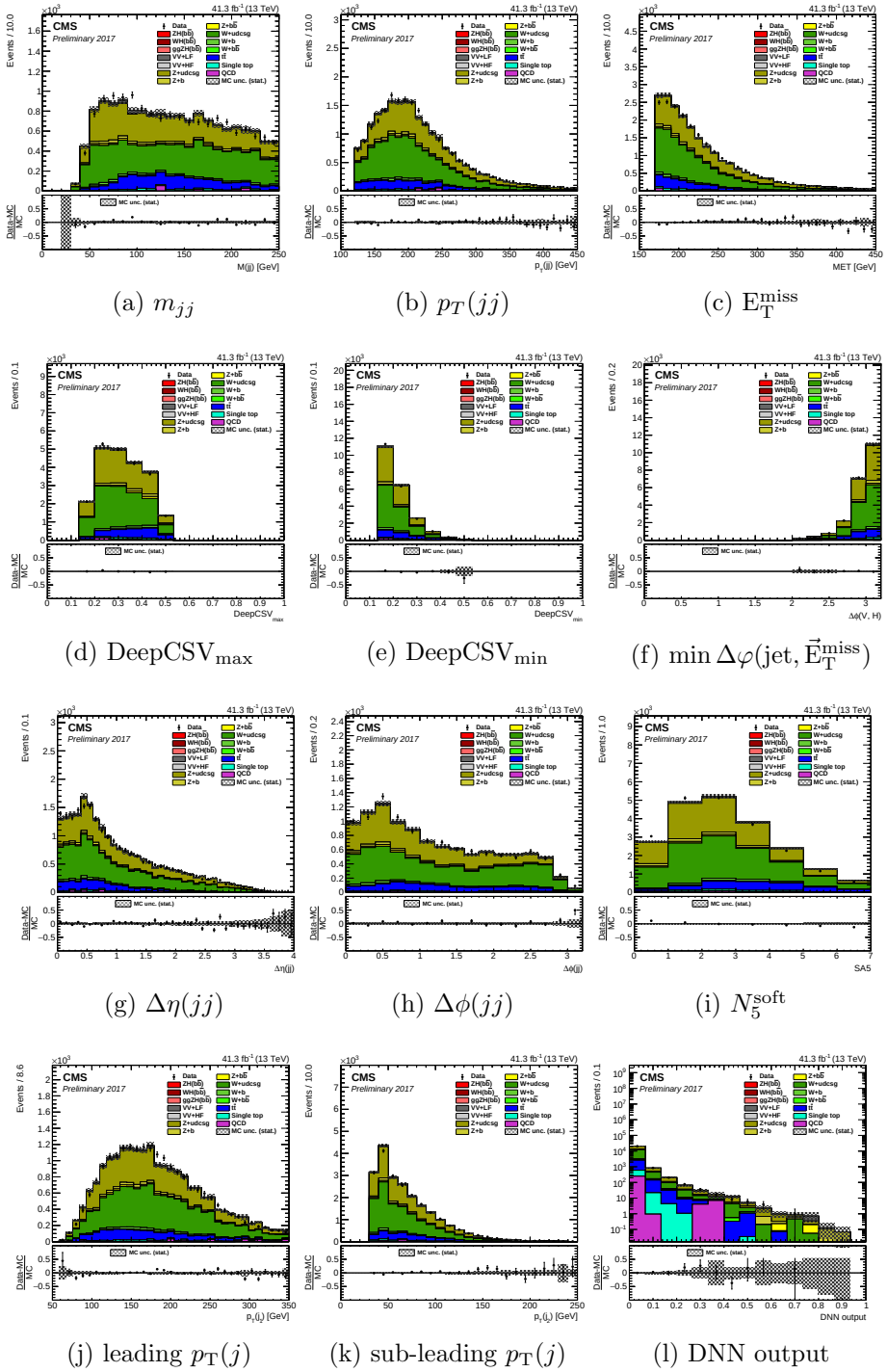
(d) $\mathrm{DeepCSV}_{\mathrm{max}}$ (e) $\mathrm{DeepCSV}_{\mathrm{min}}$ (f) $\min \Delta\varphi(\mathrm{jet}, \vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}})$

(g) $\Delta\eta(jj)$ (h) $\Delta\phi(jj)$ (i) N_5^{soft}

(j) leading $p_{\mathrm{T}}(j)$ (k) sub-leading $p_{\mathrm{T}}(j)$ (l) DNN output

Figure 9.1: Analysis observables for the $t\bar{t}$-enriched control region in the Z($\nu\nu$)H channel.

Comparisons of the data with the prediction from simulation are shown in Fig. 9.2 for the Z+light-jets control region in the Z($\nu\nu$)H channel.



(a) m_{jj} (b) $p_T(jj)$ (c) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$

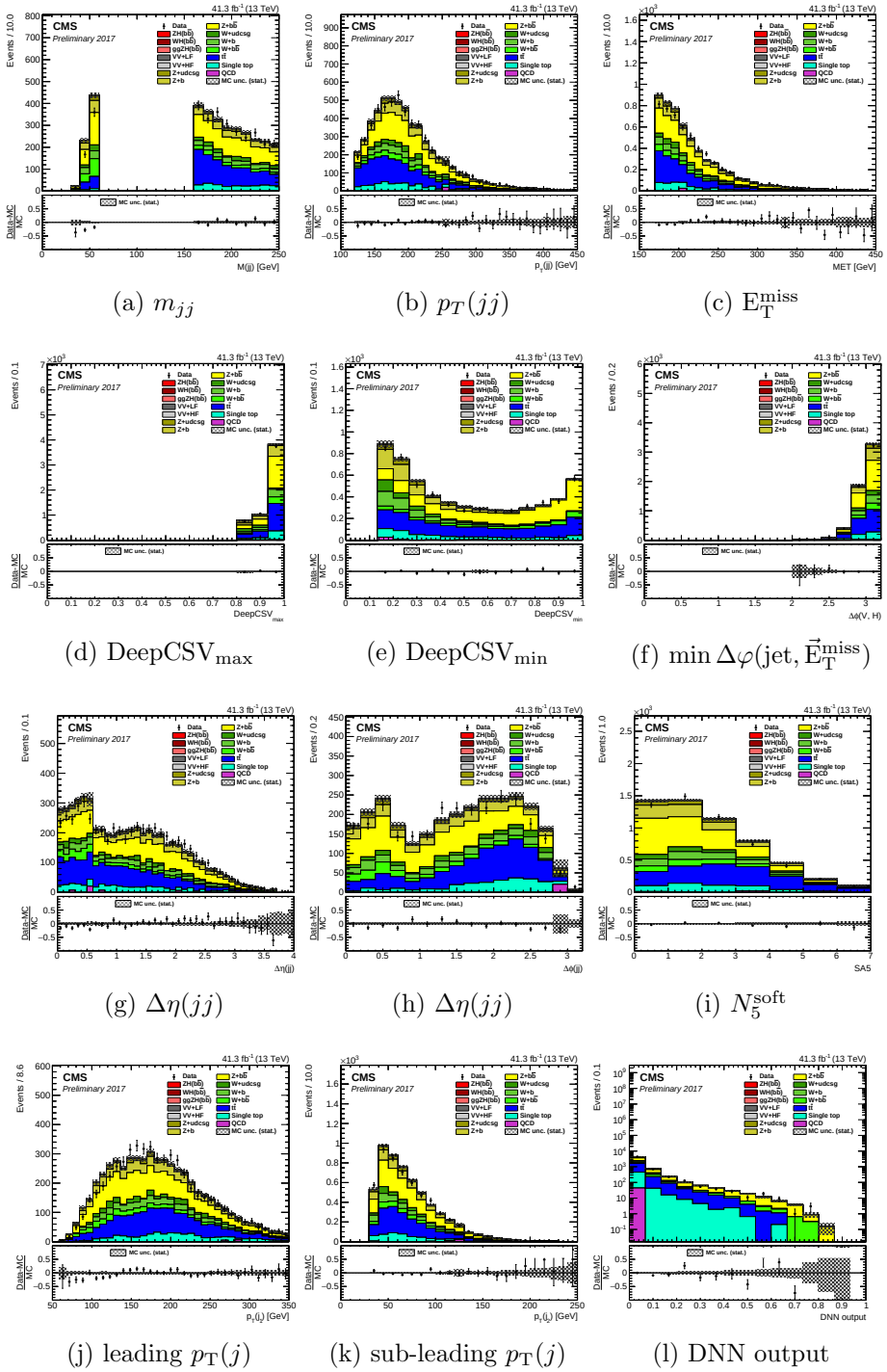
(d) $\mathrm{DeepCSV}_{\mathrm{max}}$ (e) $\mathrm{DeepCSV}_{\mathrm{min}}$ (f) $\min \Delta\varphi(\mathrm{jet}, \vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}})$

(g) $\Delta\eta(jj)$ (h) $\Delta\phi(jj)$ (i) N_5^{soft}

(j) leading $p_{\mathrm{T}}(j)$ (k) sub-leading $p_{\mathrm{T}}(j)$ (l) DNN output

Figure 9.2: Analysis observables for the Z+light-jets control region in the Z($\nu\nu$)H channel.

Comparisons of the data with the prediction from simulation are shown in Fig. 9.3 for the $\mathrm{Z} + \mathrm{b\bar{b}}$-enriched control region in the $\mathrm{Z}(\nu\nu)\mathrm{H}$ channel.



(a) m_{jj} (b) $p_T(jj)$ (c) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$

(d) $\mathrm{DeepCSV}_{\mathrm{max}}$ (e) $\mathrm{DeepCSV}_{\mathrm{min}}$ (f) $\min \Delta\varphi(\mathrm{jet}, \vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}})$

(g) $\Delta\eta(jj)$ (h) $\Delta\eta(jj)$ (i) N_5^{soft}

(j) leading $p_{\mathrm{T}}(j)$ (k) sub-leading $p_{\mathrm{T}}(j)$ (l) DNN output

Figure 9.3: Analysis observables for the $\mathrm{Z} + \mathrm{b\bar{b}}$-enriched control region in the $\mathrm{Z}(\nu\nu)\mathrm{H}$ channel.

9.2 W($\ell\nu$)H control regions

The W($\ell\nu$)H channel considers three control regions, targeting the W + udscg, W + $b\bar{b}$, and $t\bar{t}$ backgrounds. The selection criteria, common for the electron and muon channels, are summarized in Table 9.2.

- The W + $b\bar{b}$-enriched control region is similar to the signal region, except that the b-tagging requirement on the second Higgs boson candidate b jet is removed and the m_{jj} selection is inverted.

- The W+light-jets control region is defined by inverting the b-tagging requirements on the Higgs boson candidate jets, to enhance the light flavor (udcsg) jet contribution.

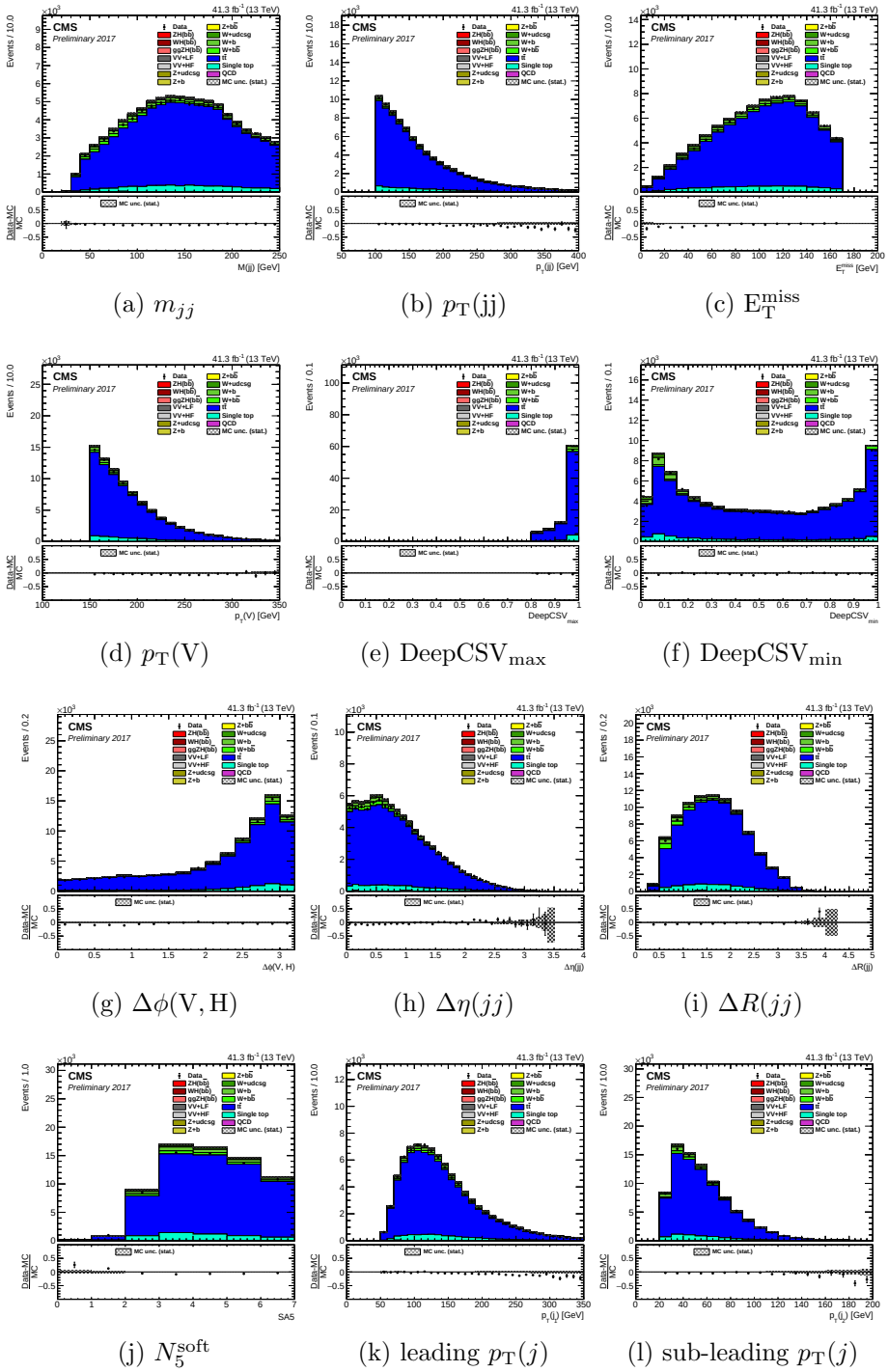
- The $t\bar{t}$ control region is defined by requiring at least two additional jets other than the Higgs boson candidate b jets.

Comparions of data with simulation for the W($\ell\nu$)H analysis observables in these control regions are shown in Figures 9.4–9.6.

Table 9.2: Definition of control regions for the W($\ell\nu$)H channel, common for the electron and muon categories. The values listed for kinematic variables are in units of GeV.

Variable	W + udscg	$t\bar{t}$	W + $b\bar{b}$
$p_T(j_1)$	> 25	> 25	> 25
$p_T(j_2)$	> 25	> 25	> 25
p_T(jj)	> 100	> 100	> 100
p_T(V)	> 150	> 150	> 150
DeepCSV$_{max}$	<Medium	>Tight	>Tight
N_{aj}	–	> 1	< 2
N_{al}	$= 0$	$= 0$	$= 0$
E_T^{miss} significance	> 2	–	> 2
$\Delta\phi(\vec{E}_T^{miss}$, lepton)	< 2	< 2	< 2
m_{jj}	< 250	< 250	< 250, veto $[90 - 150]$

Comparisons of the data with the prediction from simulation are shown in Fig. 9.4 for the $\mathrm{t\bar{t}}$-enriched control region in the W($\ell\nu$)H channel.



(a) m_{jj} (b) $p_{\mathrm{T}}(\mathrm{jj})$ (c) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$

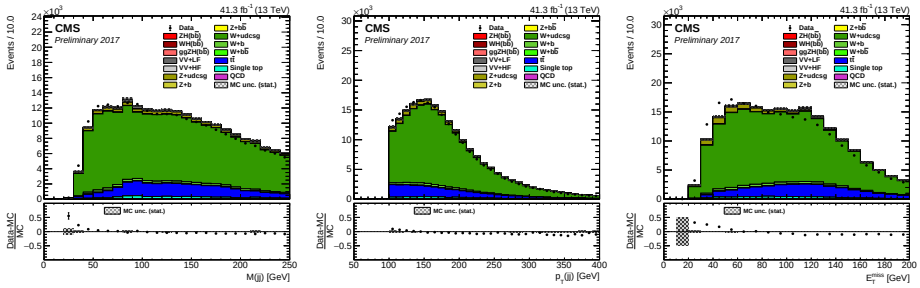
(d) $p_{\mathrm{T}}(\mathrm{V})$ (e) $\mathrm{DeepCSV}_{\mathrm{max}}$ (f) $\mathrm{DeepCSV}_{\mathrm{min}}$

(g) $\Delta\phi(\mathrm{V,H})$ (h) $\Delta\eta(jj)$ (i) $\Delta R(jj)$

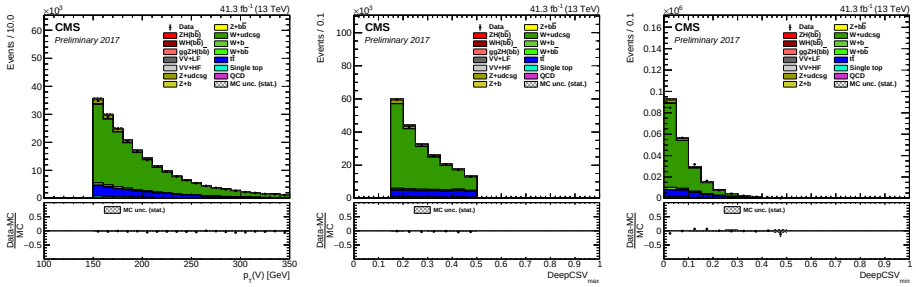
(j) N_5^{soft} (k) leading $p_{\mathrm{T}}(j)$ (l) sub-leading $p_{\mathrm{T}}(j)$

Figure 9.4: Analysis observables for the $\mathrm{t\bar{t}}$-enriched control region in the W($\ell\nu$)H channel.

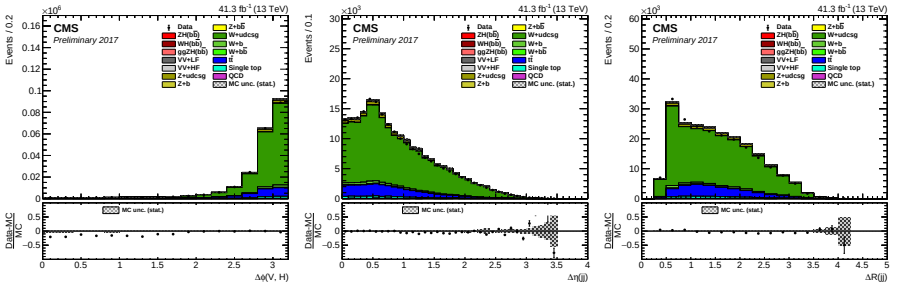
Comparisons of the data with the prediction from simulation are shown in Fig. 9.5 for the W+light-jets control region in the W($\ell\nu$)H channel.



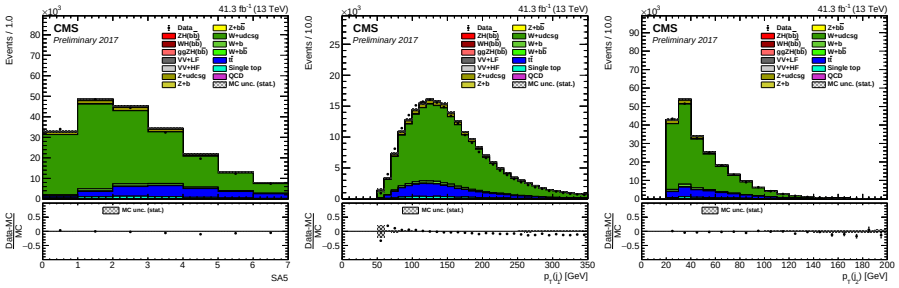
(a) m_{jj} (b) p_{T}(jj) (c) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$



(d) p_{T}(V) (e) $\mathrm{DeepCSV}_{\mathrm{max}}$ (f) $\mathrm{DeepCSV}_{\mathrm{min}}$



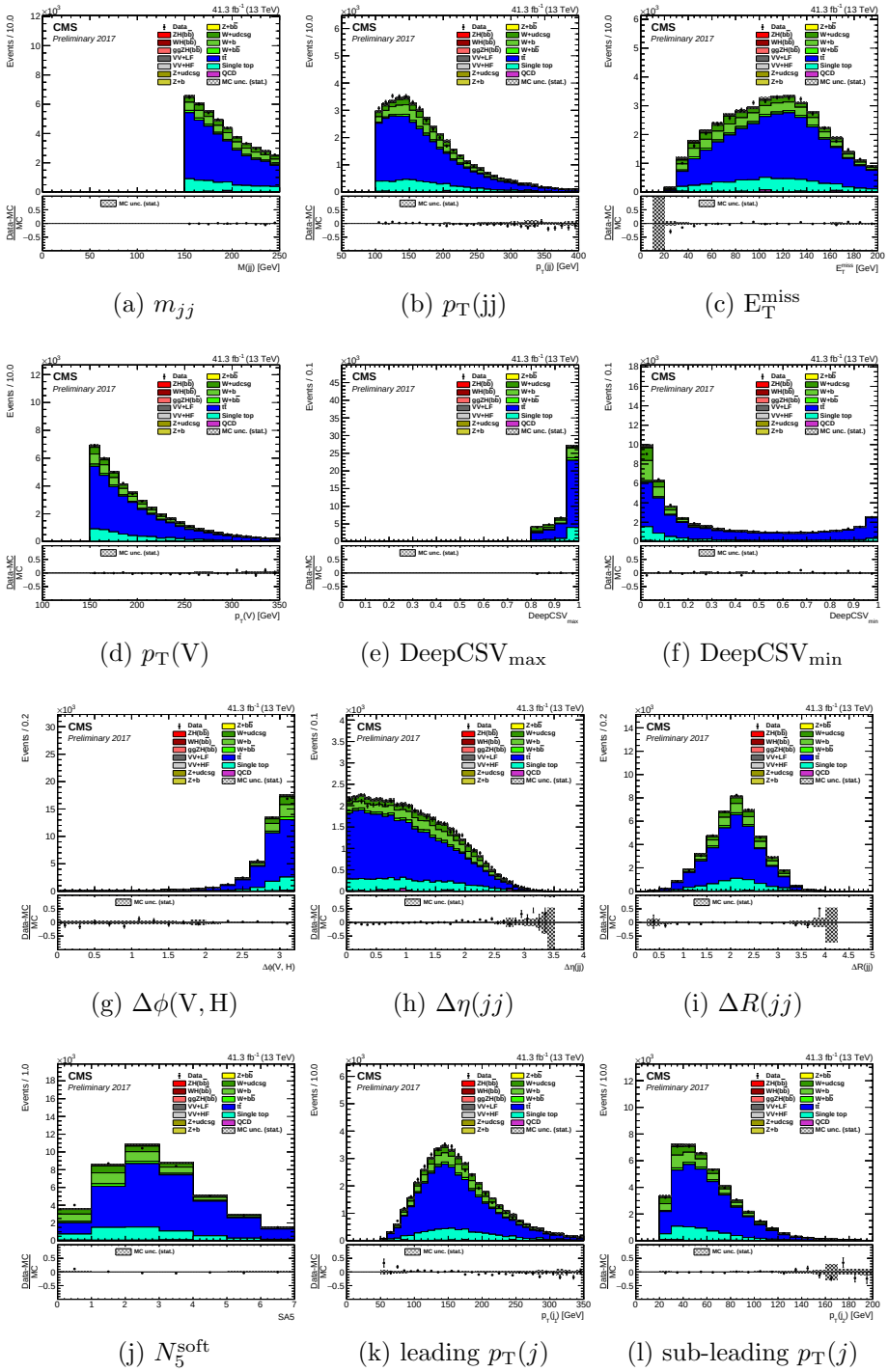
(g) $\Delta\phi$(V, H) (h) $\Delta\eta(jj)$ (i) $\Delta R(jj)$



(j) N_5^{soft} (k) leading $p_{\mathrm{T}}(j)$ (l) sub-leading $p_{\mathrm{T}}(j)$

Figure 9.5: Analysis observables for the W+light-jets control region in the W($\ell\nu$)H channel.

Comparisons of the data with the prediction from simulation are shown in Fig. 9.6 for the $\mathrm{W} + \mathrm{b\bar{b}}$-enriched control region in the $\mathrm{W}(\ell\nu)\mathrm{H}$ channel.



(a) m_{jj} (b) $p_{\mathrm{T}}(\mathrm{jj})$ (c) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$

(d) $p_{\mathrm{T}}(\mathrm{V})$ (e) $\mathrm{DeepCSV}_{\mathrm{max}}$ (f) $\mathrm{DeepCSV}_{\mathrm{min}}$

(g) $\Delta\phi(\mathrm{V,H})$ (h) $\Delta\eta(jj)$ (i) $\Delta R(jj)$

(j) N_5^{soft} (k) leading $p_{\mathrm{T}}(j)$ (l) sub-leading $p_{\mathrm{T}}(j)$

Figure 9.6: Analysis observables for the $\mathrm{W} + \mathrm{b\bar{b}}$-enriched control region in the $\mathrm{W}(\ell\nu)\mathrm{H}$ channel.

9.3 Z($\ell\ell$)H control regions

Three control regions are developed in the Z($\ell\ell$)H channel to target separately the Z + udscg, $t\bar{t}$, and Z + $b\bar{b}$ backgrounds. The selection criteria, common for the electrons and muons, are summarized in Table 9.3.

- The Z + $b\bar{b}$ control region is the most similar to the signal region, with an inverted m_{jj} selection. Note that for the Z($\ell\ell$)H channel it is possible to achieve high Z + $b\bar{b}$ purity in this control region, unlike in the other channels.

- The Z+light-jets control region is defined by inverting the b-tagging requirements on the Higgs boson candidate jets in order to enhance the light flavor (udcsg) jet contribution.

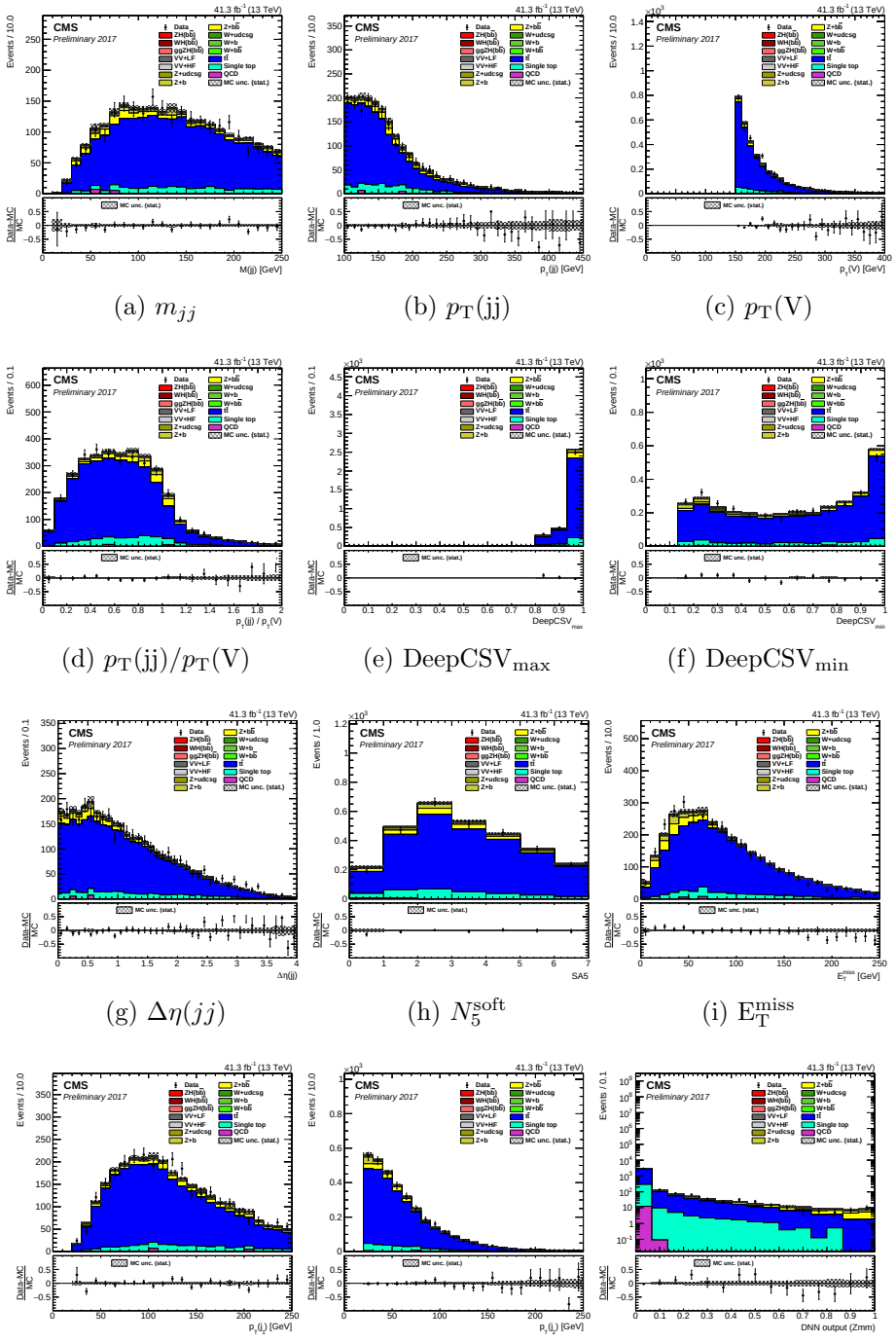
- The $t\bar{t}$ control region is defined by inverting the dilepton invariant mass cut to reject events consistent with M_Z.

Comparisons between data and simulation for the analysis observables in these control regions are shown in Fig. 9.7–9.9

Table 9.3: Definition of control regions for the Z($\ell\ell$)H channel, common for the electron and muon categories. The values listed for kinematic variables are in units of GeV.

Variable	$t\bar{t}$	Z + udscg	Z + $b\bar{b}$
$p_T(j_1)$	> 20	> 20	> 20
$p_T(j_2)$	> 20	> 20	> 20
$p_T(V)$	$[50, 150]$,> 150	$[50, 150]$,> 150	$[50, 150]$,> 150
$\text{DeepCSV}_{\text{max}}$	>Tight	<Loose	>Tight
$\text{DeepCSV}_{\text{min}}$	>Loose	<Loose	>Loose
E_T^{miss}	–	–	< 60
$\Delta\phi(V, H)$	–	> 2.5	> 2.5
$m_{\ell\ell}$	$\notin [0, 10]$, $\notin [75, 120]$	$[75, 105]$	$[85, 97]$
m_{jj}	–	$[90, 150]$	$\notin [90, 150]$

Comparisons of the data with the prediction from simulation are shown in Fig. 9.7 for the $\mathrm{t\bar{t}}$-enriched control region in the $\mathrm{Z}(\ell\ell)\mathrm{H}$ high-p_{T} channel.



(a) m_{jj} (b) $p_{\mathrm{T}}(\mathrm{jj})$ (c) $p_{\mathrm{T}}(\mathrm{V})$

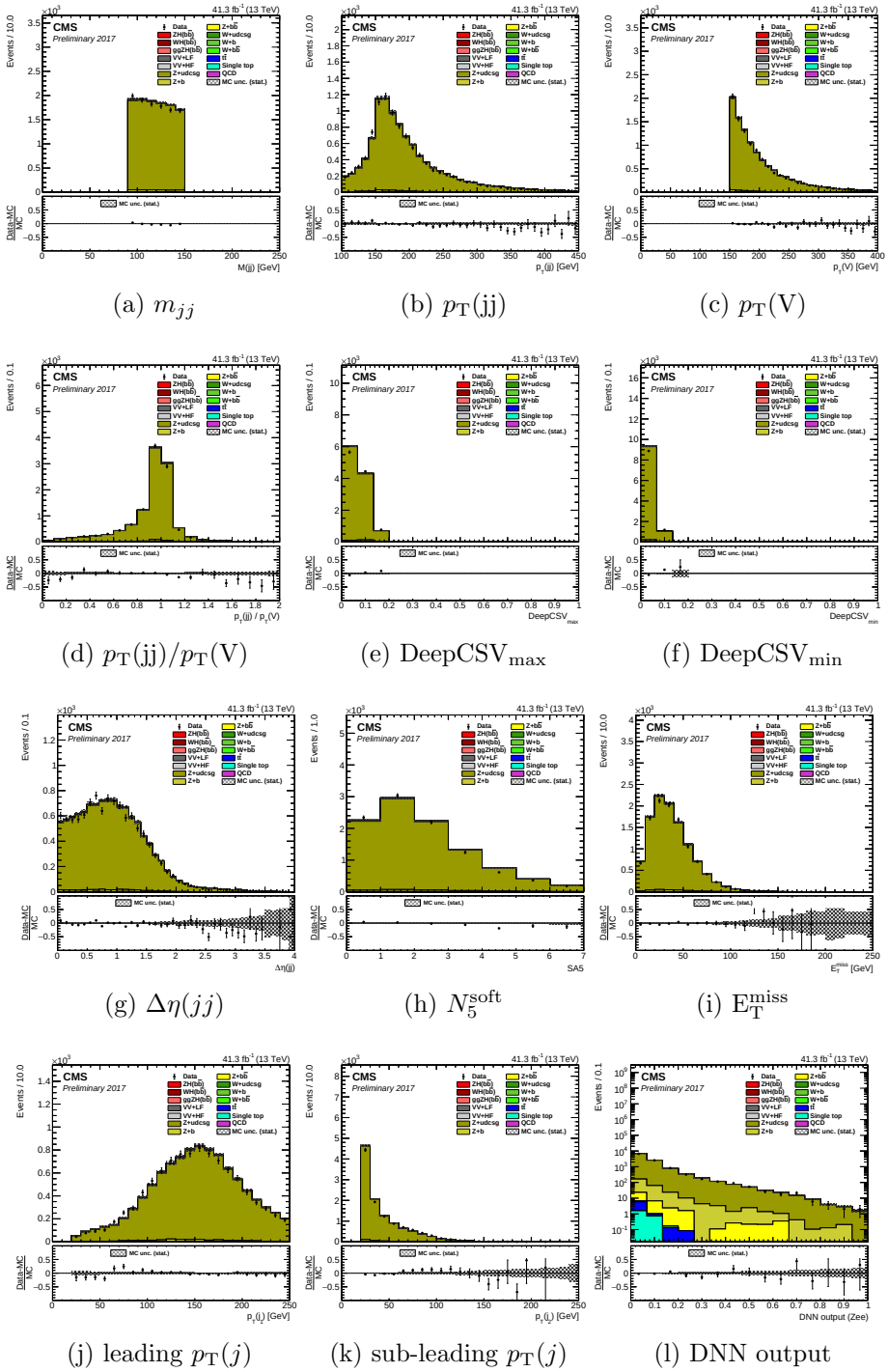
(d) $p_{\mathrm{T}}(\mathrm{jj})/p_{\mathrm{T}}(\mathrm{V})$ (e) $\mathrm{DeepCSV}_{\mathrm{max}}$ (f) $\mathrm{DeepCSV}_{\mathrm{min}}$

(g) $\Delta\eta(jj)$ (h) N_5^{soft} (i) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$

(j) leading $p_{\mathrm{T}}(j)$ (k) sub-leading $p_{\mathrm{T}}(j)$ (l) DNN output

Figure 9.7: Analysis observables in data and simulated samples in the $\mathrm{t\bar{t}}$ control region for the $\mathrm{Z}(\ell\ell)\mathrm{H}$ high-p_{T} channel.

Comparisons of the data with the prediction from simulation are shown in Fig. 9.8 for the Z+light-jets control region in the Z($\ell\ell$)H high-p_{T} channel.



(a) m_{jj} (b) p_{T}(jj) (c) p_{T}(V)

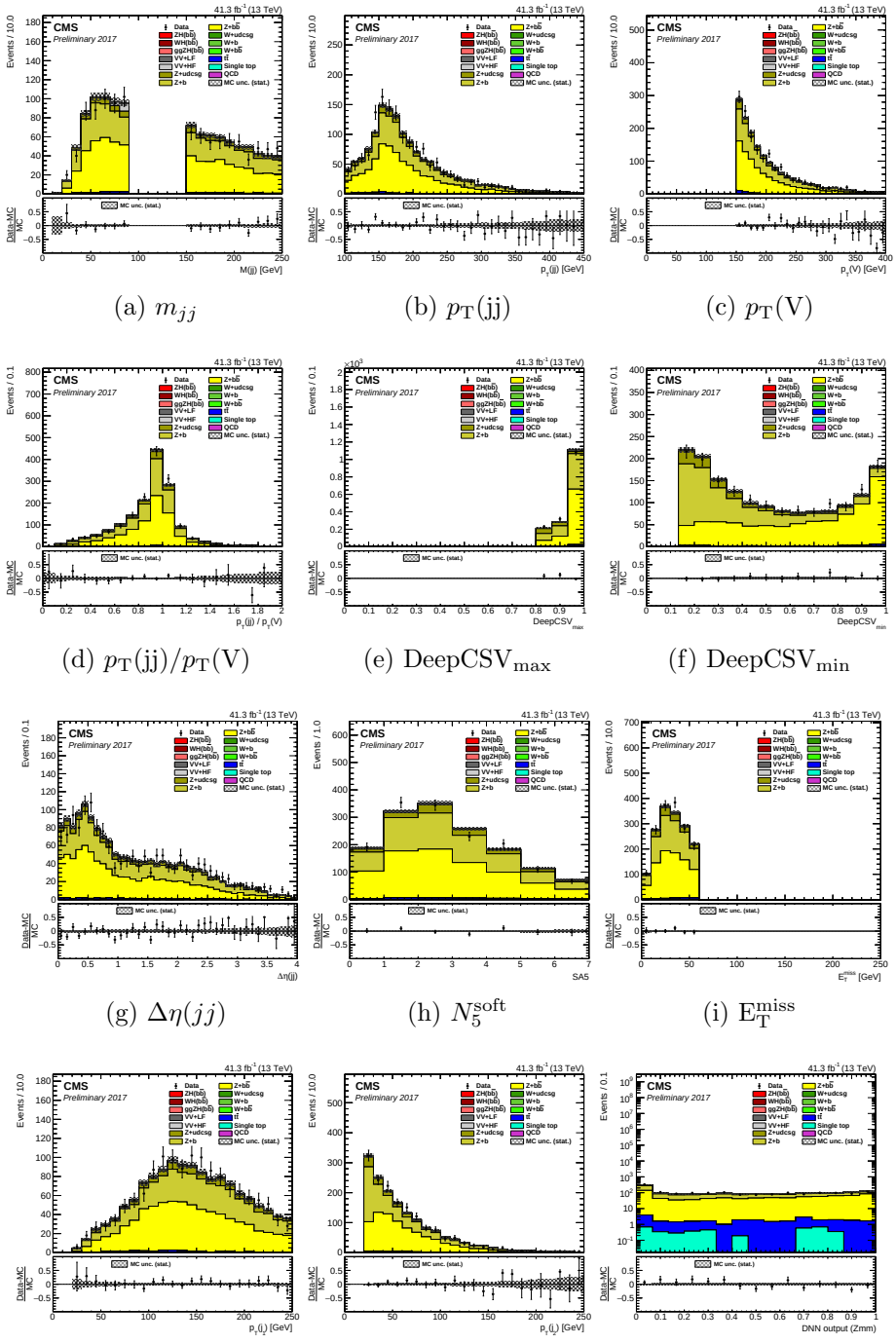
(d) p_{T}(jj)/p_{T}(V) (e) DeepCSV$_{\mathrm{max}}$ (f) DeepCSV$_{\mathrm{min}}$

(g) $\Delta\eta(jj)$ (h) N_5^{soft} (i) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$

(j) leading $p_{\mathrm{T}}(j)$ (k) sub-leading $p_{\mathrm{T}}(j)$ (l) DNN output

Figure 9.8: Analysis observables in data and simulated samples in the Z+light-jets control region for the Z($\ell\ell$)H high-p_{T} channel.

Comparisons of the data with the prediction from simulation are shown in Fig. 9.9 for the $\mathrm{Z} + \mathrm{b\bar{b}}$ control region in the $\mathrm{Z}(\ell\ell)\mathrm{H}$ high-p_{T} channel.



(a) m_{jj} (b) $p_{\mathrm{T}}(\mathrm{jj})$ (c) $p_{\mathrm{T}}(\mathrm{V})$

(d) $p_{\mathrm{T}}(\mathrm{jj})/p_{\mathrm{T}}(\mathrm{V})$ (e) $\mathrm{DeepCSV}_{\mathrm{max}}$ (f) $\mathrm{DeepCSV}_{\mathrm{min}}$

(g) $\Delta\eta(jj)$ (h) N_5^{soft} (i) $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$

(j) leading $p_{\mathrm{T}}(j)$ (k) sub-leading $p_{\mathrm{T}}(j)$ (l) DNN output

Figure 9.9: Analysis observables in data and simulated samples in the $\mathrm{Z} + \mathrm{b\bar{b}}$ control region for the $\mathrm{Z}(\ell\ell)\mathrm{H}$ high-p_{T} channel.

9.4 DNN background multi-classifier

As mentioned in Section 8.1, the production of a vector boson in association with one or more b jets is a particularly difficult background to distinguish from signal. As demonstrated in the previous section, it is possible to define a control region with high purity in $Z + b\bar{b}$ and $Z + b$ in the $Z(\ell\ell)H$ channel. In the $W(\ell\nu)H$ and $Z(\nu\nu)H$ channels, however, it is not possible to obtain a high-purity $V + b\bar{b}$ control region with selections on physical analysis observables alone. A DNN multi-classifier is trained to maximize the separation power between the varied backgrounds in the $V + b\bar{b}$-enriched control regions in the $W(\ell\nu)H$ and $Z(\nu\nu)H$ channels. The resulting improved ability to separate the $V + b\bar{b}$ and $V + b$ components from other backgrounds allows for an improved constraint and therefore a reduced background uncertainty in the signal extraction fits.

The multi-output DNN is trained with the same input variables as for the discrimination between signal and background, as given in Table 8.2. The background multi-classifier returns a set of five probabilities p for an event to belong to the following categories: $V + b\bar{b}$, $V + b$, $t\bar{t}$, single top, and V+light-jets (which is mostly V+light-jets, but also includes small contributions from diboson and QCD multijet events). Each event is assigned to the bin associated with the background process with highest probability, following the binning convention given in Table 9.4.

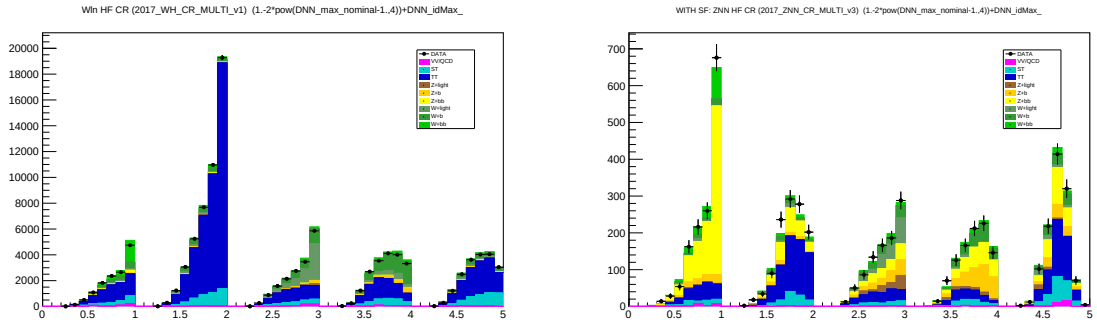
Table 9.4: Binning convention for the DNN background multi-classifier.

category	$V + b\bar{b}$	$t\bar{t}$	$V +$ udscg	$V + b$	single t
bin	0	1	2	3	4

For validation only, the individual bins are further split into regions of varied purity with the formula

$$x = \text{argmax}(p) + (1 - 2(1 - \max(p))^4), \tag{9.1}$$

where the argmax function returns the bin number corresponding to the most probable background type. This distribution is shown with 50 bins in Fig. 9.10.



(a) W($\ell\nu$)+bb CR classifier (50 bins)

(b) Z($\nu\nu$)+bb CR classifier (50 bins)

Figure 9.10: Output of the DNN background multi-classifier in the W($\ell\nu$)H (left) and Z($\nu\nu$)H (right) channels. The fine binning shown is used for validation purposes only.

9.5 Background normalization fits

The predicted normalizations for the $t\bar{t}$, W+jets, and Z+jets backgrounds in the high-p_T(V) phase space relevant for this analysis are known to be unreliable, in particular for the production of W and Z bosons at high p_T(V) in association with one or more heavy flavor quarks. The normalizations for these background components are therefore fit to data as free parameters in the signal extraction (Sec. 10.1.1), in which all the control and signal regions are fit simultaneously. The ratios of the fitted normalizations to those predicted by the simulation are hereby referred to as scale factors (SF).

A binned maximum likelihood fit (Sec. 10.1.1) is performed on data using templates derived from simulation. Systematic uncertainties, described in Sec. 10.2, are included to take into account potential shape differences in the fitted distributions between data and simulation. Separate SF are considered for vector bosons produced with a single b quark (Z + b, W + b) or a $b\bar{b}$ pair (Z+$b\bar{b}$, W + $b\bar{b}$), because the ratio of the predicted normalization to data is expected to be in general different for each of these processes.

In the Z($\nu\nu$)H channel, the W+jets SF are correlated with the SF obtained from the W($\ell\nu$)H channel, which has much better statistics to constrain the W+jets SF. In the W($\ell\nu$)H and Z($\ell\ell$)H channels, the scale factors are correlated between the muon and electron categories. Any difference in lepton efficiencies is taken into account by the systematic uncertainties.

The following control region fit strategy is used in order to ensure a stable fit and to maximize the precision on the fitted background component normalizations:

- **V+light-jets and $t\bar{t}$ CRs:**

  - Fit the yield only for each channel (these CR are generally very pure in the targeted background)

- **V + $b\bar{b}$-enriched CRs:**

  - Z($\ell\ell$)H channel: fit 2-bin DeepCSV distribution

  - W($\ell\nu$)H, Z($\nu\nu$)H channels: fit DNN background multi-classifier (Sec. 9.4), reduced to 5 bins (1 bin per background category)

Table 9.5 summarizes the fitted normalization adjustments from the fit to all CR and SR simultaneously.

Table 9.5: Fitted background normalization adjustments from a simultaneous fit of all control regions and signal regions. The errors include both statistical and systematic uncertainties.

Process	Z($\nu\nu$)H	W($\ell\nu$)H	Z($\ell\ell$)H low-p_{T}	Z($\ell\ell$)H high-p_{T}
W + udscg	1.04 ± 0.07	1.04 ± 0.07	–	–
W + b	2.09 ± 0.16	2.09 ± 0.16	–	–
W + b$\bar{\mathrm{b}}$	1.74 ± 0.21	1.74 ± 0.21	–	–
Z + udscg	0.95 ± 0.09	–	0.89 ± 0.06	0.81 ± 0.05
Z + b	1.02 ± 0.17	–	0.94 ± 0.12	1.17 ± 0.10
Z + b$\bar{\mathrm{b}}$	1.20 ± 0.11	–	0.81 ± 0.07	0.88 ± 0.08
t$\bar{\mathrm{t}}$	0.99 ± 0.07	0.93 ± 0.07	0.89 ± 0.07	0.91 ± 0.07

Chapter 10

Results

10.1 Fit methodology

The signal extraction is performed via a simultaneous binned maximum likelihood fit of the background-enriched control regions (CR), described in Chapter 9, and the signal regions (SR), described in Sec. 8.2. The fitted distributions in the control regions are the same as those described in the background normalization fit in Section 9.5. In the signal regions, the fitted distribution is 15 bins of the reshaped DNN signal classifier (Chapter 8). The Z($\ell\ell$)H channel is split into two p_{T}(V) categories, while the Z($\nu\nu$)H and W($\ell\nu$)H channels consider one single p_{T}(V) category.

10.1.1 Fitted distributions

Figure 10.1 shows the background multi-classifier DNN distributions in the W($\ell\nu$)H (top) and Z($\nu\nu$)H (bottom) V + b$\bar{\mathrm{b}}$-enriched control regions. Figure 10.2 shows the DNN signal classifier output in the signal regions. Note that the remaining fitted distributions are either one bin (yield only) or two bins only (the Z($\ell\ell$)+bb-enriched control region fits two bins of the minimum b-tagging discriminator).

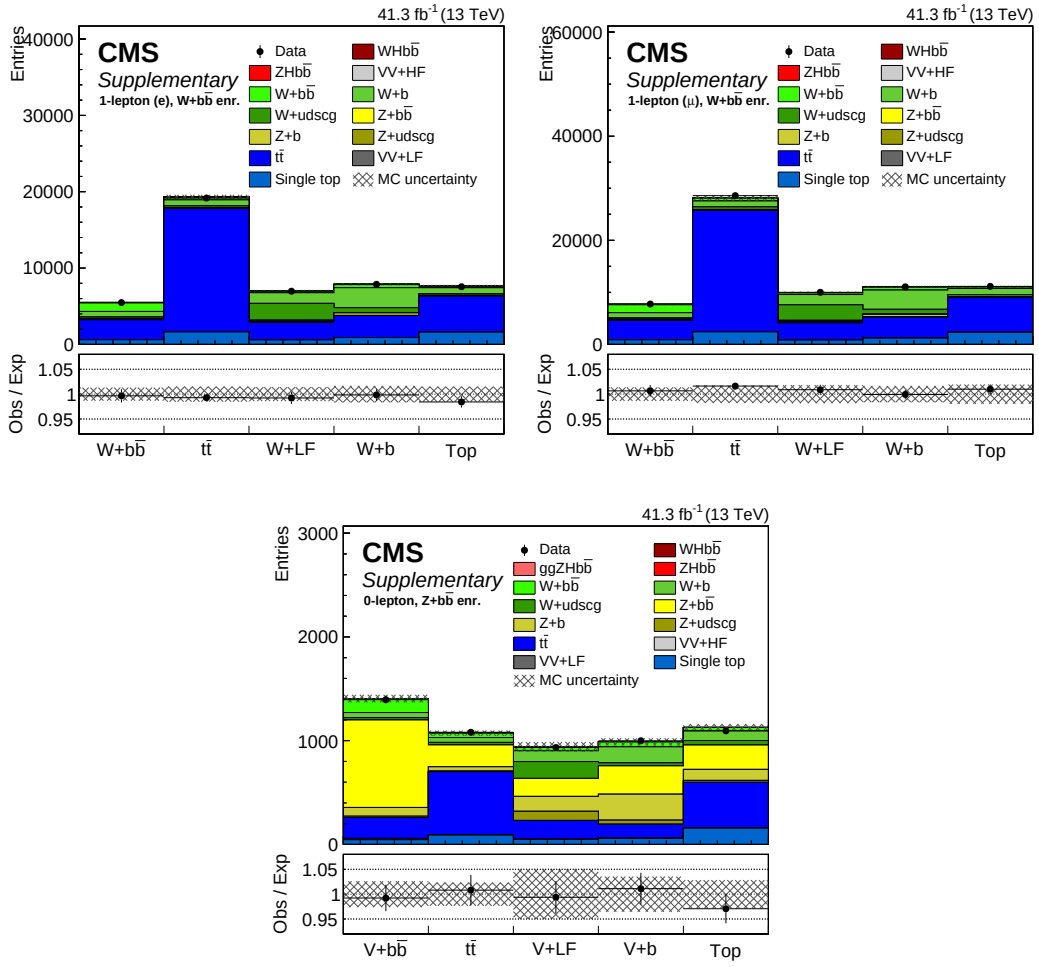


Figure 10.1: DNN background multi-classifier categories (Sec. 9.4) in the $\mathrm{V}+\mathrm{b\bar{b}}$-enriched CR for the $\mathrm{W}(\ell\nu)\mathrm{H}$ channel (top row) for the muon (left) and electron (right) categories, and for the $\mathrm{Z}(\nu\nu)\mathrm{H}$ channel (bottom row).

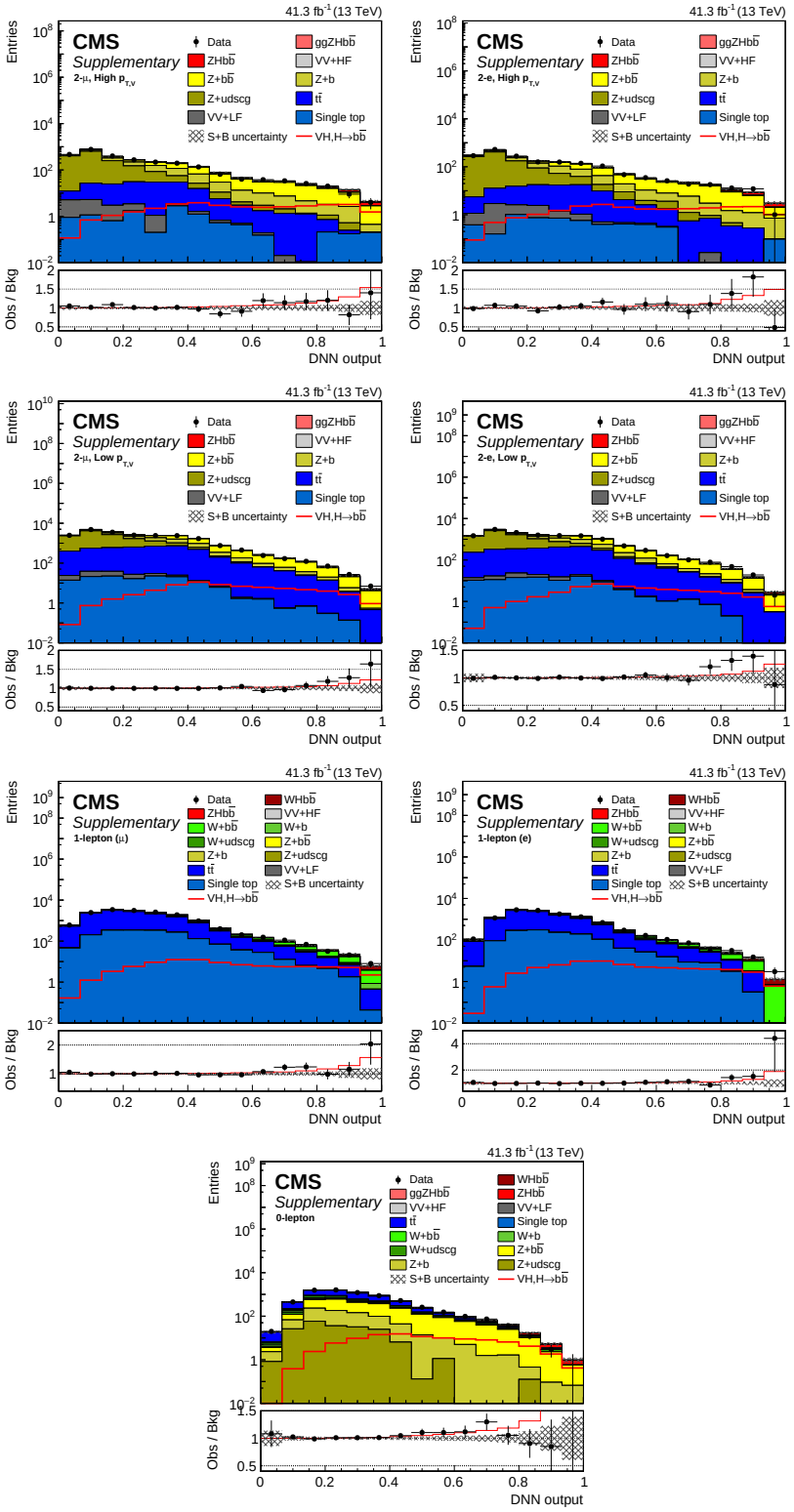


Figure 10.2: DNN signal classifier output in each of the signal regions. First row: Z($\ell\ell$)H muon (left) and electron (right) categories for high p_{T}(V), in the second row the low p_{T}(V) channels are shown. Third row: W($\ell\nu$)H muon (left) and electron (right) categories. Fourth row: Z($\nu\nu$)H channel.

10.2 Systematic uncertainties

Systematic uncertainties can affect both the normalization of a process and the shape of a predicted distribution. The following uncertainties are considered in the signal extraction fit.

- **Background normalization**: a mix of data-driven, experimental, and theory uncertainties contribute to the total uncertainty on the background normalizations. An uncertainty of 15% is assigned for the single top and diboson normalizations (approximately the uncertainty on the measured cross sections). The normalization of the $t\bar{t}$ and V+jets backgrounds are taken directly from data, with the associated uncertainties from the fits, as described in Section 9.5. These uncertainties include the other systematic uncertainties as well the statistics in data.

- **Monte Carlo statistics**: The shape of the DNN is allowed to vary independently per bin within the statistical uncertainty of the simulated samples.

- **b-jet tagging**: The b-tagging discriminator (DeepCSV) output is corrected to match data in a dileptonic $t\bar{t}$ control sample orthogonal to any regions used in this analysis. The uncertainty on this correction is evaluated as a function of the discriminator value for several independent uncertainty sources. The average uncertainty is 6% per b jet, 12% per c jet, and 15% per fake tag (light quarks and gluons).

- **Jet energy scale**: the energy scale for each jet is varied within one standard deviation independently for each of 27 uncertainty sources. The uncertainties are further decorrelated based on the jet p_{T} and η.

- **Jet energy resolution**: The energy resolution for each b jet after applying the b-jet energy regression (Sec. 7.3.1) is calibrated to match data. The b-jet

resolution is varied within an uncertainty of 10%, based on dedicated post-regression b-jet resolution studies that have been performed on a Z+1-jet control sample in data.

- **Signal cross section**: the signal cross section is calculated at next-to-next-to-leading order accuracy, with a total uncertainty of 4% [65], including the effect of QCD scale and PDF variations.

- **$\mathbf{H \to b\bar{b}}$ branching ratio**: an uncertainty of 0.5% is assigned [57].

- **Signal $\boldsymbol{p_T}$ spectrum**: Higher-order effects not taken into account by the simulation can lead to differences in the $p_{\mathrm{T}}(\mathrm{V})$ and $p_{\mathrm{T}}(H)$ spectrum between simulation and data. This can introduce systematic effects in the signal acceptance and efficiency estimates in the high-p_{T} region of this analysis. Two calculations are available that estimate the NLO electroweak [7, 8] and NNLO QCD [66] corrections to VH production in the boosted regime. The estimated effect from NNLO electroweak corrections is 2% for both ZH and WH. An uncertainty of 5% for both ZH and WH is estimated to account for the higher order QCD corrections.

- **$\boldsymbol{\Delta\eta}$(jj) reweighting**: the full LO V+jets $\Delta\eta$(jj) reweighting (Sec. 6.2.2) is taken as a systematic uncertainty on the DNN shape.

- **W boson and $\mathbf{t\bar{t}}$ $\boldsymbol{p_T}$(V) reweighting**: the uncertainties on the $p_{\mathrm{T}}(\mathrm{V})$ corrections (Sec. 7.4.2) are taken from the statistical uncertainties on the fitted slopes, corresponding to 13% for $\mathrm{t\bar{t}}$ and 6% for both $\mathrm{W} + \mathrm{udscg}$ and the combination of $\mathrm{W} + \mathrm{b\bar{b}}$, $\mathrm{W} + \mathrm{b}$, and single top.

- **PDF uncertainties**: the imperfect knowledge of the proton quark content is encoded in a set of NNPDF MC replicas. For each process, the root mean square over all replicas is evaluated for each bin of the DNN distribution and

the largest variation is considered as a normalization uncertainty for the given process.

- **pQCD scale variations**: The perturbative QCD renormalization (μ_R) and factorization (μ_F) scales are varied by 1/2 and 2 times the nominal values separately for each process.

- **Luminosity**: 2.3% uncertainty on the total integrated luminosity recorded by CMS in 2017 [67].

- **Lepton efficiency**: muon and electron trigger, isolation, and identification efficiencies are determined in data using tag-and-probe techniques with Drell-Yan data events, as described in Section 7.4.1. Efficiency corrections are applied to simulation to match the data. The systematic uncertainty on these corrections is evaluated by considering the statistical uncertainties in each efficiency measurement bin as well as efficiency differences obtained by using alternative samples and selections.

- **Unclustered $\mathbf{E_T^{miss}}$** : 3% uncertainty on the calibration of unclustered E_T^{miss} (missing energy associated with particles not clustered into jets).

- **$\mathbf{E_T^{miss}}$+jets trigger**: an uncertainty of 1% is estimated by varying the parameters describing the trigger efficiency curve within statistical uncertainties.

10.3 Statistical interpretation

It is necessary to define a statistical formalism with which to interpret the observed data. In particular, a metric is needed to quantify the level of compatibility between the observed data and

- the expectation including only SM backgrounds

- the expectation including both SM backgrounds and a SM Higgs boson with $m_H = 125$ GeV decaying to $b\bar{b}$

A modified Frequentist approach is used based on the likelihood function, which for a given observed dataset returns a likelihood that a given model would yield the observed data. For a simple counting experiment with N observed data events the likelihood function is given by

$$L(m) = p(N|m) = \frac{m^N}{N!}e^{-m}, \tag{10.1}$$

where m is the mean number of events predicted by the model. Note that the likelihood function L(m) is different than the probability density function (PDF) p(N|m), which is normalized to unity and in this example would be the Poisson distribution. Whereas the PDF is a function of the observed data, the likelihood is a function of the parameters of the assumed model. The likelihood function also has the advantage of invariance under variable transformation, unlike the PDF. Neglecting systematic uncertainties, this analysis can be considered as a set of independent counting experiments in each individual bin of the fitted distributions. In each bin N_i data events are observed whereas b_i events are expected from SM backgrounds and s_i events from a SM Higgs boson decaying to $b\bar{b}$. The likelihood function can then be written as

$$L(\mu) = \prod_{i=1}^{\#\text{ bins}} \frac{(\mu s_i + b_i)^{N_i}}{N_i!} e^{-(\mu s_i + b_i)}, \tag{10.2}$$

where μ is the "signal strength", with $\mu = 0$ corresponding to a model with only SM backgrounds and $\mu = 1$ for the expectation including a SM Higgs boson. As described in Section 10.2, there are also systematic uncertainties which must be considered in the predictions for s_i and b_i. These systematic uncertainties are parametrized by a set of independent nuisance parameters θ_j such that $s_i, b_i \to s_i(\theta), b_i(\theta)$, where the vector of all θ_j is denoted by θ. A prior assumption is made for the value of each

nuisance parameter, $\tilde{\theta_j}$. For example, the signal cross section is assumed to be exactly the theoretically predicted value. An additional term is then added to the likelihood of the form

$$\rho(\theta_j|\tilde{\theta_j},\kappa) = \frac{1}{\sqrt{2\pi}\ln(\kappa)} e^{-\frac{(\ln(\theta_j/\tilde{\theta_j}))^2}{2(\ln(\kappa))^2}}, \tag{10.3}$$

where κ is the assumed uncertainty value and θ_j is a free parameter. The term given in Equation 10.3 is referred to as the log-normal distribution, which for small κ is a Gaussian but has the advantage of properly describing positively defined observables since $\rho(0|\tilde{\theta_j},\kappa) = 0$. The log-normal distribution is assumed for nuisance parameters which affect the normalization of the predictions. As discussed in Section 10.2, systematic uncertainties can also affect the shape of a predicted distribution. For a detailed discussion of the form of $\rho(\theta_j|\tilde{\theta_j},\kappa)$ in this case refer to [68]. The likelihood function including systematic uncertainties can thus be written as

$$L(\mu,\theta) = \prod_{i=1}^{\#\text{ bins}} \frac{(\mu s_i(\theta) + b_i(\theta))^{N_i}}{N_i!} e^{-(\mu s_i(\theta)+b_i(\theta))} \prod_{j=1}^{\#\text{ nuis.}} \rho(\theta_j|\tilde{\theta_j},\kappa). \tag{10.4}$$

For a given set of observed data N_i, the likelihood is maximized with respect to μ and θ to obtain the best-fit signal strength $\hat{\mu}$ and set of nuisance parameter values $\hat{\theta}$. When measuring an established signal, the value of $\hat{\mu}$ is typically the measured quantity of interest, corresponding to the measured signal yield relative to the expectation. In order to establish the presence of a signal, another quantity of interest is the level of compatibility of the observed data with the expectation including only the SM background predictions ($\mu = 0$). A test statistic q_0 is introduced,

$$q_0 = -2\ln\frac{L(\mu=0,\hat{\theta_0})}{L(\hat{\mu},\hat{\theta})}, \tag{10.5}$$

where $\hat{\theta_0}$ is the set of θ_j that maximize the likelihood function with $\mu = 0$ fixed. The probability (p-value) that the observed data are consistent with the background-

only prediction is quantified as the probability that under the assumption $\mu = 0$ the measured q_0 would be greater than or equal to the observed value q_0^{obs}. This p-value is typically converted to number of standard deviations (assuming a Gaussian distribution) via

$$p = \int_Z^\infty \frac{1}{2\pi} e^{-x^2/2} dx. \tag{10.6}$$

An observed excess in the data over the SM background-only prediction can thus be quantified in terms of a number Z of standard deviations (σ). By convention the observation of a new process can be claimed when the p-value for the background-only prediction is less than 2.8 x 10^{-7}, corresponding to at least 5σ.

10.4 Results with 2017 data

10.4.1 VZ, Z $\to$ b$\bar{\text{b}}$ cross-check

As discussed in Section 8.1, VZ, Z $\to$ b$\bar{\text{b}}$ events are nearly indistinguishable from the VH, H $\to$ b$\bar{\text{b}}$ signal other than the distinct m_{jj} peak position. These similarities are exploited by performing a parallel cross-check analysis following the same methodology as the nominal analysis except that the DNN signal classifiers are trained to extract VZ, Z $\to$ b$\bar{\text{b}}$ as signal and the m_{jj} requirements in the signal regions are adjusted to include the full Z boson mass peak. A full description of this cross-check is given in Appendix A.1. The sensitivity of this cross-check measurement benefits from the higher VZ, Z $\to$ b$\bar{\text{b}}$ cross section with respect to the VH, H $\to$ b$\bar{\text{b}}$ signal.

The significance of the observed (expected) VZ, Z $\to$ b$\bar{\text{b}}$ excess is 5.2σ (5.0σ). The results per channel are summarized in Table 10.1. The corresponding signal strength relative to the SM expectation is $\mu_{VZ} = \sigma/\sigma_{\text{SM}} = 1.05^{+0.22}_{-0.21}$. The good agreement

between the result of this cross-check analysis and the SM expectation gives additional confidence in the analysis strategy and the modeling of the backgrounds.

Table 10.1: Expected and observed significances over the SM background, for the combined fit as well as the individual channels, for the VZ, $Z \to b\bar{b}$ cross-check analysis.

Channel	$Z(\nu\nu)Z(b\bar{b})$	$W(\ell\nu)Z(b\bar{b})$	$Z(\ell\ell)Z(b\bar{b})$	comb.
Expected σ	3.5	2.1	3.0	5.0
Observed σ	2.7	2.1	3.4	5.2

10.4.2 VH, $H \to b\bar{b}$

An excess over the SM backgrounds is observed in the data with a significance of 3.3σ, where 3.1σ is expected for a SM Higgs boson with $m_H = 125$ GeV decaying to $b\bar{b}$. Table 10.2 summarizes the expected and observed significances for the combined fit as well as for the individual channels.

Table 10.2: Expected and observed significances over the SM background, for the combined fit as well as the individual channels.

Channel	$Z(\nu\nu)H$	$W(\ell\nu)H$	$Z(\ell\ell)H$	comb.
Expected σ	1.9	1.9	1.9	3.1
Observed σ	1.3	2.6	1.9	3.3

The best-fit signal strength for the excess is $\mu = \sigma/\sigma_{\mathrm{SM}} = 1.08^{+0.35}_{-0.33}$, in good agreement with the expectation for a SM Higgs boson. Two additional fits are performed, one in which the signal strengths for the WH and ZH production modes are decoupled, and one in which the signal strengths in each channel are decoupled. Figure 10.3 summarizes the result of these fits. The signal strength from the combined fit is compatible with the per-channel signal strength fit result with a p-value of 96%.

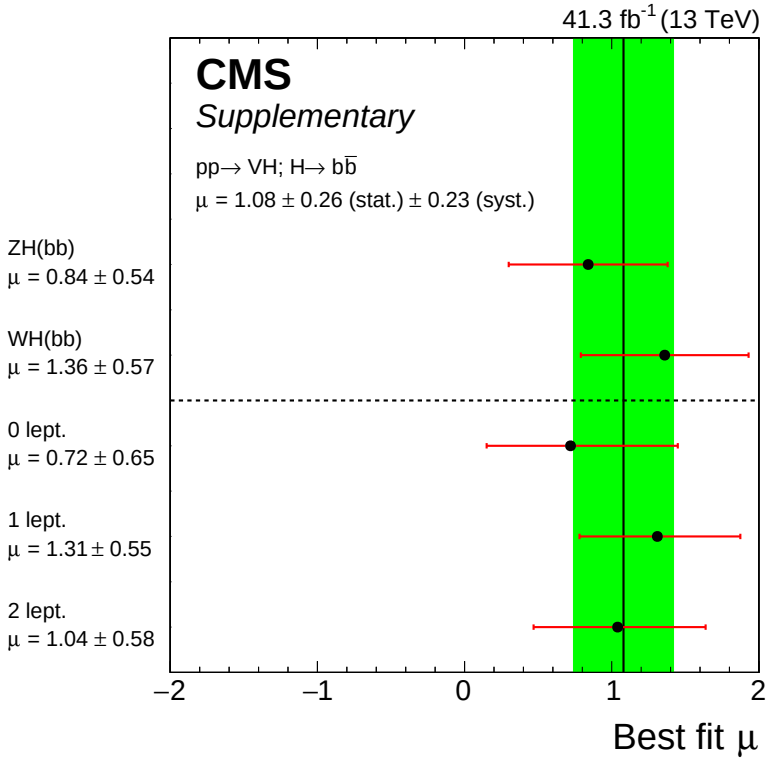


Figure 10.3: Result of additional signal extraction fits with the signal strength decoupled per production mode and per channel. The black vertical line shows the common signal strength fit result with the uncertainty in shaded green. The signal strength from the combined fit is compatible with the per-channel signal strength fit result with a p-value of 96%.

10.5 Combination with previous measurements

10.5.1 Combination with VH, H $\rightarrow$ b$\bar{\text{b}}$ measurements on Run-1 and 2016 datasets

The analysis on 41.3 fb^{-1} of 2017 data is combined with similar measurements performed on 35.9 fb^{-1} of 2016 data [69] as well as Run-1 measurements at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV [61]. An excess over the SM backgrounds is observed (expected) in data with a significance of 4.8σ (4.9σ). The best-fit signal strength from the combined fit is $\mu = \sigma/\sigma_{\text{SM}} = 1.01^{+0.22}_{-0.22}$.

Figure 10.4 combines the multivariate outputs of all channels and analyses, where the events are gathered in bins of similar signal-to-background ratio. A clear excess is visible in the data over the SM backgrounds that is well compatible with the expectation for the SM Higgs boson decaying to b$\bar{\text{b}}$ (bottom inset, red line).

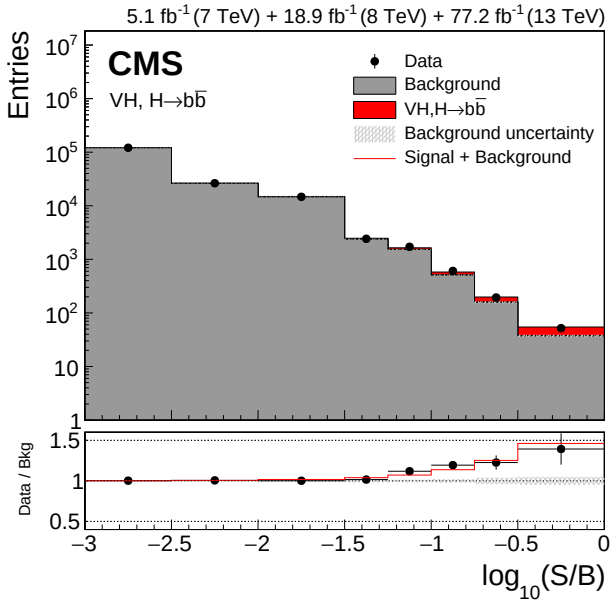


Figure 10.4: Combination of all channels into a single distribution. Events are sorted in bins of similar expected signal-to-background ratio, as given by the value of the output of the corresponding multivariate discriminant. The bottom inset shows the ratio of the data to the predicted sum of backgrounds as well as the expectation including a SM Higgs boson signal with a mass of 125 GeV (red line).

The contribution of each of the main uncertainty sources, described in Sec. 10.2, to the total measurement uncertainty is reported in Table 10.3. The total uncertainty is decomposed into four components: theory, size of simulated samples, experimental, and statistical. Detailed decompositions into specific sources are included for the theory, experimental, and statistical components. Due to correlations in the combined fit between nuisance parameters from different sources, the sum in quadrature for each source does not in general equal the total uncertainty of each component.

Table 10.3: The contributions of the main uncertainty sources to the combined measurement of the Run 1, 2016 and 2017 VH, $H \to b\bar{b}$ analyses. The total uncertainty is decomposed into four components: theory, size of simulated samples, experimental and statistical. Within the theory, experimental and statistical components a more detailed decomposition into specific sources is given.

Uncertainty source	$\Delta\mu$	
Statistical	$+0.18$	-0.17
Normalization of backgrounds	$+0.08$	-0.07
Experimental	$+0.10$	-0.09
b-tagging efficiency	$+0.05$	-0.05
Jet energy scale and resolution	$+0.03$	-0.03
Lepton identification	$+0.01$	-0.01
Luminosity	$+0.03$	-0.02
Other experimental uncertainties	$+0.05$	-0.05
Size of simulated samples	$+0.06$	-0.06
Theory	$+0.09$	-0.08
Signal modeling	$+0.05$	-0.03
Background modeling	$+0.07$	-0.07
Total	$+0.23$	-0.22

10.5.2 Combination of all CMS $H \to b\bar{b}$ searches

Although this thesis has focused on the search for $H \to b\bar{b}$ via the VH production mode, which is the most sensitive channel at the LHC, CMS has performed searches for $H \to b\bar{b}$ in all of the main Higgs boson production modes (Sec. 4.1.2), namely ttH [10], VBF [70], and ggH [59]. These searches on Run-1 and Run-2 data are

combined with the VH, H $\rightarrow$ b$\bar{\text{b}}$ analyses into a global fit for one common signal strength. An excess over the SM backgrounds is observed (expected) in the data with a significance of 5.6σ (5.5σ). This constitutes the observation of Higgs boson decay to bottom quarks by the CMS Collaboration. The best-fit signal strength from the combined fit is $\mu = 1.04 \pm 0.20 = 1.04^{+0.10}_{-0.09}(\text{th.})^{+0.06}_{-0.06}(\text{MC})^{+0.09}_{-0.09}(\text{exp.})^{+0.14}_{-0.14}(\text{stat.})$.

Figure 10.5 summarizes the fit result for the VH, H $\rightarrow$ b$\bar{\text{b}}$ combination (left) and for the full H $\rightarrow$ b$\bar{\text{b}}$ combination (right), where the combined fit result is compared to a fit with independent signal strength modifiers per Higgs boson production mode.

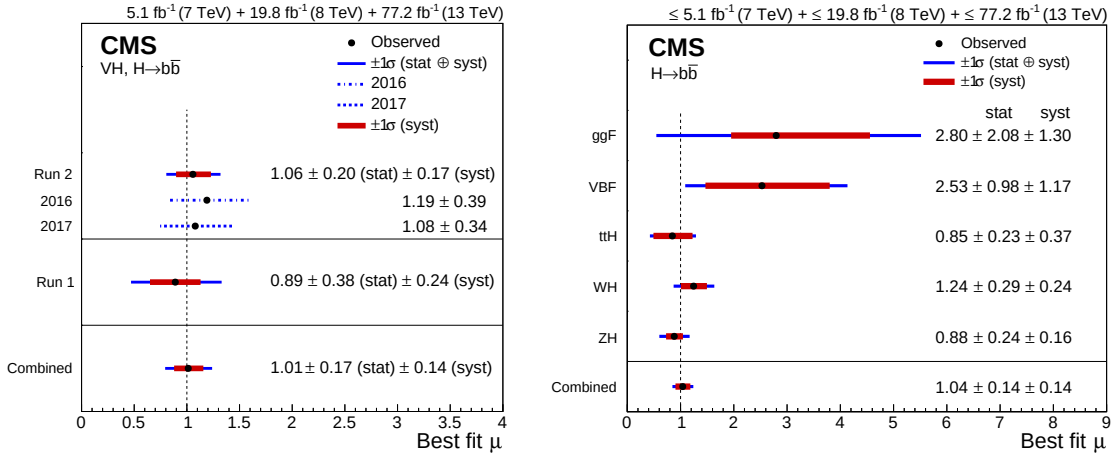


Figure 10.5: Best-fit signal strength per dataset and combined for the VH, H $\rightarrow$ b$\bar{\text{b}}$ combination (left) and a comparison for the full H $\rightarrow$ b$\bar{\text{b}}$ combination of the combined fit result with a fit with individual signal strengths per Higgs boson production mode (right).

10.6 Invariant mass analysis

An alternative approach to extracting the signal via fits to the DNN output is to fit directly the m_{jj} distribution, which is the most discriminating physical observable between signal and background. This approach is not as sensitive to the H $\rightarrow$ b$\bar{\text{b}}$ signal because the DNN more optimally identifies the highest-S/B regions, but allows for the visualization of the excess over the SM backgrounds with a physical observ-

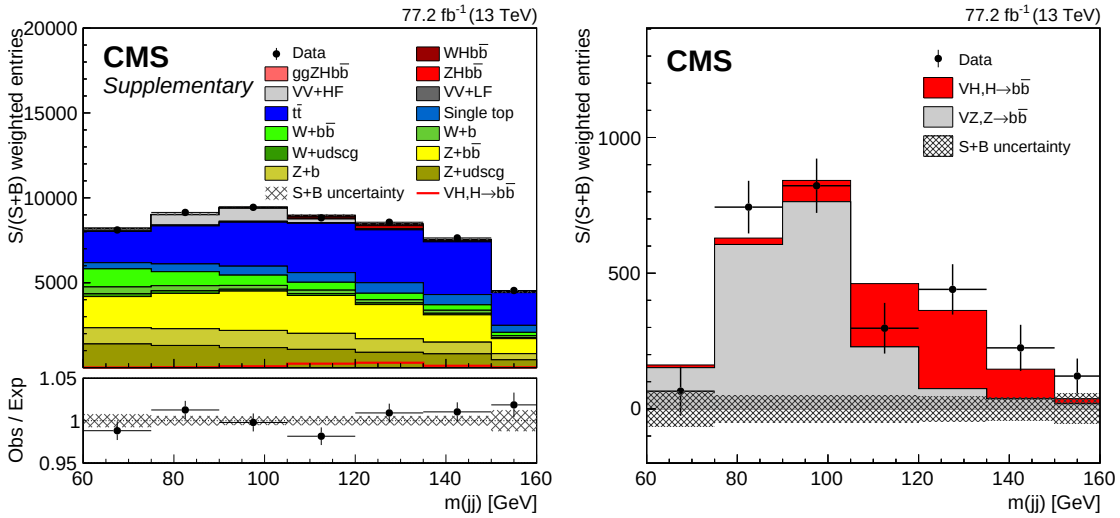


Figure 10.6: Dijet invariant mass distribution for events weighted by S/(S+B) in all channels combined in the 2016 and 2017 data sets. Weights are derived from a fit to the m(jj) distribution, as described in the text. Shown are data (points) and the fitted VH signal (red) and VZ background (grey) distributions, as well as all other backgrounds. The right plot shows the same distribution with nonresonant backgrounds subtracted.

able. Events in the signal regions are further categorized based on the output of the DNN signal classifier, with correlations between the DNN score and m_{jj} removed by fixing the m_{jj}-correlated DNN input variables to mean background values. A full description of the m_{jj} cross-check analysis is given in Appendix A.2. Figure 10.6 (left) shows the combined m_{jj} plot using both 2016 and 2017 data (77.2 fb^{-1}), where the m_{jj} distribution in each fitted signal region is included with a per-category weight of S/S+B. Figure 10.6 (right) shows the same distribution with the nonresonant background subtracted. A clear excess over the SM backgrounds is visible in the data at m_{jj}= 125 GeV, further validating the observed H $\rightarrow$ b$\bar{\mathrm{b}}$ signal.

Chapter 11

Conclusion

A measurement is presented of the Higgs boson, produced in association with a W or Z boson, decaying to a bottom quark-antiquark pair. Despite the very large branching fraction (58%) for Higgs boson decay to $b\bar{b}$, this measurement is extremely experimentally challenging at the LHC due to a variety of much larger background processes yielding similar final state signatures to the expected signal. The data sample corresponds to 41.3 fb^{-1} of $\sqrt{s} = 13$ TeV proton-proton collision data recorded by the CMS experiment in 2017. An excess in data is observed over the SM backgrounds with a significance of 3.3σ, where 3.1σ is expected for a SM Higgs boson with $m_H = 125$ GeV. The measured signal strength is $\mu = 1.08^{+0.35}_{-0.33}$, in good agreement with the SM expectation. The result combined with similar previous searches by CMS yields an observed (expected) excess of 5.6σ (5.5σ) with a signal strength of $\mu = 1.04 \pm 0.20$. This constitutes the observation of Higgs boson decay to bottom quarks [14]. A similar result was jointly submitted for publication by the ATLAS experiment [71]. This observation has been achieved earlier than originally expected by exploiting multiple sophisticated analysis techniques including several applications of the latest machine learning developments.

The Yukawa coupling to bottom quarks has thus been firmly and directly established, with a value consistent with the SM expectation within current 20% experimental precision. With the observation of ttH production achieved last year [10, 11] and the earlier observation of Higgs boson decay to tau leptons [9], the measurement of all the third generation Yukawa couplings has thus been achieved. A large variety of remarkable precision measurements of the Higgs boson at the LHC since discovery have all been in good agreement with the SM expectation. This impressive range of high precision measurements achieved, including the measurement of m_H to nearly per-mille precision, is a testament to the remarkable achievement of the LHC and the experiments.

And yet it is well established that the SM is not a full description of Nature. It may very well be that there are BSM particles just beyond the current experimental reach, whether via direct searches or indirect SM precision measurements. As described in Section 2.6, measurements of the Higgs boson properties are an excellent avenue to probe physics beyond the SM. An additional 59.7 fb^{-1} of $\sqrt{\text{s}} = 13$ TeV data recorded by CMS in 2018 is already available, allowing for improved measurements of the Higgs boson properties. The first precision differential measurements of the VH production mode are now possible using the newly accessible $\text{H} \to \text{b}\bar{\text{b}}$ decay. BSM particles with masses above the LHC direct experimental reach of roughly several TeV can instead be probed by considering potential deviations in the Higgs boson properties due to Higgs boson couplings to high-mass BSM particles. Such effects tend to increase as the square of the ratio of the SM particle momentum scale to the BSM mass scale. The high-p_{T} kinematic regime considered in the VH, $\text{H} \to \text{b}\bar{\text{b}}$ analysis is therefore an especially interesting probe of BSM physics. Now that the confirmation of the Yukawa coupling to bottom quarks has been achieved, the focus for the VH, $\text{H} \to \text{b}\bar{\text{b}}$ analysis will shift to such constraints on BSM couplings.

With Run-3 of the LHC scheduled to deliver roughly 150 fb^{-1} of $\sqrt{s} = 14$ TeV from 2021 to 2023, a total dataset of more than 300 fb^{-1} will have been delivered by the LHC. With this enormous dataset the first measurement of second generation Yukawa couplings may be possible by CMS via the measurement of the extremely rare decay of the Higgs boson to two muons. By the end of the High-Luminosity LHC, scheduled to begin in 2026 and run for roughly ten years, a factor twenty times more data will be delivered with respect to the currently recorded datasets. With this dataset, it may be possible to measure the shape of the Higgs potential, a critical (and unmeasured) aspect of the SM, by measuring HH production. This rich dataset will also make possible high-precision measurements of Higgs bosons produced in the most extreme kinematic topologies.

The Higgs boson has very rapidly transformed from an unobserved particle to a standard candle of SM precision measurements. We must not forget, however, that we are at just the beginning an era of probing the Higgs boson to make stringent experimental tests of the SM. With the observation of Higgs boson decay to bottom quarks presented in this thesis, the Higgs boson couplings with the third generation fermions are confirmed to be fully SM-like within the current 15-20% experimental precision. Such achievements are nonetheless only one step towards an extended program of high-precision experimental tests of the SM description of the Higgs sector. Only through the dedicated pursuit of physics beyond the Standard Model via unexplored avenues can we hope to discover the fundamental physics laws Nature has so far kept hidden.

Appendix A

Appendix

A.1 VZ, Z $\to$ b$\bar{\text{b}}$ cross-check analysis

As described in Section 10.4.1, a cross-check analysis is performed to extract a VZ, Z $\to$ b$\bar{\text{b}}$ signal using all the objects, corrections and analysis techniques as in the VH, H $\to$ b$\bar{\text{b}}$ measurement. The CR and SR definitions are very similar to those listed in Tables 9.1, 9.2, 9.3 (CR), and 8.1 (SR). The only difference in selections is in the SR invariant mass boundaries for the W($\ell\nu$)H and Z($\ell\ell$)H channels, which are lowered to 60 GeV (from 90 GeV), The corresponding CR mass selections are also shifted to preserve orthogonality with the SR. The DNN signal classifier, as described in Sec. 8.3, is trained to extract the VZ, Z $\to$ b$\bar{\text{b}}$ signal using the same set of training inputs, with VH, H $\to$ b$\bar{\text{b}}$ included as a background.

Table A.1 shows the background normalization scale factors (Sec. 9.5) obtained from the VZ cross-check analysis simultaneous fit of all CR and SR. The fitted scale factors are in good agreement with those obtained from the nominal analysis. The unblinded distributions resulting from the combined fit of all channels are shown in Figure A.1 for the VZ DNN score in the signal regions.

Table A.1: Background normalization scale factors from the VZ, $Z \to b\bar{b}$ cross-check analysis SR+CR fit. The errors include both statistical and systematic uncertainties. Compatible values are obtained from the nominal VH, $H \to b\bar{b}$ fit.

Process	$Z(\nu\nu)H$	$W(\ell\nu)H$	$Z(\ell\ell)H$ low-p_T	$Z(\ell\ell)H$ high-p_T
W + udscg	1.04 ± 0.01	1.04 ± 0.01	–	–
W + b	2.02 ± 0.09	2.02 ± 0.09	–	–
W + $b\bar{b}$	2.02 ± 0.13	2.02 ± 0.13	–	–
Z + udscg	0.86 ± 0.05	–	0.88 ± 0.01	0.80 ± 0.01
Z + b	1.07 ± 0.14	–	0.89 ± 0.05	1.13 ± 0.08
Z + $b\bar{b}$	1.20 ± 0.07	–	0.84 ± 0.03	0.95 ± 0.05
$t\bar{t}$	0.97 ± 0.02	0.93 ± 0.01	0.88 ± 0.01	0.90 ± 0.02

The best-fit VZ signal strength is found to be $\mu_{VZ} = \sigma/\sigma_{\mathrm{SM}} = 1.05^{+0.22}_{-0.21}$. The observed (expected) significance of the excess over the background-only prediction is 5.2σ (5.0σ).

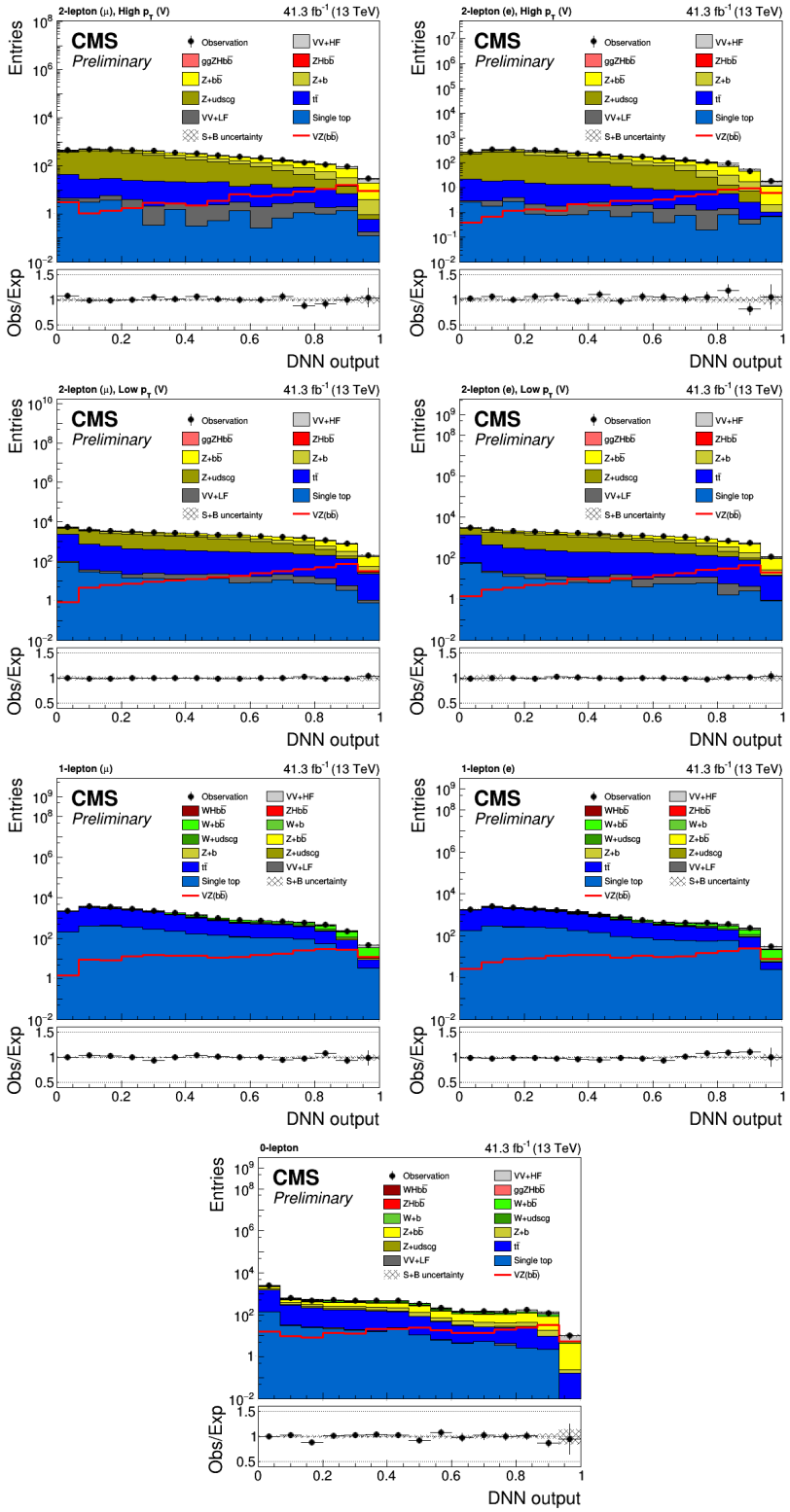


Figure A.1: VZ DNN score in all signal regions for the VZ, Z $\rightarrow$ b$\bar{\text{b}}$ cross-check analysis, First row: Z($\ell\ell$)H muon (left) and electron (right) categories for high p_{T}(V), in the second row the low p_{T}(V) channels are shown. Third row: W($\ell\nu$)H muon (left) and electron (right) categories. Fourth row: Z($\nu\nu$)H channel.

A.2 Invariant mass cross-check analysis

As mentioned in Section 10.6, an alternative approach to the DNN signal extraction is to fit directly the invariant mass distribution. One benefit of this approach is the direct physical interpretation of the fit, showing the $Z \to b\bar{b}$ and $H \to b\bar{b}$ mass peaks, and further validating the results observed with the DNN fit.

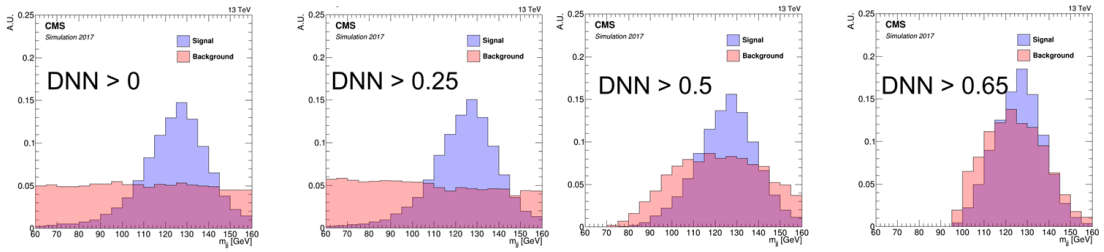
In this cross-check analysis the signal regions are defined according to Sec. 8.2, except that the invariant mass selections have been widened to [60-160] GeV for the W($\ell\nu$)H and Z($\ell\ell$)H channels. In order to increase the signal over background ratio, events are categorized according to the output of a DNN signal classifier. The DNN, as described in Sec. 8.3, is trained with the invariant mass as an input variable. Therefore, the DNN score is highly correlated with the invariant mass and a categorization of events with the DNN score would yield biased distributions of the invariant mass. This bias should be avoided for the invariant mass signal extraction fits, such that the resonant signal can be observed over the nonresonant backgrounds.

The correlations between the invariant mass and the other DNN inputs (listed in Table. 8.2) are investigated. The variables significantly correlated with the mass are fixed to the central values of the background distributions, as listed in Table A.2. The resulting DNN evaluated with the mass-correlated inputs fixed is hereby referred to as the "massless evaluated DNN" (MEDNN).

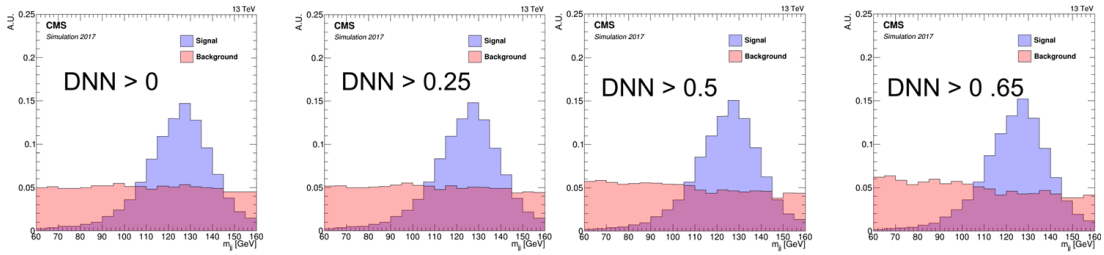
Table A.2: DNN output variables correlated with the invariant mass, separated by channel. When the variable is found to be correlated with the invariant mass, the mean value of the background distribution is used in the MEDNN evaluation. All values listed are in units of GeV.

Channel	Z($\nu\nu$)H		W($\ell\nu$)H (μ)		W($\ell\nu$)H (e)		Z($\ell\ell$)H (high p_T(V))		Z($\ell\ell$)H (low p_T(V))	
	Correlated	Mean	Correlated	Mean	Correlated	Mean	Correlated	Mean	Correlated	Mean
m_{jj}	✓	110.8	✓	120.1	✓	121.0	✓	118.4	✓	117.5
$\sigma(m_{jj})$							✓	21.3	✓	7.0
$\Delta\eta$(jj)	✓	0.62	✓	0.76	✓	0.75	✓	0.93	✓	1.36
Leading jet p_T	✓	154.3	✓	140.5	✓	142.4	✓	172.9	✓	85.7
Subleading jet p_T	✓	68.3	✓	59.0	✓	59.5	✓	55.3	✓	41.3
p_T(jj)	✓	206.3								
p_T(V)	✓	220.6								

The distributions of the invariant mass for signal (blue) and background (red) for the nominal DNN (top) described in Sec. 8.3 and the massless evaluated DNN (bottom) are shown in Fig. A.2. Each signal region is split into four MEDNN categories, with boundaries listed in Table A.3. The invariant mass distribution for the backgrounds is much less biased in the MEDNN categories.



(a) Standard DNN



(b) Massless evaluated DNN (MEDNN)

Figure A.2: Signal (blue) and background (red) invariant mass distributions for the nominal DNN (top) described in Sec. 8.3 and the massless evaluated DNN (bottom).

Table A.3: The optimized MEDNN category boundaries for each channel.

Channel	Boundaries
0-lepton	[0.0 ; 0.558 ; 0.768 ; 0.838 ; 1.0]
1-lepton	[0.0 ; 0.550 ; 0.856 ; 0.941 ; 1.0]
2-lepton high p_{T}(V)	[0.0 ; 0.486 ; 0.836 ; 0.899 ; 1.0]
2-lepton low p_{T}(V)	[0.0 ; 0.662 ; 0.875 ; 0.927 ; 1.0]

The same procedure is applied to the 2016 data with the same training inputs as for 2017 data apart from the b-tagging discriminator, where an older algorithm is used instead of DeepCSV. A combined 2016+2017 fit of all signal regions and control regions is performed to extract a common VH, H $\to$ b$\bar{\mathrm{b}}$ signal strength. The background normalization scale factors obtained from the combined fit are shown in Table A.4 for 2016 data and Table A.5 for 2017 data.

Table A.4: Background normalization scale factors for the 2016 MEDNN analysis from the SR+CR fit. The errors include both statistical and systematic uncertainties.

Process	Z($\nu\nu$)H	W($\ell\nu$)H	Z($\ell\ell$)H low-p_{T}	Z($\ell\ell$)H high-p_{T}
W + udscg	1.09 ± 0.08	1.09 ± 0.08	–	–
W + b	1.72 ± 0.13	1.72 ± 0.13	–	–
W + b$\bar{\mathrm{b}}$	1.31 ± 0.15	1.31 ± 0.15	–	–
Z + udscg	1.30 ± 0.11	–	0.92 ± 0.06	0.95 ± 0.06
Z + b	2.25 ± 0.32	–	0.82 ± 0.08	0.74 ± 0.11
Z + b$\bar{\mathrm{b}}$	1.74 ± 0.14	–	0.94 ± 0.08	1.09 ± 0.10
t$\bar{\mathrm{t}}$	0.89 ± 0.07	0.87 ± 0.07	0.85 ± 0.07	0.85 ± 0.07

The results are summarized in Table A.6. The corresponding fitted signal strength is $\mu = \sigma/\sigma_{\mathrm{SM}} = 0.91^{+0.35}_{-0.34}$, in agreement with the result obtained from the nominal fits to multivariate discriminants.

The fitted invariant mass distributions are merged into a single distribution with each category assigned a weight of S/(S+B). The weighted combined mass distribution including 2016 and 2017 data is shown in Figure 10.6 without (left) and with (right)

Table A.5: Background normalization scale factors for the 2017 MEDNN analysis from the SR+CR fit. The errors include both statistical and systematic uncertainties.

Process	Z($\nu\nu$)H	W($\ell\nu$)H	Z($\ell\ell$)H low-p_{T}	Z($\ell\ell$)H high-p_{T}
W + udscg	1.02 ± 0.07	1.02 ± 0.07	–	–
W + b	1.82 ± 0.14	1.82 ± 0.14	–	–
W + b$\bar{\mathrm{b}}$	2.05 ± 0.22	2.05 ± 0.22	–	–
Z + udscg	0.95 ± 0.08	–	0.88 ± 0.06	0.81 ± 0.05
Z + b	1.16 ± 0.16	–	0.99 ± 0.13	1.12 ± 0.11
Z + b$\bar{\mathrm{b}}$	1.00 ± 0.08	–	0.72 ± 0.06	0.82 ± 0.07
t$\bar{\mathrm{t}}$	0.97 ± 0.08	0.90 ± 0.07	0.89 ± 0.07	0.88 ± 0.07

Table A.6: Expected and observed significances over the SM background for the invariant mass cross-check analysis, for the combined fit as well as the individual channels.

Channel	0-lepton	1-lepton	2-lepton	comb.
2016 analysis				
Expected	1.0	1.0	1.5	2.1
Observed	0.0	2.4	1.3	2.5
2017 analysis				
Expected	1.4	1.0	1.5	2.2
Observed	1.5	0.0	0.7	1.3
2016 + 2017 analysis				
Expected	–	–	–	3.0
Observed	–	–	–	2.7

background subtraction. The background-subtracted plot is shown for 2016 and 2017 separately in Fig. A.3.

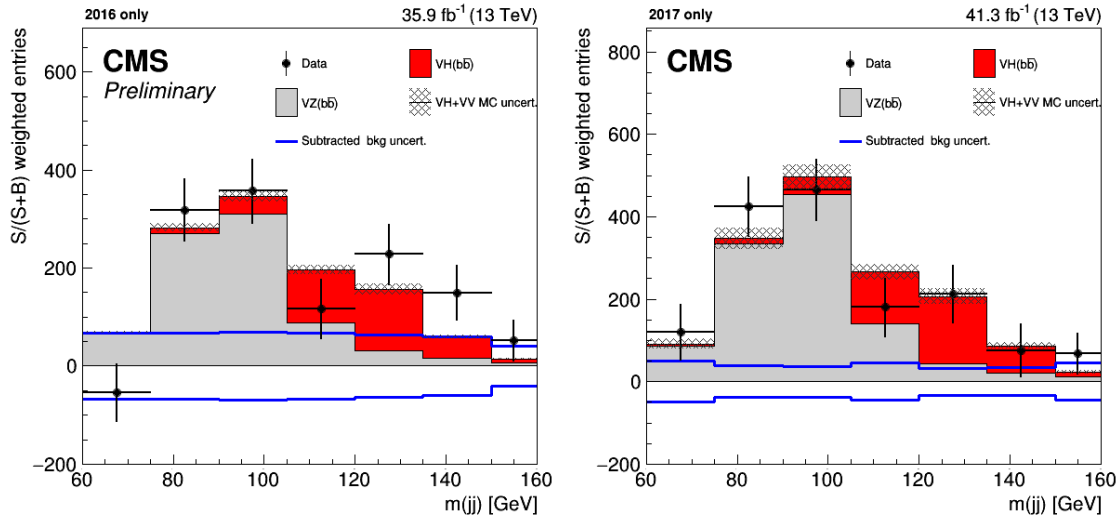


Figure A.3: Combined S/(S+B)-weighted invariant mass distribution with nonresonant backgrounds subtracted for 2016 data (left) and 2017 data (right).

Bibliography

[1] ALEPH, DELPHI, L3, and OPAL Collaborations, "Search for the Standard Model Higgs boson at LEP," *Physics Letters B*, vol. 565, pp. 61 – 75, 2003.

[2] ALEPH, DELPHI, L3, OPAL, and SLD Collaborations, "Precision electroweak measurements on the Z resonance," *Physics Reports*, vol. 427, no. 5, pp. 257 – 454, 2006.

[3] CDF and D0 Collaborations, "Evidence for a Particle Produced in Association with Weak Bosons and Decaying to a Bottom-Antibottom Quark Pair in Higgs Boson Searches at the Tevatron," *Phys. Rev. Lett.*, vol. 109, p. 071804, 2012.

[4] E. Mobs, "The CERN accelerator complex," 2016. General Photo.

[5] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *JINST*, vol. 9, no. 10, p. P10009., 2014.

[6] CMS Collaboration, "CMS Technical Design Report for the Pixel Detector Upgrade," Tech. Rep. CERN-LHCC-2012-016. CMS-TDR-11, 2012.

[7] M. Ciccolini, A. Denner, and S. Dittmaier, "Strong and electroweak corrections to the production of Higgs+2jets via weak interactions at the LHC," *Phys. Rev. Lett.*, vol. 99, p. 161803, 2007.

[8] M. Ciccolini, A. Denner, and S. Dittmaier, "Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC," *Phys. Rev.*, vol. D77, p. 013002, 2008.

[9] ATLAS Collaboration, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ =7 and 8 TeV," *Journal of High Energy Physics*, vol. 2016, no. 8, p. 45, 2016.

[10] CMS Collaboration, "Observation of $t\bar{t}h$ production," *Phys. Rev. Lett.*, vol. 120, p. 231801, 2018.

[11] ATLAS Collaboration, "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector," *Physics Letters B*, vol. 784, pp. 173 – 191, 2018.

[12] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 2.* Technical Design Report ATLAS, Geneva: CERN, 1999.

[13] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, "Jet substructure as a new Higgs search channel at the LHC," *Phys. Rev. Lett.*, vol. 100, p. 242001, 2008.

[14] CMS Collaboration, "Observation of Higgs Boson Decay to Bottom Quarks," *Phys. Rev. Lett.*, vol. 121, p. 121801, 2018.

[15] R. P. Feynman, *The Strange Theory of Light and Matter.* Princeton University Press, 1988.

[16] P. Higgs, "Broken symmetries, massless particles and gauge fields," *Physics Letters*, vol. 12, no. 2, pp. 132 – 133, 1964.

[17] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964.

[18] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles," *Phys. Rev. Lett.*, vol. 13, pp. 585–587, Nov 1964.

[19] P. W. Higgs, "Spontaneous symmetry breakdown without massless bosons," *Phys. Rev.*, vol. 145, pp. 1156–1163, May 1966.

[20] T. W. B. Kibble, "Symmetry Breaking in Non-Abelian Gauge Theories," *Phys. Rev.*, vol. 155, pp. 1554–1561, Mar 1967.

[21] ALEPH Collaboration, "Observation of an excess in the search for the Standard Model Higgs boson at ALEPH," *Physics Letters B*, vol. 495, no. 1, pp. 1 – 17, 2000.

[22] CDF and D0 Collaborations, "Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with up to 6.7 fb^{-1} of Data," in *Proceedings, 35th International Conference on High energy physics (ICHEP 2010): Paris, France, July 22-28, 2010*, 2010.

[23] CDF Collaboration, "Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab," *Phys. Rev. Lett.*, vol. 74, pp. 2626–2631, 1995.

[24] D0 Collaboration, "Observation of the Top Quark," *Phys. Rev. Lett.*, vol. 74, pp. 2632–2637, 1995.

[25] CDF Collaboration, "Observation of $B_s^0 - \overline{b}_s^0$ oscillations," *Phys. Rev. Lett.*, vol. 97, p. 242003, 2006.

[26] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys.Lett.*, vol. B716, pp. 30–61, 2012.

[27] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys.Lett.*, vol. B716, pp. 1–29, 2012.

[28] CMS Collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS, Geneva: CERN, 1997.

[29] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *JINST*, vol. 9, no. 10, p. P10009, 2014.

[30] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV," *JINST*, vol. 10, no. 06, p. P06005, 2015.

[31] CMS Collaboration, "Performance of muon identification in pp collisions at $\sqrt{s}$ = 7 TeV," no. CMS-PAS-MUO-10-002, 2010.

[32] CMS Collaboration, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV," *JINST*, vol. 7, p. P10002, 2012.

[33] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector," *Journal of Instrumentation*, vol. 12, no. 10, pp. P10003–P10003, 2017.

[34] M. Cacciari and G. P. Salam, "Dispelling the N^3 myth for the k_t jet-finder," *Phys. Lett. B*, vol. 641, p. 57, 2006.

[35] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

[36] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS," *JINST*, vol. 6, p. P11002, 2011.

[37] CMS Collaboration, "Pileup jet identification," CMS Physics Analysis Summary CMS-PAS-JME-13-005, CERN, 2013.

[38] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV," *JINST*, vol. 13, no. 05, p. P05011, 2018.

[39] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector," *JINST*, vol. 12, no. 10, p. P10003, 2017.

[40] CMS Collaboration, "Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV," *JHEP*, vol. 11, p. 047, 2017.

[41] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual," *Eur. Phys. J.*, vol. C72, p. 1896, 2012.

[42] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *Journal of High Energy Physics*, vol. 2014, no. 7, p. 79, 2014.

[43] P. Nason, "A new method for combining NLO QCD with shower monte carlo algorithms," *Journal of High Energy Physics*, vol. 2004, no. 11, pp. 040–040, 2004.

[44] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method," *Journal of High Energy Physics*, vol. 2007, no. 11, pp. 070–070, 2007.

[45] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX," *Journal of High Energy Physics*, vol. 2010, no. 6, p. 43, 2010.

[46] T. Sjostrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Computer Physics Communications*, vol. 191, pp. 159 – 177, 2015.

[47] GEANT4 Collaboration, "Geant4, a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250 – 303, 2003.

[48] K. Hamilton, P. Nason, and G. Zanderighi, "MINLO: multi-scale improved NLO," *Journal of High Energy Physics*, vol. 2012, no. 10, p. 155, 2012.

[49] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano, "H$W^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO," *Journal of High Energy Physics*, vol. 2013, no. 10, p. 83, 2013.

[50] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO," *Journal of High Energy Physics*, vol. 2012, no. 12, p. 61, 2012.

[51] J. Alwall, S. Höche, F. Krauss, N. Lavesson, L. Lönnblad, F. Maltoni, M. Mangano, M. Moretti, C. Papadopoulos, F. Piccinini, S. Schumann, M. Treccani, J. Winter, and M. Worek, "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions," *The European Physical Journal C*, vol. 53, no. 3, pp. 473–500, 2008.

[52] S. Frixione, G. Ridolfi, and P. Nason, "A positive-weight next-to-leading-order monte carlo for heavy flavour hadroproduction," *Journal of High Energy Physics*, vol. 2007, no. 09, pp. 126–126, 2007.

[53] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method," *The European Physical Journal C*, vol. 71, no. 2, p. 1547, 2011.

[54] R. Frederix, E. Re, and P. Torrielli, "Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO," *Journal of High Energy Physics*, vol. 2012, no. 9, p. 130, 2012.

[55] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions," *Journal of High Energy Physics*, vol. 2009, no. 09, pp. 111–111, 2009.

[56] R. D. Ball, V. Bertone, S. Carrazza, L. D. Debbio, S. Forte, P. Groth-Merrild, A. Guffanti, N. P. Hartland, Z. Kassabov, J. I. Latorre, E. R. Nocera, J. Rojo, L. Rottoli, E. Slade, and M. Ubiali, "Parton distributions from high-precision collider data," *The European Physical Journal C*, vol. 77, no. 10, p. 663, 2017.

[57] LHC Higgs Cross Section Working Group, C. Anastasiou, D. de Florian, C. Grojean, F. Maltoni, C. Mariotti, A. Nikitenko, M. Schumacher, and R. Tanaka (Eds.), "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector," *arXiv*, vol. 1610.07922, 2016.

[58] S. Kallweit, J. M. Lindert, P. Maierhofer, S. Pozzorini, and M. Schnherr, "NLO QCD+EW predictions for V + jets including off-shell vector-boson decays and multijet merging," *JHEP*, vol. 04, p. 021, 2016.

[59] CMS Collaboration, "Inclusive Search for a Highly Boosted Higgs Boson Decaying to a Bottom Quark-Antiquark Pair," *Phys. Rev. Lett.*, vol. 120, p. 071802, 2018.

[60] T. Aaltonen, A. Buzatu, B. Kilminster, Y. Nagai, and W. Yao, "Improved b-jet Energy Correction for $H \to b\bar{b}$ Searches at CDF," 2011.

[61] CMS Collaboration, "Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks," *Phys. Rev.*, vol. D89, no. 1, p. 012003, 2014.

[62] CMS Collaboration, "Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with $h \to \tau\tau$," *Physics Letters B*, vol. 755, pp. 217 – 244, 2016.

[63] M. Tanabashi et al. *Phys. Rev. D*, vol. 98, p. 030001, 2018.

[64] J. B. Diederik P. Kingma, "Adam: A method for stochastic optimization," *arXiv*, vol. 1412.6980, 2014.

[65] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka (Eds.), *et al.*, "Handbook of LHC Higgs Cross Sections: Inclusive Observables," 2011.

[66] G. Ferrera, M. Grazzini, and F. Tramontano, "Associated WH production at hadron colliders: a fully exclusive QCD calculation at NNLO," 2011.

[67] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV," Tech. Rep. CMS-PAS-LUM-17-004, CERN, 2018.

[68] J. Conway, "Incorporating nuisance parameters in likelihoods for multisource spectra," *arXiv*, vol. 1103.0354, 2011.

[69] CMS Collaboration, "Evidence for the Higgs boson decay to a bottom quark-antiquark pair," *Physics Letters B*, vol. 780, pp. 501 – 532, 2018.

[70] CMS Collaboration, "Search for the standard model higgs boson produced through vector boson fusion and decaying to $b\bar{b}$," *Phys. Rev. D*, vol. 92, p. 032008, 2015.

[71] ATLAS Collaboration, "Observation of H$\rightarrow$b$\bar{\text{b}}$ decays and VH production with the ATLAS detector," *Physics Letters B*, vol. 786, pp. 59 – 86, 2018.