Title

Search for Exotic Decays of the Higgs Boson to a Pair of New Light Bosons in the Final States with Four Bottom Quarks

Author

Ghiasi Shirazi, Seyyed Mohammad Amin

Publication Date

2018

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Search for Exotic Decays of the Higgs Boson to a Pair of New Light Bosons in the Final States with Four Bottom Quarks

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Seyyed Mohammad Amin Ghiasi Shirazi

March 2019

Dissertation Committee:

Dr. Owen Long, Chairperson
Dr. William Gary
Dr. Hai-Bo Yu

The Dissertation of Seyyed Mohammad Amin Ghiasi Shirazi is approved:

Committee Chairperson

University of California, Rivers

ABSTRACT OF THE DISSERTATION

Search for Exotic Decays of the Higgs Boson to a Pair of New Light Bosons in the Final States with Four Bottom Quarks

by

Seyyed Mohammad Amin Ghiasi Shirazi

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, March 2019
Dr. Owen Long, Chairperson

In this dissertation, we present a search for non-standard decay of the Standard Model-like Higgs boson to a pair of pseudoscalar lighter bosons $H \rightarrow aa$ with $12\, Gev < m_a < 60\, GeV$, as predicted in models with extended Higgs sectors like the next-to-minimal supersymmetric standard model. The search is done using 35.9 fb^{-1} of 13 TeV center of mass pp collision data recorded in 2016 by the Compact Muon Solenoid experiment at the Large Hadron Collider.

In this analysis, only the W Higgs associated production is considered to characterize the final state by one isolated high p_T lepton. Then, the MVA BDT classifier is used to distinguish between signal and background. And finally, The CLs method is used in the three and four reconstructed b-quark regions to set upper limits on the branching ratio $Br(H \rightarrow aa) \times Br^2(a \rightarrow b\bar{b})$. Unfortunately, at the time of writing this dissertation, the analysis is not unblinded yet. Therefore, the upper limit plot lacks the observed curve.

Table of Contents

Table of Figures

Table of Tables

List of Acronyms

ANN	**Artificial Neural Network**
APD	**Avalanche PhotoDiodes**
BC	**Before Christ**
BDT	**Boosted Decision Tree**
BSM	**Beyond Standard Model**
CMS	**Compact Muon Solenoid**
CP	**Charge conjugation and Parity**
CR	**Control Region**
CSC	**Cathode Strip Chambers**
CSV	**Combined Secondary Vertex**
CTF	**Combinatorial Track Finder**
DGLAP	**Dokshitzer-Gribov-Lipatov-Altarelli-Parisi**
DY	**Drell–Yan**
DT	**Drift Tube**
ECAL	**Electromagnetic CALoremeter**
FSR	**Final-State Radiation**
GEANT	**GEometry ANd Tracking**
HB	**Hadron Barrel**
HCAL	**Hadronic CALoremeter**
HE	**Hadron Endcap**

HF	**Heavy Flavor/Hadron Front**
HLT	**High Level Trigger**
HO	**Hadron Outer**
HPD	**Hybrid PhotoDiodes**
ISR	**Initial-State Radiation**
JHEP	**Journal of High Energy Physics**
L1T	**Level-1 trigger**
LF	**Light Flavor**
LHC	**Large Hadron Collider**
LHCf	**Large Hadron Collider forward**
LSP	**Lightest Supersymmetric Particle**
MC	**Monte Carlo**
MFV	**Minimal Flavor Violation**
MoEDAL	**Monopole and Exotics Detector at the LHC**
MSSM	**Minimal Supersymmetric Standard Model**
NMSSM	**Next-to-Minimal Supersymmetric Standard Model**
NP	**Nuisance Parameter**
POI	**Parameter Of Interest**
POWHEG	**POsitive-Weight Hardest Emission Generator**
PDF	**Parton Density Function also Probability density function**
PF	**Particle Flow**
PL	**Profile Likelihood**

pMSSM	**Phenomenological Minimal Supersymmetric Standard Model**
POG	**Physics Object Group**
PQ	**Peccei-Quinn**
PV	**Primary Vertices**
QCD	**Quantum ChromoDynamics**
QED	**Quantum ElectroDynamic**
ROC	**Receiver Operating Characteristics**
RPC	**Resistive Plate Chambers**
SF	**Scale Factor**
SM	**Standard Model**
SR	**Signal Region**
SUSY	**SUperSYmmetry**
TEC	**Tracker Inner Endcaps**
TIB	**Tracker Inner Barrel**
TID	**Tracker Inner Discs**
TOB	**Tracker Outer Barrel**
ToTeM	**TOTal cross-section, Elastic scattering and diffraction dissociation Measurement at the LHC**
VPT	**Vacuum PhotoTriodes**
WIMP	**Weakly Interacting Massive Particle**
WLS	**WaveLength Shifting**

Introduction

The idea that matter is made of discrete units goes back to round 440 BC in ancient Greece as a philosophical concept[1]. After around 2200 years, chemists in the early 19th century observed in their laboratories that matter indeed behaves as if it were made up of atoms. The word “atom” originates from the Ancient Greek adjective “atomos”, meaning "indivisible".[2] The theoretical explanation of the Brownian motion by Albert Einstein in 1995 can be considered the first proof of the atomic theory[3]. However, discovering the subatomic world started by the discovery of the electron by J.J. Thomson[4] in 1897 and continued by the discovery of the proton (in 1911 by Ernest Rutherford)[5] and the neutron (in 1932 by James Chadwick)[6]. The fairly simple picture of the world made up of three particles, namely electrons, protons, and neutrons had transformed entirely by the mid-1960s, when the experimentalists discovered so many different species that scientists started talking about a particle zoo. Alongside experimentalists, the theorists also continued shedding light on the mysteries of the universe and developed the Quantum Field Theory and then the Standard Model. In the second half of the twentieth century, high energy physics was driven by a sequence of increasingly more powerful particle accelerators; the LHC accelerator being the newest, most extensive, and highest energy so far. [7]

Chapter 1 Theory and Motivation

1.1 The Standard Model

The Standard Model is our best current understanding of the fundamental nature of the universe [8] It tries to explain all the measurements in particle physics using the properties and interactions of a handful of particles of four distinct types:

- Two families of fermions named leptons and quarks; Fermions are particles that have half-integer spin
- One family of bosons - named gauge bosons – Bosons are ‘force carriers’ and have integer spin;
- And a spin-0 particle called the Higgs boson, which is responsible for the origin of mass within the theory. Without the Higgs boson leptons, quarks, and gauge bosons would all be massless according to the SM.

All the particles in the standard model are assumed to be elementary: that is they are treated as point particles, without internal structure or excited states[9]. Figure 1-1 depicts the elementary particles that the SM embodies. All these particles have been experimentally observed with the last being the discovery of the Higgs Boson at CERN in 2012.

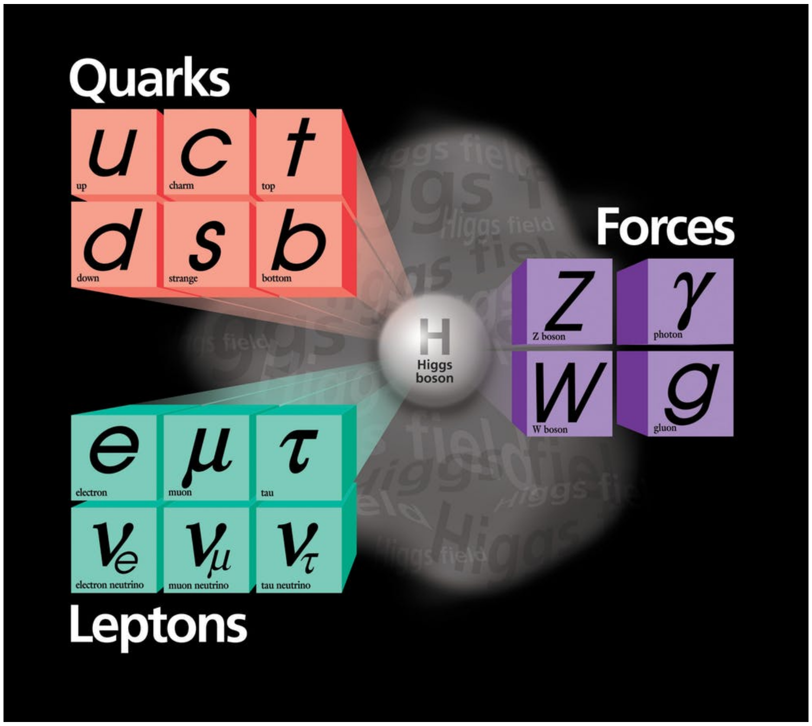


Figure 1-1: The Standard Model tries to explain the world using only a handful of elementary particles. This image is taken from [10]

The SM describes three forces between particles: Electromagnetic, weak and strong forces. Although it does not formulate gravity, it is not an urgent problem since gravity is negligible at subatomic scales. Quantum Chromodynamics (QCD) is the formulation of the strong force based on the SU(3) non-Abelian gauge group, and the electroweak theory is a unified formulation for the weak and electromagnetic forces based on an $SU(2) \times SU(1)$ symmetry. Therefore, the Standard Model belongs to the symmetry group:

$$SU(3)_C \times SU(2)_L \times SU(1)_L \qquad (1)$$

Where C stands for color, L denotes that only the left-handed fermions transform nontrivially, and Y shows weak hypercharge; the generator of U(1) [11]

1.1.1 Quantum Chromodynamics (QCD)

In early days of nuclear physics, it was a puzzle how protons could stay so close when they should repel each other electromagnetically. This puzzle was solved by proposing a new force called the strong force. In 1935, Hideki Yukawa proposed a massive boson mediator responsible for the short range of the strong interaction.[12] However, this picture was entirely changed by mid-1960s when the QCD theory was developed. [7]

Nowadays, we know that not only is the strong force responsible for the binding of protons and neutrons (long range about 1 to 3 fm) but also it is responsible for keeping quarks together in protons and neutrons (less than about 0.8 fm, the radius of a nucleon). In the latter case, it is also called the color force [13]

The theory of the strong interactions between quarks and gluons is called Quantum ChromoDynamics (QCD). The underlying symmetry is non-Abelian SU(3), and the charge is called color which can be red, green, and blue. So far, we have not observed any free color charges. In other words, quarks create groups that are neutral together in color. Any particle that is strongly interactive is called a Hadron. Two groups of Hadrons have been observed so far[1] [7]:

- Mesons are composed of a quark-antiquark pair
- Baryons are composed of three quarks or three antiquarks

---

[1] The LHCb collaboration at CERN confirmed the observation of pentaquarks, i.e. a particle with five quarks (14. Aaij, R., et al., *Observation of J/ψ p Resonances Consistent with Pentaquark States in Λ b 0→ J/ψ K− p Decays*. 2015. **115**(7): p. 072001.) in 2015. Even more quarks binding in one particle is still theatrically possible. However, in this thesis, we will only talk about Mesons and Baryons.

Gluons are the mediators of the strong force. They are massless and electrically neutral. Gluons are linear combinations of the eight generators of the SU(3) group corresponding to eight gauge fields $G_{a\mu}$. [15]

1.1.1.1 The strong coupling constant

The Lagrangian density of the QCD reads:

$$\begin{aligned} \mathcal{L}_{QCD}(x) = & -\frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a + i\sum_{j=1}^{n}\bar{\psi}_j^\alpha\gamma^\mu\left(D_\mu\right)_{\alpha\beta}\psi_j^\beta - \sum_{j=1}^{n}m_j\bar{\psi}_j^\alpha\psi_{j,\alpha} \\ & -\frac{1}{2\alpha_G}\partial^\mu A_\mu^a\partial_\mu A_a^\mu - \partial_\mu\bar{\varphi}_a D^\mu\varphi^a \end{aligned} \tag{2}$$

Where:

- $G_{\mu\nu}^a = [\partial_\mu A_v^a - \partial_v A_\mu^a + g_S f_{abc}A_\mu^b A_v^c]; a \equiv 1,2,\dots,8$ are the Yang-Mills field strength constructed from the gluon fields A_μ^a; f_{abc} are real structure constants which close the SU(3) Lie Algebra
- ψ_j is the field of the quark flavor j
- $\varphi_a(x)$ are eight anti-commuting scalar fields in the 8 of SU(3)
- $\left(D_\mu\right)_{\alpha\beta} \equiv \delta_{\alpha\beta}\partial_\mu - ig\sum_a \frac{1}{2}\lambda_{\alpha\beta}^a A_\mu^a$ are the covariant derivatives acting on the quark color components $\alpha, \beta \equiv \text{red, blue, and yellow}$; $\lambda_{\alpha\beta}^a$ are the eight 3×3 color matrices[16]

The strong coupling constant is defined as:

$$\alpha_S = \frac{g_S^2}{4\pi} \tag{3}$$

If we use the perturbation theory to calculate a process, it will include loop diagrams of higher-order contributions. The 4-dimensional integrations over those loop momenta will diverge. These divergences are called ultraviolet divergences. To sweep the divergences under the carpet, we need a renormalization procedure. The bare parameters of the theory, like the masses, field strengths, and coupling constants are modified, such that they contain counter terms to absorb these divergences.

No matter what renormalization method we use, the coupling constants change with scale. In other words, if we normalize a theory in a scale μ_0, we need to calculate the value of coupling constants in another scale μ. This means that the coupling constants are not actually constant anymore and they run with momentum, a phenomenon commonly called running coupling constants [17, 18] If we calculate the dependency at the leading order, we have:

$$\alpha_S(\mu^2) = \frac{\alpha_S(\mu_0^2)}{1 + \alpha_S(\mu_0^2)\beta_0 \ln\left(\frac{\mu^2}{\mu_0^2}\right)} \tag{4}$$

Where $\beta_0 = 11 - \frac{2}{3}n_f$ where n_f is the number of quark flavors

Note how $\alpha_S \rightarrow 0\ and\ g_S \rightarrow 0$ when $\mu \rightarrow \infty$. This means that any strong interaction is proportional to the powers of g_S^2. When $g_S \rightarrow 0$, it means that the strong interaction disappears, and the quark or gluon becomes free. This behavior is called asymptotic freedom, and it allows us to use the perturbation theory at high energies. However, at low energies where QCD is non-perturbative, the behavior of the strong force is more complex.

1.1.2 Electroweak Theory

The electroweak theory is a perfect example of how the forces we consider different can be unified. Before Maxwell's work in the 1850s, the electric and magnetic forces were deemed to be separate. However, his work unified the forces in one formula:

$$F = q\left\{E + \left[\frac{v}{c} \times H\right]\right\} \tag{5}$$

The relative ratio of electric and magnetic forces depends on the velocity and $|E| \sim |H|$ only at relativistic speeds. Therefore, we can say that the scale in which the unification of electric and magnetic forces happens is the ultrarelativistic energies of particles. About 100 years after Maxwell, the Quantum Electrodynamics (QED) could explain the electromagnetic interactions very accurately and have the most accurate agreement between theory and experiment so far. [19]

After the unification of the electric and magnetic forces, it was the turn of the weak force to be unified with the electromagnetic force. Glashow, Weinberg, and Salam proposed the structure which led to their noble prize in 1979. The scale at which the Electroweak unification happens is called the electroweak scale (Order of 100 GeV). It seems that the next step could be to unify the electroweak and strong forces and Supersymmetry is one of the theories that have a structure for this unification as we will discuss later.

1.1.2.1 Massive Electroweak Force Carriers

The Electroweak Lagrangian density can be written as:

$$\mathcal{L}_{EW} = \mathcal{L}_g + \mathcal{L}_f \tag{6}$$

Where

$$\mathcal{L}_g = -\frac{1}{4}W^{i,\mu v}W^i_{\mu v} - \frac{1}{4}B^{\mu v}B_{\mu v} \tag{7}$$

describes the interaction between the three W vector bosons and the B vector boson. And the kinetic term for the Standard Model fermions is:

$$\mathcal{L}_f = \bar{\psi}_L\gamma^\mu\left(i\partial_\mu - g\frac{1}{2}\sigma^i.W^i_\mu - g'\frac{Y}{2}B_\mu\right)\psi_L + \bar{\psi}_R\gamma^\mu\left(i\partial_\mu - g'\frac{Y}{2}B_\mu\right)\psi_R \tag{8}$$

With

$$W^i_\mu = \partial_\mu W^i_v - \partial_v W^i_\mu - g\varepsilon_{ijk}W^j_\mu W^k_v \text{ and } B_{\mu v} = \partial_\mu B_v - \partial_v B_\mu \tag{9}$$

Where W^i_μ are the $SU(2)_L$ gauge fields, $\psi_{L/R}$ are left/right-handed fermion fields, B_μ is the $U(1)_Y$ gauge field, and g and g' are the coupling constants of the gauge to the fermion fields. The hypercharge (Y) is the generator of U(1) algebra. [20, 21]

Although photon was known to be massless at the time of electroweak theory development, W and Z were expected to have mass due to the short range of the weak force. Also, some experiments supported this idea. For example, measurements of the lifetime of the muon allowed us to calculate the mass of the W.[22]

We can include the mass of W and Z in the Lagrangian manually:

$$\mathcal{L}_m = M_Z^2 Z_\mu Z^\mu + M_W^2 W^+_\mu W^{-,\mu} - \left[\bar{u}_L M^u u_R + \bar{d}_L V^\dagger M^d d_R + \bar{e}_L M^e e_R + h.c.\right] \tag{10}$$

Where V is the 3 × 3 unitary CKM matrix.[20]

However, three problems arise with this Lagrangian:

- The mass terms that we added to the Lagrangian violates the gauge invariance.
- The Feynman propagator of a massive spin-1 particle is

$$-i\frac{g_{\mu\nu}+q_{\mu}q_{\nu}/M^2}{q^2-M^2+i\varepsilon} \tag{11}$$

  The longitudinal parts, which increase together with q, yield the disturbing property of a theory of massive spin-1 particles, namely that Feynman graphs including closed loops diverge faster than logarithmically. As a result, renormalization constants cannot absorb the divergences, that is, the massive spin-1 quantum field theory is not renormalizable.[23]
- The scattering amplitude for longitudinally polarized W bosons, W_L, scales as E^2 for high energies E.[20]

Therefore, adding the mass terms manually to the Lagrangian does not give us a consistent theory.

1.1.2.2 The Electroweak Spontaneous Symmetry Breaking

Spontaneous symmetry breaking is the trick to let the Lagrangian retain the symmetry invariance but drop it for physical states. The Higgs mechanism, also known as the Brout-Englert-Higgs mechanism, is the simplest model capable of generating massive intermediate bosons using spontaneous symmetry breaking. [23]

In the Higgs mechanism, there is a complex scalar field ϕ:

$$\phi=\frac{1}{\sqrt{2}}(\phi_1+i\phi_2) \tag{12}$$

With the corresponding Lagrangian being:

$$\mathcal{L} = \left(\partial_\mu \phi\right)^\dagger \left(\partial_\mu \phi\right) - V(\phi)\ with\ V(\phi) = \mu^2(\phi^\dagger\phi) - \lambda(\phi^\dagger\phi)^2 \qquad (13)$$

For $\mu^2 < 0$ and $\lambda > 0$, the minimum of the potential does not happen at $\phi = (0,0)$. (Figure 1-2) In other words, the potential gives us the spontaneous symmetry breaking while the Lagrangian respects the symmetry. By expanding the potential around the vacuum state, we will get quadratic terms in Lagrangian accounting for the mass of particles and some other interaction terms.[22]

Another way to look at the Higgs mechanism is using the Goldstone theorem. According to the Goldstone theorem, since we have spontaneously broken continuous symmetries, we get four extra massless particles in the Lagrangian. In the Higgs mechanism, three of the four fields pair up with $W^+,\ W^- and\ Z$ to give them mass and the last one is responsible for the Higgs boson.[24] The Higgs Boson gains its mass from a self-interaction term in Lagrangian.

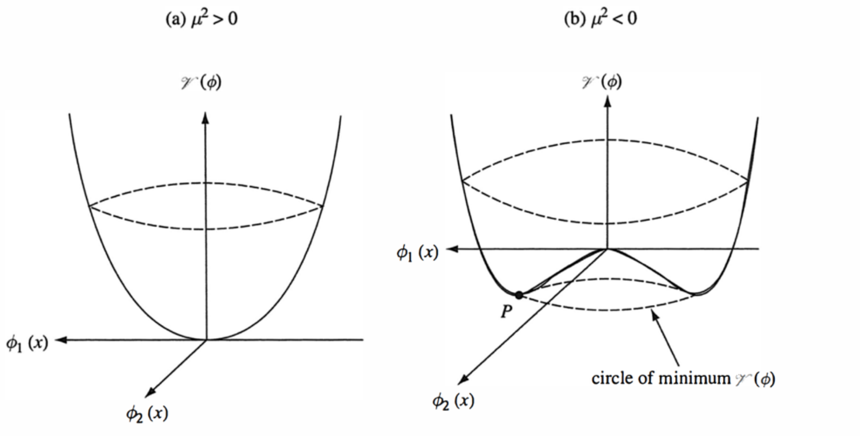


Figure 1-2: If $\mu^2 < 0$ while $\lambda > 0$ in the Higgs mechanism, we get the Mexican hat potential on the right with a non-zero potential minimum. This image is taken from [25]

1.1.3 Proton-Proton collisions at LHC

In this thesis, we have used the data taken by the CMS detector from proton-proton collisions happening at the LHC. Hence, a decent knowledge of both the structure of the proton and the physics of the proton-proton collisions are necessary.

1.1.3.1 The Parton Model

The proton is not an elementary particle and is composed of two up and one down quark ($|p> = |uud>$). However, at the time of collisions, there might be gluons inside the proton binding the quarks and fleeting pairs of quark-antiquark that the gluons have produced. Any strongly-interacting particle inside the proton is called a parton. Due to the substantial momentum transfer at LHC, partons have asymptotic freedom (explained in section 1.1.1.1 The strong coupling constant), and therefore they are almost free and have minimum interaction with other partons of the proton. As a result, we can forget protons and think of LHC as a parton collider when looking at a single collision. [26]

Although we know that the energy of each proton is 6.5 TeV, we cannot know for sure what portion of the momentum is carried by each parton and especially by the colliding parton(s). Instead, we can use probability. The Parton Density Function (PDF) is defined as the probability density for finding a parton with a particular longitudinal momentum fraction x at resolution scale μ_F^2 [2]. [27] PDFs cannot be calculated theoretically due to the non-perturbative nature of QCD at lower energies and need to be measured experimentally

---

2- This is the leading-order picture of the PDFs. Generalized Parton Distributions (GPDs) are a more recent approach to better understand the hadron structure. [27. Wikipedia_contributors, *Parton (particle physics)*, in *Wikipedia, The Free Encyclopedia*. 2018.]

at a given scale. [20] However, when we have the PDF for a given scale μ_F, the evolution is described by the DGLAP[3] formula:

$$\mu_F^2 \frac{\partial f_a(x,\mu_F^2)}{\partial \mu_F^2} = \frac{\alpha_s}{2\pi} \int_x^1 \frac{dz}{z} P_{ab}(z) f_b\left(\frac{x}{z}, \mu_F^2\right) \tag{14}$$

Where $f_q(x,\mu_F^2)$ is the PDF and $P_{ab}(z)$ is the Altarelli-Parisi splitting function representing the probability of parton splitting $a \to bc$ with longitudinal momentum $p_b = zp_a$ [28, 29] There are several PDFs based on slightly different sets of data, and with some variation in the theoretical assumptions. [30] By knowing PDFs, the inclusive cross-section of the production of a final state X in the collision of hadrons h_1 and h_2 is:

$$\sigma_{h_1 h_2 \to X} = \sum_{a.b \in \{q,g\}} \int dx_a \int dx_b \, f_a^{h_1}(x_a, \mu_F^2) \, f_a^{h_2}(x_b, \mu_F^2) \int d\phi_{ab\to X} \, \frac{d\sigma_{ab}(\phi_{ab\to X}, \mu_F^2)}{\phi_{ab\to X}} \tag{15}$$

Where $f_a^{h_1}(x_a, \mu_F^2)$ are the PDFs. [29]

3- Dokshitzer–Gribov–Lipatov–Altarelli–Parisi

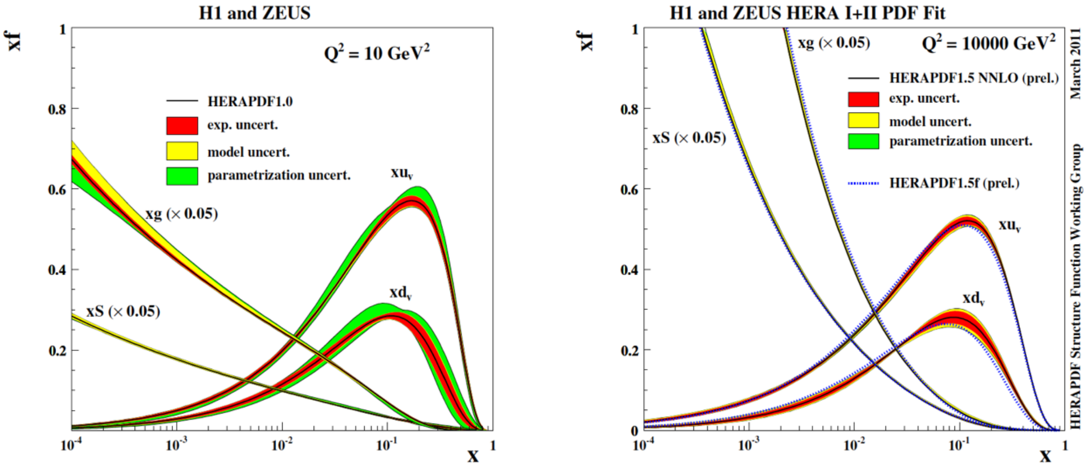


Figure 1-3: The PDFs from HERAPDF at $\mu_F^2 = 10\ GeV$ *(Left) and at* $\mu_F^2 = 10{,}000\ GeV$ *(right). The images are taken from [31]*

1.1.3.2 Physics of the proton-proton collisions

Many different phenomena can happen in a proton-proton collision. Figure 1-4 shows the most important correction processes that can happen in a proton-proton collision.

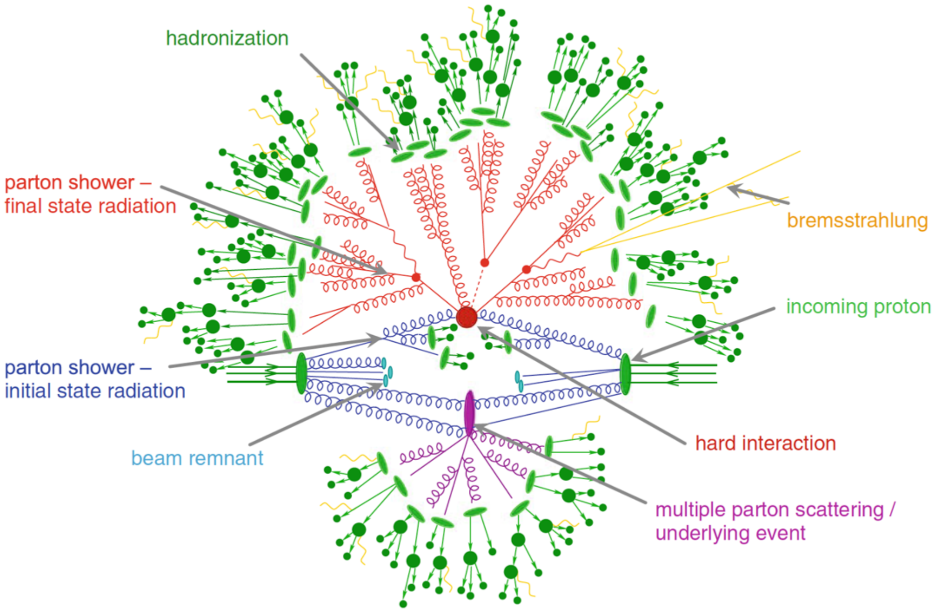


Figure 1-4: Illustration of a typical proton-proton collision event. The incoming partons are in blue, red shows hard scattering, underlying events are in purple, and hadronization is shown in green. The image is taken from [20]

Typical evolution of a collision can be considered as follows. This part is taken from [30]

1. Initially, two beam particles are coming in towards each other. Normally each particle is characterized by a set of parton distributions, which defines the partonic substructure in terms of flavor composition and energy sharing.
2. One shower initiator parton from each beam starts off a sequence of branchings, such as q → qg, which build up an initial-state shower.
3. One incoming parton from each of the two showers enters the hard process, where then a number of outgoing partons are produced, usually two. It is the nature of this process that determines the main characteristics of the event.
4. The hard process may produce a set of short-lived resonances, like the Z0/W± gauge bosons, whose decay to normal partons has to be considered in close association with the hard process itself.

5. The outgoing partons may branch, just like the incoming did, to build up final-state showers.
6. In addition to the hard process considered above, further semi hard interactions may occur between the other partons of two incoming hadrons.
7. When a shower initiator is taken out of a beam particle, a beam remnant is left behind. This remnant may have an internal structure and a net color charge that relates it to the rest of the final state.
8. The QCD confinement mechanism ensures that the outgoing quarks and gluons are not observable, but instead fragment to color neutral hadrons.
9. Normally the fragmentation mechanism can be seen as occurring in a set of separate color singlet subsystems, but interconnection effects such as color rearrangement or Bose-Einstein may complicate the picture.
10. Many of the produced hadrons are unstable and decay further.

1.1.3.2.1 Initial- and final-state radiation

When a process contains colored or charged particles in the initial or final state, partons emit gluon, photon or split into quarks. This process is called initial (final) state radiation (ISR/FSR), and the emitted photons are also called "bremsstrahlung". Categorizing the radiation into ISR and FSR is arbitrary, but very convenient. ISR and FSR can play an important role in the overall topology of the event. For instance, they can change a $2 \rightarrow 2$ basic process to a $2 \rightarrow 3$, $2 \rightarrow 4$ and more final-state topologies. [20, 30]

There are two conventional approaches to model the perturbative corrections:

- One is the matrix-element method is which the Feynman diagrams are calculated, order by order. This method is the correct approach in principle. However, the calculations become increasingly heavier in higher orders. [30]
- The second approach is called the parton showers. There are many details to be considered in the parton shower method. However, the main idea is to use the DGLAP equation (eq.14) to convert the inclusive prediction for the occurrence of parton a in the beam hadron h into an exclusive prediction for parton a and a certain number of additional particles, which are resolved at smaller and smaller momentum transfer. [29] (Figure 1-5)

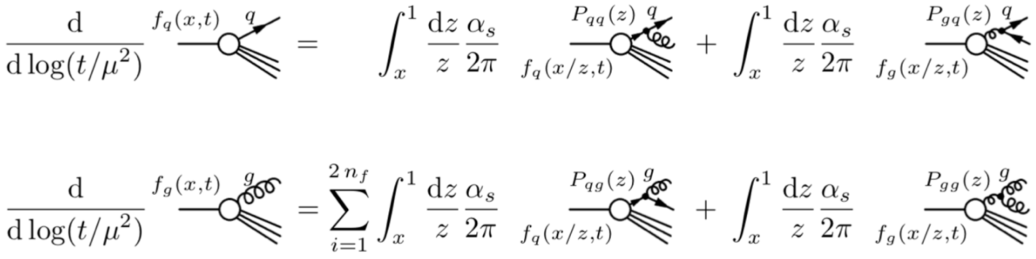


Figure 1-5: Pictorial representation of the DGLAP evolution of PDFs. The white blobs are the incoming hadron. The image is taken from [29]

1.1.3.2.2 Beam remnants

For simplicity, let's assume that the only partons in a proton are two up quarks and one down quark (uud). If in a collision, an up quark collides with a parton from the other proton, we will have a ud diquark beam remnant. Since this diquark is not color-neutral, it is color-connected to the hard interaction and forms part of the same fragmentation system. [30]

1.1.3.2.3 Multi-parton interactions

So far in this part, we have assumed that when proton A and proton B collide, always one single parton in proton A collides with a single parton in proton B. However, the picture can be more complex. Meaning that we can have two simultaneous collisions between protons A and B. In other words, we can have two different pairs of partons in protons A and B collide independently. Alternatively, we can have a single parton in proton A scatters against several different partons in the other beam. These interactions are called “Multi-parton interactions”. [30]

1.1.3.2.4 Hadronization

The next step in an event evolution is what happens to the single partons from the abovementioned phenomena. At this point, the partons have low virtualities (μ_F^2), and they cannot be considered free anymore. In other words, the strong coupling constant is large again and brings the partons together to build color-neutral hadrons. This process is called “Hadronization” or “Fragmentation”. [20, 30, 32]

If the hadrons are still unstable, they decay further to more hadrons, leptons, and photons until the particles are stable[4]. This process creates a cluster of particles moving more or less in the direction of the mother parton. In experiments, this cluster is called a “jet”. In order to simulate hadronization, we cannot use perturbation theory anymore (Because the strong

---

4- In an experiment, we might observe an unstable hadron if its lifetime is large enough to reach to the detector.

coupling constant is large) and have to use the phenomenological models: The String model or the cluster model.[5] [20, 30, 32]

- The string model comes from the observation that the quark-antiquark potential is linear. i.e., $V(r) = \kappa r$ where r is the distance between the quark and anti-quark pair and $\kappa \approx 1\ GeV/fm$. In a pair of quark-antiquark, the potential energy stored in the string stretched between them can result in a spontaneous creation of a new quark-antiquark pair and therefore a new hadron. While meson production is explained immediately in this model, baryon creation is more complicated. [20, 30, 32]
- The cluster model is based on the so-called preconfinement property of QCD. In other words, the cluster model requires all gluons at the end of the shower to split into qq pairs. Then, clusters are formed by connecting neighboring quarks to color-singlets. The clusters decay isotopically into a pair of hadrons. The energy needed for the hadron formation by the qq production is extracted from the cluster field. Clusters with too low energy will decay into only one hadron, while massive clusters can split into lighter daughter clusters. [20, 30, 32]

5- There is a third school called the independent model that is not used anymore. Therefore, we will not talk about it in this dissertation.

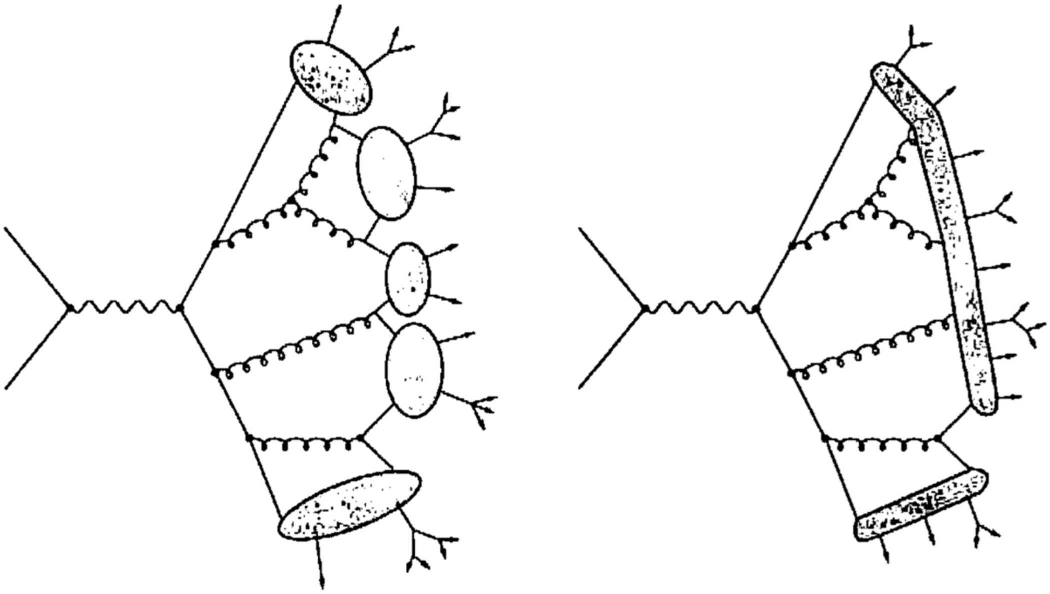


Figure 1-6: Sketches of the Hadronization models. Left: The cluster model and right: The string model [32]

1.1.3 Limitations of the Standard Model

The Standard model has been put to the test to unrivaled precision, and there has almost always been agreement with the experiment. Figure 1-7 shows the level of agreement between the predicted and observed cross-sections of various SM processes. However, now we know that the SM is not perfect. We will briefly explain the reasons in this section.

There are many candidate theories for beyond the standard model physics. However, none of them is experimentally verified so far. Among all BSM theories, we will only discuss MSSM and NMSSM that are related to this dissertation.

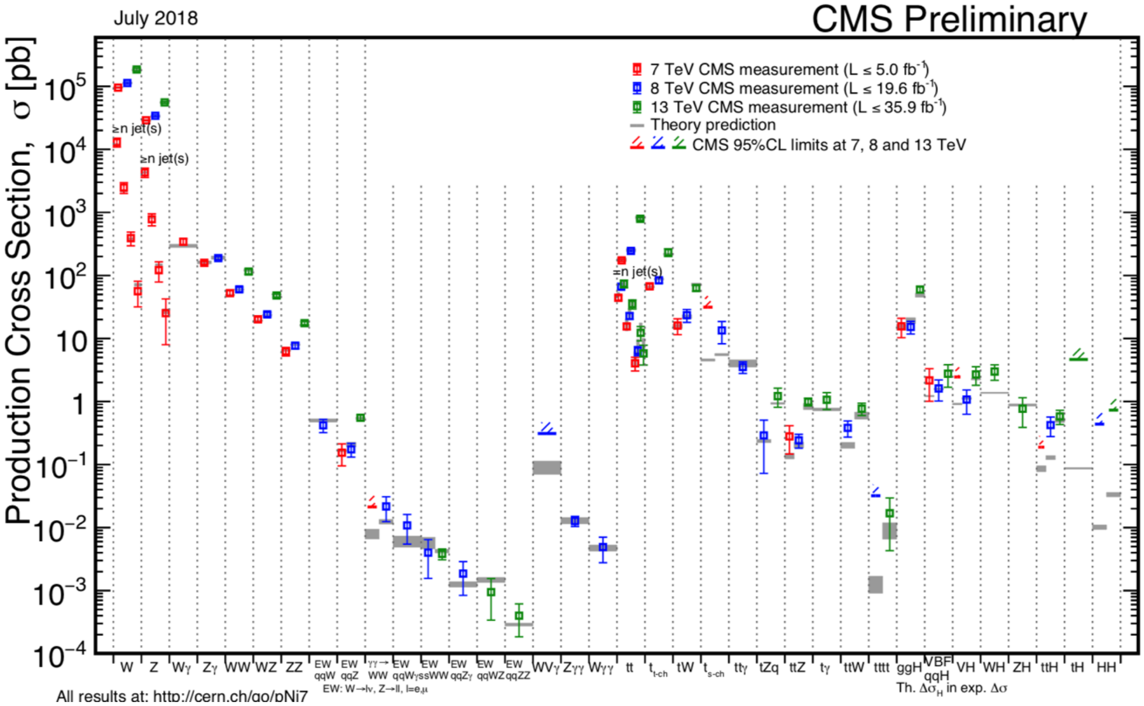


Figure 1-7: Summary of the cross-section measurements of the Standard Model processes by the CMS collaboration. [33]

1.1.3.1 Theoretical problems

Hierarchy problem

When we calculate the mass of the Higgs boson in the SM, we have loop corrections from all the particles that couple to the Higgs Boson. Figure 1-8 shows example loop corrections from fermions and scalars.

If we calculate the mass squared of the Higgs boson (m_H^2), the correction due to diagrams in Figure 1-8 to the first order is

$$\Delta m_H^2 = \frac{|\lambda_f|^2}{8\pi^2}\Lambda_{cutoff}^2 + \frac{\lambda_S}{16\pi^2}\left[\Lambda_{cutoff}^2 - 2m_S^2 \ln\left(\frac{\Lambda_{cutoff}}{m_S}\right)\right] + \cdots \tag{16}$$

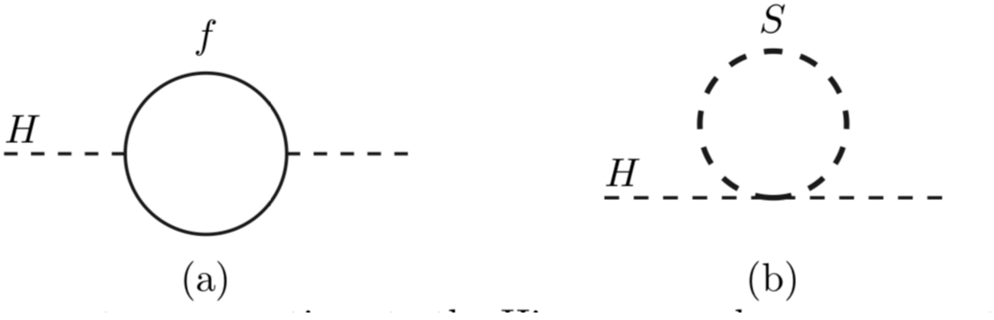


Figure 1-8: Single-loop quantum corrections to the Higgs mass squared (m_H^2) due to a fermion (left) and a scalar (right). This Image is taken from [34]

Where λ_f and λ_S are the Higgs couplings to the fermion and the scalar, respectively and Λ_{cutoff} is the cut off energy at which the SM breaks down and a new theory takes over. Examples of Λ_{cutoff} are the grand unification scale ($O(10^{14})GeV$) and the Plank scale ($O(10^{19})GeV$). Therefore, the corrections to the Higgs mass are orders of magnitude higher than its mass (125 GeV). Solving the problem by extreme fine-tuning of the SM parameters is possible but not favorable and the SM in incapable of giving us a more natural explanation. [34]

The hierarchy problem is not necessarily the most puzzling open question of the SM, however, it is certainly the most relevant to the collider experiments, since most likely, its resolution lies in the TeV energy range. [35]

Quantum triviality

As [36] exaplains: “In quantum field theory, charge screening can restrict the value of the observable renormalized charge of a classical theory. If the only resulting value of the renormalized charge is zero, the theory is said to be "trivial" or noninteracting. Substantial evidence supports the idea that a field theory involving only a scalar Higgs boson is trivial in four spacetime dimensions.”

Strong CP problem

The theory of QCD allows a CP[6] violation phase. However, measurements show that this phase is zero or very close to zero. This fine tuning seems unnatural.

1.1.3.2 Experimental results not explained

The anomalous magnetic dipole moment of the muon

The experimentally measured value of muon's anomalous magnetic dipole moment seems significantly different from the Standard Model prediction.

1.1.3.3 Phenomena not explained

Gravity

Although the SM is very successful in describing the electroweak and strong forces, it is incapable of describing the macroscopic world and the gravity. A hypothetical particle called graviton is proposed to be added to the SM; however, the SM is incompatible with the most successful theory of gravity so far, general relativity.[36]

Dark matter and Dark energy

Cosmological observations tell us that only 5% of the universe is made of ordinary matter, 23% dark matter and 72% dark energy. The nature of both dark matter and dark energy is unknown, and the SM fails to give us a candidate for dark matter. It also doesn’t have any explanation for the dark energy. [37]

6- The combination of the discrete symmetry operations charge conjugation and parity transformation

Neutrino masses

Although neutrinos are massless in the SM, the observation of neutrino flavor oscillation is a strong evidence that at least two neutrino generations have mass.

Matter-antimatter asymmetry

According to the SM during the creation of the universe an (almost) equal amount of matter and antimatter has been produced. SM cannot explain this large asymmetry between matter and antimatter that we see today.

1.2 Supersymmetry

As explained in the previous part, one of the major issues with the SM is the enormous corrections to the Higgs boson mass-squared parameter – the so-called hierarchy problem. One solution to this problem is to add a hypothetical symmetry – very different from what we have seen in nature so far - which relates bosons and fermions. In other words, Lagrangian is invariant under the operator that changes the spin of particles by ½. In other words, this operator transforms bosons to fermions and vice versa.

$$Q|boson> = |fermion> \quad ; \quad Q^{\dagger}|fermion> = |boson> \qquad (17)$$

Where Q is the generator of the supersymmetric transformation.

Quantum corrections of bosons have the opposite sign to the ones due to fermions. Therefore, if the masses and coupling constants of the particles are related, it is possible that the corrections cancel each other out.

1.2.1 Minimal Supersymmetric Standard Model

The first realistic supersymmetry extension to the SM was proposed in 1981 by Dimopoulos, Rabi, and Wilczek [38]. This theory proposed a partner for each elementary particle in the SM with the same quantum numbers except for spin, leading to what is called the Minimal Supersymmetric Standard Model (MSSM).

In MSSM, leptons and quarks receive scalar superpartners, sleptons and squarks. Also, there are two Higgs bosons in MSSM which get fermionic superpartners, the higgsinos. And the electroweak gauge bosons get fermionic superpartners. (Table 1-1)

Table 1-1: Supersymmetric particles in MSSM. This table is adapted from [39]

NAME	SPIN	GAUGE EIGENSTATES	MASS EIGENSTATES
HIGGSINO	0	$H_u^0\ H_d^0\ H_u^+\ H_d^-$	$h^0\ H^0\ A^0\ H^\pm$
SQUARKS	0	$\tilde{u}_L\ \tilde{u}_R\ \tilde{d}_L\ \tilde{d}_R$ $\tilde{s}_L\ \tilde{s}_R\ \tilde{c}_L\ \tilde{c}_R$ $\tilde{t}_L\ t_R\ \tilde{b}_L\ \tilde{b}_R$	(same) (same) $\tilde{t}_1\ \tilde{t}_2\ \tilde{b}_1\ \tilde{b}_2$
SLEPTONS	0	$\tilde{e}_L\ \tilde{e}_R\ \tilde{\nu}_e$ $\tilde{\mu}_L\ \mu_R\ \tilde{\nu}_\mu$ $\tilde{\tau}_L\ \tilde{\tau}_R\ \tilde{\nu}_\tau$	(same) (same) $\tilde{\tau}_1\ \tilde{\tau}_2\ \tilde{\nu}_\tau$
NEUTRALINOS	½	$\tilde{B}^0\ \widetilde{W}_0\ \widetilde{H}_u^0\ \widetilde{H}_d^0$	$\tilde{\chi}_1^0\ \tilde{\chi}_2^0\ X_3^0\ X_4^0$
CHARGINOS	½	$\widetilde{W}^\pm\ \widetilde{H}_u^+\ \widetilde{H}_d^-$	$\tilde{\chi}_1^\pm\ \tilde{\chi}_2^\pm$
GLUINO	½	$\tilde{g}$	(same)
GOLDESTINO-GRAVITINO	½ - 1 ½	$\tilde{G}$	(same)

1.2.1.1 Breaking the symmetry

After the addition of the supersymmetric particles to the SM, equation 16 changes to

$$\Delta m_H^2 = \frac{1}{8\pi}\left(\lambda_S - \left|\lambda_f\right|^2 \Lambda_{cutoff} + m_s^2\left(\frac{\lambda}{16\pi^2}\ln\frac{\Lambda_{cutoff}}{m_s}\right)\right) \tag{18}$$

Therefore, the SM term and supersymmetry terms cancel each other when $\lambda_S = \left|\lambda_f\right|^2$. However, a question remains here. If the masses of sparticles are so low, then why have not we observed any supersymmetric particle so far? The only answer to this question is to assume that sparticles are heavier than the SM particles and we have not probed higher energies to find sparticles yet. Therefore, a symmetry breaking has to happen here. Like in the electroweak theory, we want to keep the Lagrangian invariant under the supersymmetric transformation and have the vacuum state break the symmetry and at the same time keep $\lambda_S = |\lambda_f|^2$. This symmetric breaking is called "soft symmetry breaking" [34]

1.2.1.2 R-parity

R-Parity is defined as:

$$R = (-1)^{3B+L+2S} \tag{19}$$

Where B is the baryon number, L the lepton number, and S the spin. All the SM particles have R = +1, and all the supersymmetric particles have R=-1. Allowing for the R-parity violation has severe consequences. For example, the proton can decay via squark exchange, like $p \rightarrow e^+\pi^0$. Assuming R-parity conservation has a few consequences:

- Both the baryon and lepton numbers are conserved
- sparticles can only be produced in pairs
- sparticles can only decay into an odd number of other sparticles
- The lightest supersymmetric particle (LSP) is stable

1.2.1.4 Theoretical motivations

Naturalness

Naturalness is the property of a theory that the dimensionless ratios between free parameters or physical constants appearing in a physical theory should take values "of order 1" and that free parameters are not fine-tuned.[40] As explained before, Supersymmetry can solve the hierarchy problem and bring naturalness to the Higgs mass calculation.

Gauge coupling unification

Equation 4 shows the evolution of the strong coupling constant α_S as a function of Q^2. If we plot the gauge coupling constants (for Strong, weak and electromagnetic forces), there is no scale Q^2 found for which all three coupling constants intersect in SM. However, in MSSM and at $M_{GUT} = 10^{15} - 10^{16}$, all three coupling constants meet allowing for the unification of the forces. Such a theory is called Grand Unification Theory (GUT).

Dark matter

If Supersymmetry is a true theory and the nature preserves R-parity, then the lightest superparticle (LSP) falls into the category of cold dark matter (CDM) candidates. In other words, R-parity conservation makes LSP stable and a Weakly Interacting Massive Particle (WIMP) which are the requirements for a dark matter candidate.

1.2.1.5 Phenomenological MSSM (pMSSM)

The unconstrained MSSM has more than 100 parameters in addition to the Standard Model parameters. This makes any phenomenological analysis (e.g., finding regions in parameter space consistent with observed data) impractical. Under the following three assumptions:

- No new source of CP-violation
- No Flavor Changing Neutral Currents

- First- and second-generation universality

one can reduce the number of additional parameters to the following 19 quantities of the phenomenological MSSM (pMSSM)

1.2.1.6 Deficiencies in the MSSM

The μ-problem

The supersymmetric invariant Lagrangian is constructed from functions of superfields [41]:

- The Kahler potential which is a real function of the superfields
- The superpotential which is a holomorphic (analytic) function of the superfields
- The gauge kinetic function which appears in supersymmetric gauge theories

The superpotential term with conserved R-parity is:

$$W^{MSSM} = \bar{u} y_u Q H_u - \bar{d} y_d Q H_d - \bar{e} y_e L H_d + \bar{\mu} H_u H_d \tag{20}$$

Where $Q = (u_L, d_L)$; $L = (v, e_L)$ and y_u, y_d and y_l are the dimensionless Yukawa coupling parameters. [42]

The μ parameter in eq. 20 has the mass dimension and there is no constraint on it in MSSM. However, it should have a value close to the electroweak or SUSY breaking scale to get the correct pattern of electroweak symmetry breaking[43]. Fine-tuning its magnitude to that of the electroweak scale seems arbitrary and without theoretical support. [44]

Little hierarchy problem

Another problem in the MSSM is that the mass of the stop (the supersymmetric partner of the top quark) must be quite large - on the order of 1-10 TeV - in order for the lightest CP-even Higgs boson to have a mass greater than 115 GeV without large stop mixing. However, such large stop masses would produce large loop corrections to the Higgs mass and quartic coupling, thus reviving the need for fine-tuning. This is referred to as the “little hierarchy problem”.

Flavor universality

Since no flavor mixing in addition to that predicted by the standard model has been discovered so far, the coefficients of the additional terms in the MSSM Lagrangian must be, at least approximately, flavor invariant (i.e., the same for all flavors). This implies a fine-tuning in NMSSM.

The smallness of the CP-violating phase

Since no CP violation additional to that predicted by the standard model has been discovered so far, the additional terms in the MSSM Lagrangian must be, at least approximately, CP invariant, so that their CP violating phases are small. This implies a fine-tuning in NMSSM.

1.2.2 Next-to-Minimal Supersymmetric Standard Model

The NMSSM is MSSM with an extended Higgs sector that has an additional complex Higgs singlet, S. Therefore, the sfermion and chargino sectors of MSSM are unchanged and only the Higgs and Neutralino sectors are different in NMSSM. [42] In other words, the superpotential of the NMSSM has some extra terms compared to the MSSM one. Let's start by fixing the μ-problem in MSSM and have the NMSSM superpotential as:

$$W^{NMSSM} = W^{Yuk} + \lambda S H_u H_d \qquad (21)$$

Where W^{Yuk} gives the Yukawa couplings for the Standard Model fermions [45]. Here, we have replaced the Higgs-Higgsino term in MSSM superpotential ($\bar{\mu} H_u H_d$) by $\lambda S H_u H_d$. In other words, we are dynamically linking μ to the VEV of a Higgs filed S (which is non-zero due to spontaneous symmetry breaking) and consequently to the electroweak scale. The parameter λ is to recover the Higgs-Higgsino mass term where the $VEV\ of\ \langle S \rangle = \mu/\lambda$. [43]

Looking more carefully at (eq.21), we observe that the superpotential is invariant under the U(1) PQ global symmetry[46]:

$$\hat{\phi}_i \to e^{i q_{PQ} \theta} \hat{\phi}_i \qquad (22)$$

This symmetry carries over to the Lagrange density as well.[7] Since we want S to gain a non-zero VEV, we have to spontaneously break the symmetry and end up having a near massless pseudo-Nambu-Goldstone boson (the so-called Peccei–Quinn mode). However, there has been no observation of an axion in collider experiments knowing its rare decays (e.g. quarkonium decays).[47] This puts strict experimental bounds on the existence of the axion in models where the VEV of S (PQ scale) is of the order of the electroweak scale

---

7- In MSSM, the PQ symmetry is broken by the Higgs-Higgsino mass term ($\bar{\mu} H_u H_d$) 43. Miller, D. and R.J.a.p.h.-p. Nevzorov, *The Peccei-Quinn axion in the next-to-minimal supersymmetric standard model.* 2003.

[48, 49] Adding another term to the superpotential explicitly breaks the PQ symmetry, giving mass to the axion and avoids the constraints.

$$W^{NMSSM} = W^{Yuk} + \lambda S H_u H_d + \frac{1}{3}\kappa S^3 \tag{23}$$

Where κ the PQ-symmetry breaking parameter. We chose the cubic form of S to keep κ dimentionless (the superpotential has the mass dimension of three). Any term higher than trilinear in the fields in superpotential will disrupt the requirements of renormalization. [46]

In summary, by adding S to the theory, μ changes to:

$$\mu_{eff} = \lambda \nu_s \tag{24}$$

Where ν_s is the (non-zero) vacuum expectation value of the singlet. Therefore, μ_{eff} is of the order of the electroweak breaking scale and the μ-problem is avoided in the Next-to-Minimal Supersymmetric Standard Model (NMSSM) [42][17]. The introduction of the scalar singlet Higgs field S results in totally seven Higgs bosons in NMSSM[8]:

- A pair of charged Higgs bosons ($h^{\pm}$)
- Three CP-even neutral scalars (h_1, h_2, h_3 with $m_{h_1} \leq m_{h_2} \leq m_{h_3}$)
- Two CP-odd neutral pseudoscalars (a_1, a_2 with $m_{a_1} \leq m_{a_2}$)

In general, not only NMSSM keeps the fundamental motivations of MSSM (namely, the hierarchy problem, gauge coupling unification, dark matter and so on), it also adds more motivation:

---

[8] 8- In MSSM, the capital letters are used to denote Higgs bosons, whereas in NMSSM, small letters are used.

- It solves the μ-problem in MSSM naturally as explained above.
- The singlet field also contributes to the Higgs mass formulae of MSSM such that the tree-level mass of the lightest CP-even Higgs can be raised. This can significantly reduce the little hierarchy problem in MSSM since the tree-level Higgs mass can be larger than the maximum value of MSSM. [42]
- In NMSSM, heavier Higgs can decay into lighter ones if the mass of the heavier Higgs is at least twice the mass of the lighter Higgs. It means that the current energy scale of LHC is enough to discover this phenomenon in NMSSM by observing the Higgs-like particle that was discovered in 2012 decaying to lighter Higgs if kinematically allowed; Unlike MSSM that might need higher energy scales to be discovered
- There can be a singlino-like neutralino dark matter in NMSSM with exceptionally weak couplings to SM particles. Therefore, NMSSM brings more possibilities for DM. [50]
- NMSSM can realize electroweak baryogenesis which can explain the matter-antimatter asymmetry in the universe. Baryogenesis is the hypothetical physical process that could have happened during the early universe to produce baryonic asymmetry. [51, 52]

Chapter 2 Experimental Apparatus

The analysis presented in this dissertation uses the data collected by the Compact Muon Solenoid (CMS) detector at Large Hadron Collider (LHC) at CERN, Geneva, Switzerland. Like any other experimental research, understanding the experimental setup is crucial. Therefore, in this chapter, we will briefly introduce the LHC complex and the CMS detector.

2.1 The Large Hadron Collider

To study particles, first, we have to create them. Einstein's famous equation $E = mc^2$ shows that mass and energy are equivalent. Therefore, after accelerating particles and giving kinetic energy to them, this energy can be released in a collision and convert into mass and create more massive particles. This is what happens at the LHC. The LHC is the largest and most powerful accelerator ever built and has achieved the current highest energy of 5.02 TeV for heavy-ion collisions and 13 TeV for proton collisions. [53]

The LHC is a 27-kilometer (17 miles) ring-shaped tunnel, buried underground below the border of France and Switzerland. The protons are inside 1232 dipole superconducting magnets[9] that produce a magnetic field of up to 8.3 to steer the protons[10] around the ring [53] (Figure 2-2). Inside the magnets are two pipes, through which protons beams circulate

---

9- There are higher order magnets at LHC as well. The total number of superconducting magnets are 9593 which 1.5 tons of superfluid helium-4 keep them at very low temperature of 1.9 K (-271.3 °C) 54. Evans, L.R., *The Large Hadron Collider: a marvel of technology*. 2009: EPFL Press.

10- In general, having particle-antiparticle (for example proton-antiproton) collisions is a more convenient choice (Like HEP at CERN and Tevatron at Fermilab). However, the production of antiprotons for such a high luminosity that LHC has achieved is currently not possible.

in opposite directions and finally collide at four interaction points. Around the four interaction points are four detectors: CMS and ATLAS ([55, 56]) are two general-purpose detectors that are designed to support a broad range of particle physics research. And two other detectors are called ALICE (dedicated to heavy ion collisions) [57], and LHCb (dedicated to in B-quark physics). [58] In recent years, three minor detectors have [59]with LHCb; and LHCf[11], which shares the interaction point with the ATLAS detector. We will only explain the CMS detector since we just used the data taken by it in this dissertation.

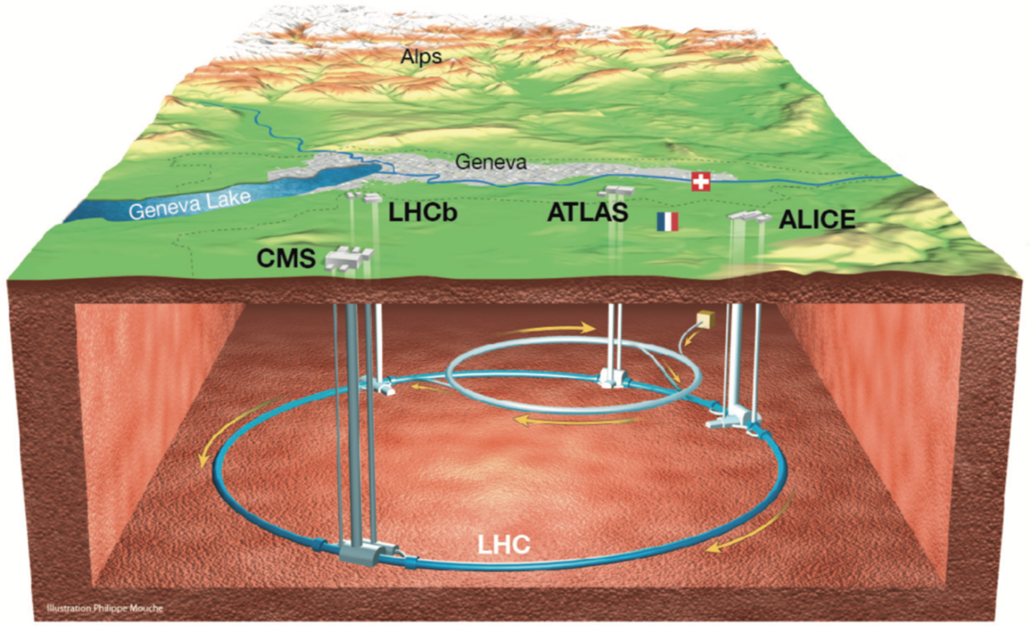


Figure 2-1: The Large Hadron Collider near Geneva, Switzerland. Depth is not to scale. This image is taken from [61]

11- Stands for "Large Hadron Collider forward" and it is "dedicated to the measurement of neutral particles emitted in the very forward region of LHC collisions. The physics goal is to provide data for calibrating the hadron interaction models that are used in the study of Extremely High-Energy Cosmic-Rays" 60. Adriani, O., et al., *The LHCf detector at the CERN large hadron collider*. 2008. **3**(08): p. S08006.

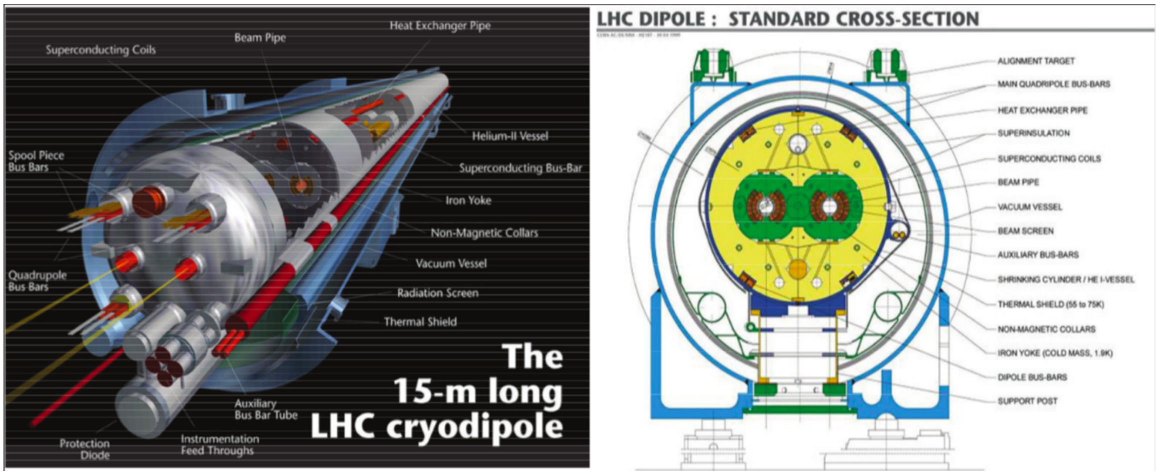


Figure 2-2: Schematic design of a dipole magnet at LHC. The image is taken from [62]

The proton beams at LHC are not continuous but grouped in bunches. The LHC can circulate up to 2808 bunches of protons in each direction, and each bunch contains up to ~120 billion protons. [63] When two bunches cross at every 25 ns, more than a pair of protons can collide. The number of pairs of protons is called "pile-up" and every proton-proton collision is called an event. [62] Luminosity is "the number of interactions per second per cm^{-2} of the beam areas" [56] The number of occurrences, N, of a given process observed is given by:

$$N = \mathcal{L}_{int}.\sigma \tag{25}$$

Where $\mathcal{L}_{int}$ is the integrated luminosity over some time (for example one year) with the unit of $\frac{1}{area}$ and σ is called the cross-section and has the unit of area. For example, a process with a cross-section of 100 fb^{-1} may be expected to occur 1000 times in 10 fb^{-1} of data[12]. This measured cross-section can be compared to the predictions of the adopted theory. [64]

Figure 2-3: The LHC tunnel, with cutaway view of the dipole magnet. This image is taken from [65]

In June 2014, the CERN Council approved the funding of the High Luminosity LHC (HL-HLC) to extend the discovery potential of the LHC. Figure 2-4 shows the energy and luminosity history and plan for the future of the LHC.

12- $1\ \mathrm{fb}^{-1} = 10^{-39}\ \mathrm{cm}^2$ 24. Baggott, J., *Higgs: the invention and discovery of the'God Particle'*. 2012: Oxford University Press.

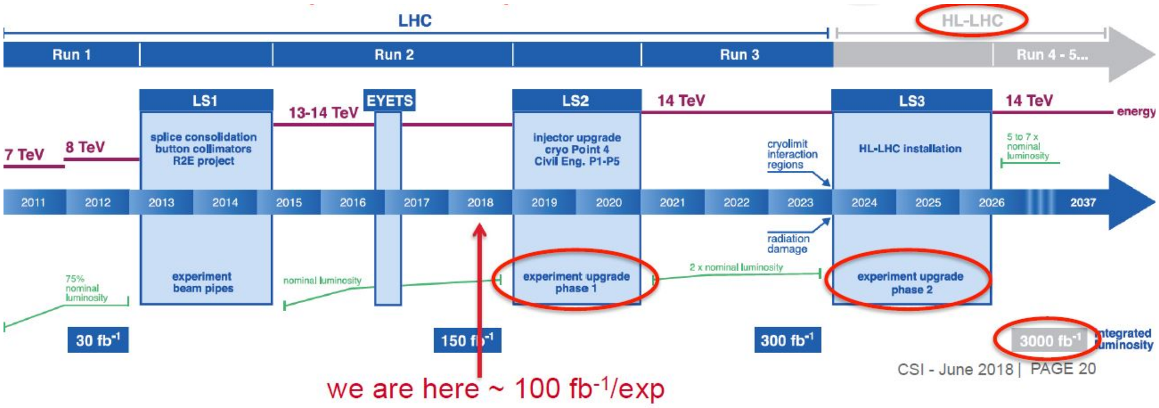


Figure 2-4: Evolution of LHC toward HL-LHC. This image is taken from [66]

2.2 The Compact Muon Solenoid Detector

The Compact Muon Solenoid is one of the two general-purpose detectors at the LHC. It has a multilayered array of sub-detectors like an onion. Each sub-detector is designed to detect a different kind of particles emerging from the interaction point. [67] The CMS consists of silicon pixel and strip trackers, an electromagnetic calorimeter, a hadronic calorimeter, a solenoidal magnet, and a muon system. We will briefly explain every layer in this section. Detector geometry and sub-detector layering are shown in Figure 2-5 and Figure 2-6 respectively. CMS is 15 meters wide and 28.7 meters long and weighs 14,000 tons. [68]

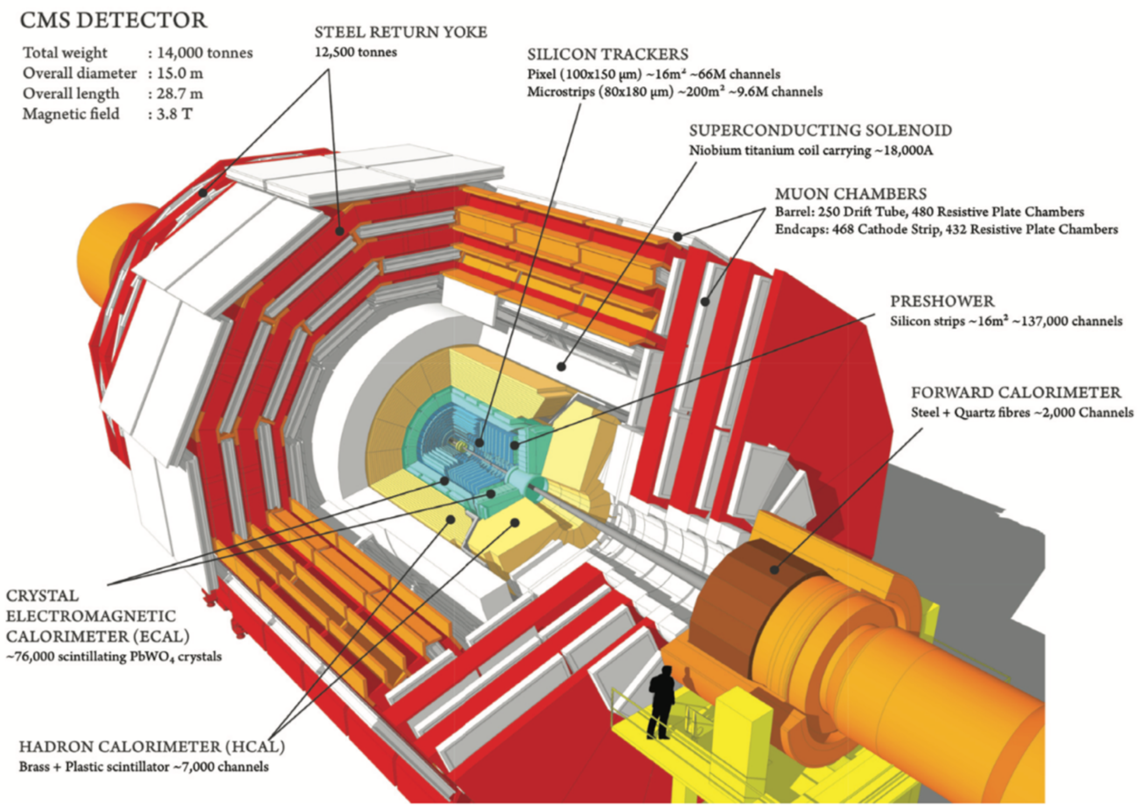


Figure 2-5: Geometry of the CMS detector

The CMS detector is divided into the barrel region which is the cylindrical body and the endcap region which is either of the two plugs that cap off the barrel on both sides. There is also an HCAL front region which is on the very front and very back of the CMS close to the proton beam.

2.2.1 Coordinate system

The origin of the coordinate system of CMS is defined to be at the interaction point. The x-axis points radially inward towards the center of LHC, the y-axis points vertically upward, and the z-axis points in the direction of the beam towards the Jura mountains. [69] Also, the radial angle is defined as:

$$r = \sqrt{x^2 + y^2} \tag{26}$$

And the azimuthal angle ϕ is measured from the z-axis.

Because collisions take place very close to the center of the CMS, it is convenient to use the polar coordinate system that gives the particle trajectories in angles. However, using the polar angle θ from the spherical coordinates is not convenient because we will need to boost between different reference frames in relativistic energies. This is why in collider physics; the rapidity is preferred:

$$y = \frac{1}{2} ln \frac{E + p_z}{E - p_z} \tag{27}$$

The rapidity difference (Δy) between two particles remains the same in different reference frames, i.e., it is Lorentz invariant. However, to measure rapidity, we need to measure two quantities (Energy and p_z). Therefore, physicists prefer a related quantity that is easier to measure, pseudorapidity. The pseudorapidity of a moving particle is

$$\eta = -\ln\left[tan\left(\frac{\theta}{2}\right)\right] \tag{28}$$

Which depends only on the polar angle of the particle. In fact, at relativistic velocities pseudorapidity converges to rapidity. Δϕ is also Lorentz invariant because ϕ is measured transverse to the z-axis. Therefore, particle physicists define

$$(\Delta R)^2 = (\Delta\eta)^2 + (\Delta\phi)^2 \tag{29}$$

to define cones around single objects or to measure the separation between two objects in a relativistically invariant way.

2.2.2 Superconducting Solenoid

‘S’ in CMS stands for superconducting solenoid which is a remarkable device in CMS in the form of a giant coil 12.9 m long and 6.3 m in diameter [70]. It is the most powerful superconducting magnet ever built capable of producing a longitudinal magnetic field up to 4 T; however, it is run at 3.8T to last longer. The strength of the magnetic field is chosen high enough that the CMS detector can determine the charge of a muon carrying 1 TeV of momentum [70]. Since it is a superconductor, liquid helium is used to cool it down to ~4 Kelvin [71].

This magnet is used to bend the charged particles emerging from the collisions to measure the transverse momentum[13] as well as the position of the interaction vertex that the particle has originated. Interaction vertices are essential in identifying b jets, the so-called b-tagging, which will be discussed further in the next chapter.

13- The magnetic field is in the z-direction; therefore, the tracker can only measure the momenta of the charged tracks in the x-y plane (transverse momentum) not in the z-direction (p_z).

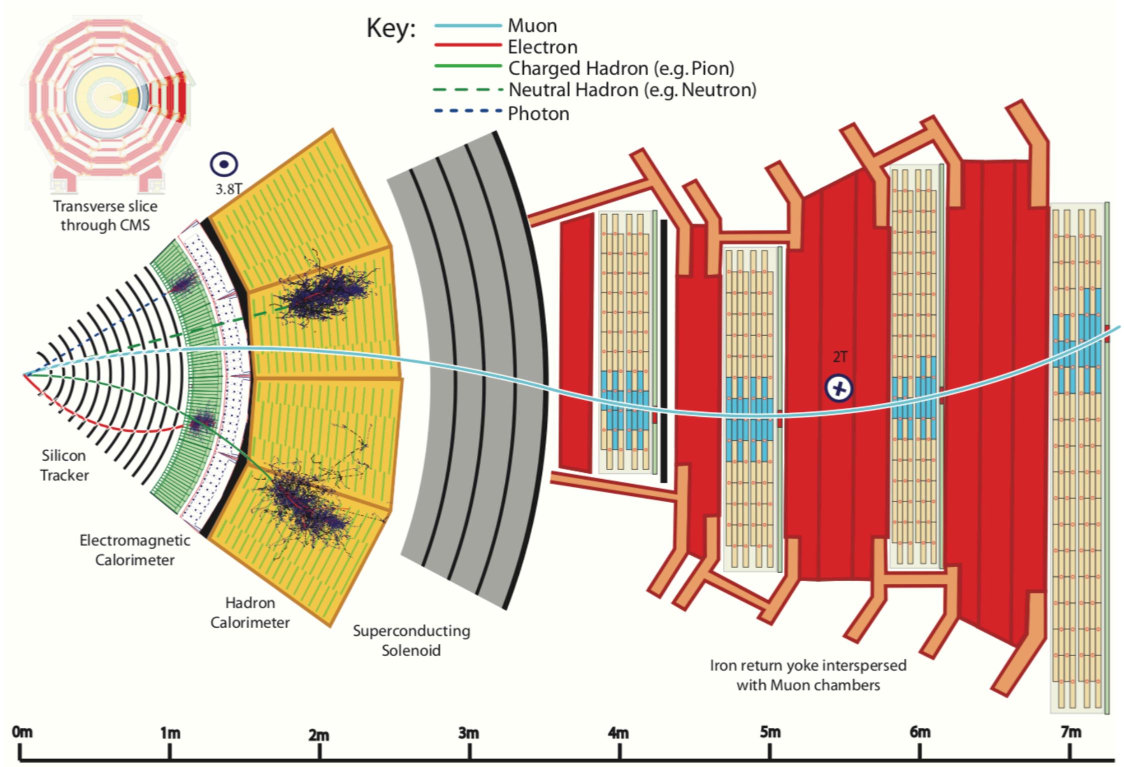


Figure 2-6: Diagram showing particles interacting with a typical cross-sectional wedge of the CMS detector [67]

2.2.3 Tracking system

The tracking system measures the trajectories or “tracks” of high p_T (> 1 GeV) charged particles which are used to [72]:

- measure particle transverse momenta from their curvatures in the magnetic field
- infer the location of interaction vertices (where multiple tracks originate)

The interaction of charged particles with silicon detectors creates a blip of electrical current at each layer that is read out electronically. Then, the path is reconstructed by “connecting the dots”. It is important to mention that neutral particles do not bend in the magnetic field and do not leave any hits in the tracker, therefore, their trajectories and momenta cannot be measured by the tracker.

There are a few requirements that the trigger has to meet. Some challenges of the tracking system are [73]:

- There are tens of collisions that produce thousands of new particles every 25 ns. This high flux of particles requires:
    - A very high-resolution trigger to be able to distinguish between the thousands of particles simultaneously going through
    - Radiation-hard to operate in a high radiation environment for a few years
    - Fast response as there are thousands of new particles every 25 ns. This necessitates on-detector electronics which in turn needs proper cooling
- The minimum amount of matter should be used to disturb the particle as little as possible.

These requirements lead to the development of an all-silicon detector. Closer to the interaction point, there is a higher particle density, and therefore, higher resolution is needed. Thus, pixel detectors were chosen to be used (Each pixel is $100 \times 150\ \mu m$ in size [70]). The inner tracker is wrapped immediately around the beam pipe [72]. In the past, there were three layers in the barrel and two in the endcaps [70]. However, the pixel tracker was upgraded in the 2016/2017 winter shutdown, and a third pixel layer was added in the endcaps and a fourth in the barrel, Also, the hardware that supports the functioning of the pixels was slimmed down [74].

The pixel detector constrains the position of interaction vertices precisely. Therefore, farther than the pixel detector, the tracker is only used to continue measuring the trajectory of particles. Since we just measure the momentum in the x-y plane and not in the z-direction, accuracy is very important only in the radial direction leading to microstrip detectors that are much longer in one direction compared to the other two. Overall, the tracker subsystem is 5.8 m long and 2.5 m in diameter making it the largest silicon tracker ever built, with $\sim 200\ m^2$ of total active silicon covering up to the pseudorapidity of $|\eta| < 2.5$ [72]. A refrigerant liquid is circulated in the tracker to maintains the temperature lower than $-10\ °C$ [71].

Figure 2.5 shows the geometry of this tracker:

- Tracker Inner Barrel (TIB) covers 20 < r < 55 cm in Barrel region with four layers of silicon microstrips, with a size of at least 10 cm × 80 μm.
- Tracker Outer Barrel (TOB) defined by r > 55 cm in Barrel region with six layers of microstrips with a size of no more than 25 cm × 180 μm.
- Tracker Inner Discs (TID) sits in the Endcap regions just outside the TIB and consist of three layers of discs of strips.
- The Tracker Endcaps (TEC) is in the Endcap regions just outside the TOB and consists of nine discs of silicon strips.

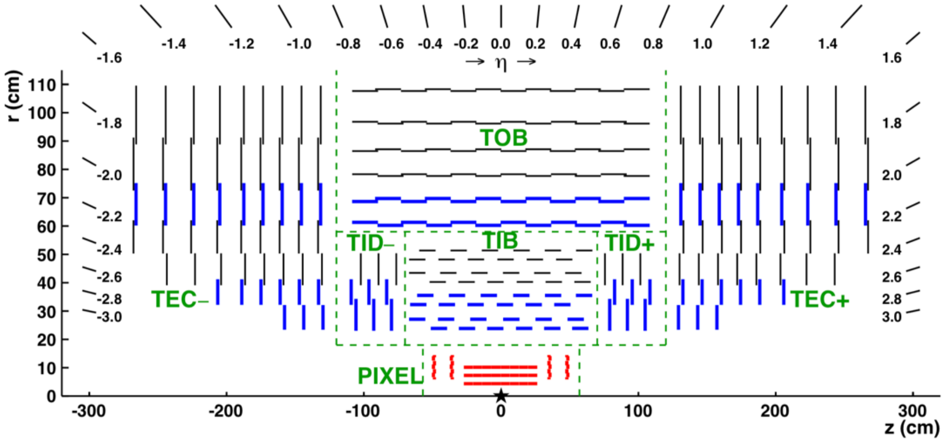


Figure 2-7: Schematic overview of a one-quarter (r, z) cross-section of the CMS tracker. This image is taken from [75].

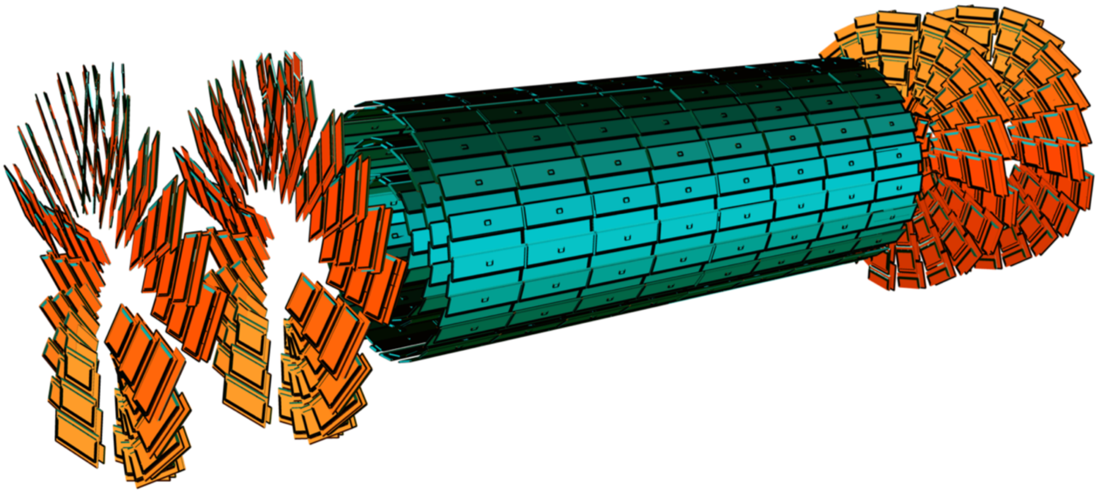


Figure 2-8: Schematic overview of barrel (green) and forward pixels (orange). This image is taken from [47]

2.2.4 Electromagnetic calorimeter

After the tracker, there are two calorimeters: the electromagnetic calorimeter and Hadronic calorimeter. The main idea in the calorimeters is to make the primary particle interact and create secondary particles. Then measure the energy deposition of all of the subsequent particles as they go through an active material.

The first calorimeter right after the tracker is called the electromagnetic calorimeter or ECAL. It measures the energy of the electromagnetic particles (electrons and photons) emerged from the collisions. High energy electrons and positrons interact with matter mostly through the bremsstrahlung radiation creating photons which in turn can create an electron-positron pair which in turn can create photons and so on. In the end, showers with thousands of electrons, positrons, and photons are created. These electromagnetic showers are relatively well-behaved which results in phenomenal resolution in the ECAL[14]. [65]

To detect electromagnetic particles, ECAL is made of almost 80,000 lead tungstate ($PbWO_4$) crystals. Lead tungstate has some desirable properties [77]:

- It is optically transparent. This helps photons travel inside the crystal
- Fast scintillation time: lead tungstate crystals release ~80% of their scintillation light in the 25 ns window between bunch crossings
- Lead tungstate is highly radiation-hard: tolerating an absorbed dose of up to 10 Mrad (100 kGy)

---

[14] This high resolution is very important in Higgs boson discovery in the diphoton final state. 76. Heller, R.E., *A search for supersymmetry with the CMS detector in the single-lepton final state using the sum of masses of large-radius jets*. 2018, UC, Santa Barbara.

- $PbWO_4$ is very dense ($8.28\ g/cm^3$) which results in short radiation length ($X_0 = 0.89\ cm$) and Moliere radius ($R_M = 2.2\ cm$).

The radiation length is defined as the mean free path for emitting Bremsstrahlung radiation [78]. And Moliere radius is the radius of a cylinder containing on average 90% of the shower's energy deposition [75]. In other words, it can be thought of the transverse distance that a particle at the critical energy goes before it dies off [78]. It is 7/9 of the mean free path for pair production by photons. The ECAL should be many radiation lengths deep to minimize the leakage of energy from the back of the calorimeter. Low Moliere radius of $PbWO_4$ makes the ECAL more compact.

On the other hand, the main disadvantage of $PbWO_4$ is its poor light yield. In other words, $PbWO_4$ only produces $O(100)$ photons per MeV of energy deposition which is two orders of magnitude lower than LYSO or NaI crystals. However, this is not a big problem at LHC high energy scale. All these properties make $PbWO_4$ a very good active material for the ECAL.

The ECAL (Figure 2-9) is made up of the barrel and endcap components which together cover the pseudorapidity of $|\eta| < 3$[15] [70]. When an electromagnetic particle goes through the lead tungsten crystal, it produces a blue-green scintillation light that is collected and converted to electrons through the photoelectric effect to determine the energy of the incident EM particles. In the barrel, this light is collected and read by silicon avalanche photodiodes (APDs) placed at the back of each crystal while in the endcaps, vacuum

---

15- Actually, there is a small gap in between the barrel and endcap that allows the electromagnetic particles that fall into this "crack" to escape undetected by ECAL phototriodes (VPTs) are used instead. VPTs are photomultipliers with a single gain stage which are used in the endcap because the magnetic field is non-uniform and the dose of radiation is higher [72]. Then the signal is amplified, digitalized and sent to computing equipment outside the detector volume [70].

The $PbWO_4$ crystals are like tapered rectangular prisms. In the barrel region, they are 230 mm (25.8 X_0) long with front (rear) faces of $22 \times 22\ mm^2$ ($26 \times 26\ mm^2$) area. And in the endcaps, the crystals are a bit larger with 220 mm of length (24.7 X_0) and $28.62 \times 28.6\ 2\ mm^2$ ($30 \times 30\ mm^2$) front (rear) face areas. After combining crystals to 3x3 or 5x5 grids (supercrystals), the fluctuations in the lateral leakage result in resolution degradation of about 2% at 1 GeV which decreases with $\sqrt{E}$ [79]. The front faces of the barrel crystals are 1.29 m from the beamline. And the endcap crystals are located at $|z| = 315.4\ cm$ from the interaction point [72]. Each endcap is divided into two half-circle structures called “Dees”. Each dee holds 138 supercrystals (groupings of 5×5 crystals) and 18 partial supercrystals.

Finally, a preshower detector is installed in front of each ECAL endcap whose main job is to distinguish boosted π_0 decays from isolated photons. π_0 that are produced in the p-p collisions decay to two photons which can mimic an isolated high-energy photon if they are closely-spaced and the ECAL picks them up together.[16] This problem mostly happens in the endcap region where the angle between two photons of the π_0 decay is likely to be small enough to cause this problem. The preshower has a much finer granularity compared

---

16- This is especially important in Higgs Boson searches. Since one of the Higgs decays is to photons.

to the ECAL and can see each of the pion-produced photons separately. In total, there are 75,848 crystals in the ECAL which provide very fine spatial granularity.

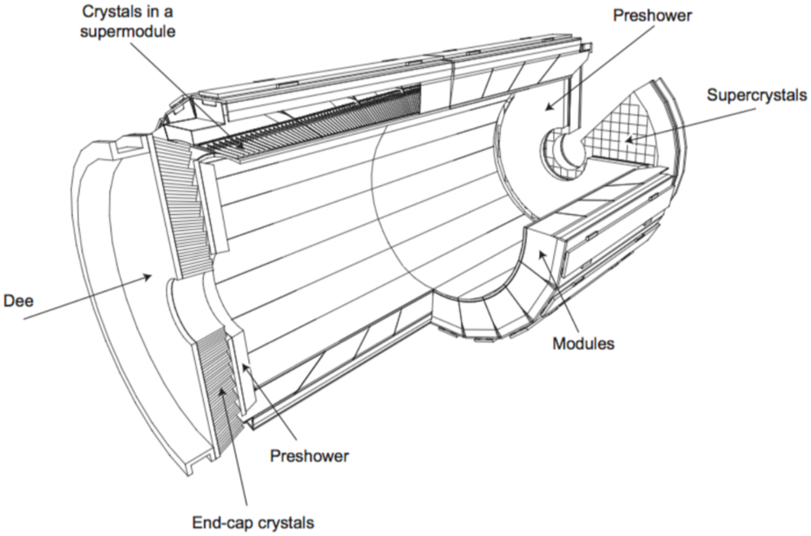


Figure 2-9: One-quarter (r, z) view of the layout of the CMS ECAL. This image is taken from [70].

2.2.5 Hadron Calorimeter

The second calorimeter is the hadron calorimeter, or HCAL. HCAL is designed to measure the energy of hadronic jets emerged from the collisions. It also contributes to measuring the pileup from secondary collisions in the event.

Hadronic showers are inherently more complex than electromagnetic showers, and therefore the resolution of hadronic calorimeters is much worse than electromagnetic calorimeters. For instance, the energy deposit by hadrons in nuclear breakup or excitation is around 30% to 50%, which are invisible or produce signals that are $O(\mu s)$ out of time [65, 80]. We can estimate the invisible energy from visible energy, however, fluctuations in this ratio deteriorate resolution. Besides, part of the hadronic jets. Therefore, every hadronic shower has a large electromagnetic component, containing one-third of the energy at low energies and rising logarithmically with energy. Since the response is very different between the two components, fluctuations in the electromagnetic fraction dramatically broaden the resolution as well.

The CMS HCAL is a sampling calorimeter, meaning that it uses alternating layers of “absorber” and fluorescent “scintillator” approximately 50 mm and 3.8 mm thick, respectively. The absorber layers, most of which are made of brass (70% Cu, 30% Zn), induce the showering. Brass[17] is relatively affordable and also does not perturb the magnetic field inside the magnet because it is non-magnetic. This brass alloy has an interaction length of λ_I = 16.42 cm and a density of $8.53\ \mathrm{g/cm^3}$. The scintillating layers interleaved with the absorber layers are made of radiation-hard plastic, mostly Kuraray SCSN81 and Bicron BC408; however, quartz fiber is also used for its superior radiation hardness. Using the energy deposited in the scintillators, we can extrapolate the invisible energy and the energy deposited in the brass layers.

Figure 2-10 shows the geometry of the HCAL. The HCAL is divided into four subcomponents:

- The Barrel (HB): The barrel covers the pseudorapidity range of $|\eta| < 1.3$ and is segmented into 36 wedges where each wedge encompasses half of the HB in η and 20° in φ corresponding to exactly one ECAL supermodule. There are 16 layers of

---

17- It is interesting to know that this brass comes from a million WWII brass shell casements from the Russian Navy absorber layers in each wedge of the Barrel region. The innermost and outermost absorber layers are made of stainless steel with 40 mm and 75 mm thickness, respectively, which also provide structural support. And the middle fourteen layers are brass interlayered with scintillators. The first eight are 50.5 mm thick, and the last six are 56.5 mm thick. The scintillators have a granularity of $(\Delta\eta, \Delta\varphi) = (0.087, 0.087)$. Every single unit of the scintillator is called a tile; the CMS HCAL detector contains about 70,000 tiles. The innermost scintillator layer in the HB is made of Bicron BC408 while the remaining 16 layers are made of Kuraray SCSN81 [71].

- The Endcaps (HE): The endcaps are two discs on positive and negative z-directions that together cover the rage of $1.3 < |\eta| < 3.0$. Each disc has 36 wedges and 18 layers of absorber and scintillator that are made of the same material as HB. However, the difference with HB is that the stainless steel is only used for the outer absorber layer, to prevent any magnetic interference. The scintillator tiles in HE are trapezoidal and provide a granularity of $(\Delta\eta, \Delta\varphi) = (0.087,0.087)$ for $|\eta| < 1.6$ (Same as HB), and approximately $(\Delta\eta, \Delta\varphi) = (0.17,0.17)$ for $|\eta| \geq 1.6$. Layers 1–17 are 3.7 mm thick, and layer 0 is 9 mm thick [71].
- The outer calorimeter (HO): The space that HCAL can take is very constrained because we want HCAL to be inside the magnet to measure hadronic showers before they interact with the magnet. This space constraint makes it very difficult to achieve enough stopping power to measure the energy of high-energy and late-developing jets reliably. Therefore, the outer calorimeter is designed to provide a

better measurement for the late-developing jets. It covers the region $|\eta| < 1.3$ (same as HB) and consists of tiles of Bicron BC408 scintillator (10 mm thick) mounted inside the iron yoke that gathers the returning magnetic field outside the solenoid. This means that the HO uses the solenoid material as the absorber. Due to the shape of the return yoke, the HO also is divided into five rings along the z-axis, with each having 12 sectors in ϕ. There are gaps between the rings because of the power lines and cryogenic that supply the magnet. The scintillator tiles roughly follow the $(\Delta\eta, \Delta\varphi) = (0.087, 0.087)$ segmentation of the HB, within the constraints of the yoke geometry.

- The forward calorimeter (HF): The forward calorimeter is placed in the most forward region of the CMS ($3.0 < |\eta| < 5$) and is very important for missing transverse energy measurement since it is the only component above $|\eta| > 3$. It is also used to measure the instantaneous luminosity of the LHC [81]. As HF is close to the beamline, it experiences enormous radiation doses, near 100 Mrad/year [82]. Thus, HF must be considerably more radiation hard than any other component of the HCAL. As a result, HF uses quartz fibers with polymer cladding as its active material and detects Cherenkov light instead of scintillation light (if the particle energy is above Cherenkov threshold of $E \geq 190\ keV$). Each end of the HF is composed of cylindrical stainless-steel absorbers of length 165 cm, from r = 12.5 cm to r = 130 cm, arranged in 18 azimuthal wedges that are interspersed with quartz fibers that run parallel to the beamline. Although HF stands for Hadronic Calorimeter Forward, it also measures the energy of electromagnetic particles.

Some of the fibers penetrate halfway (short fibers) through the steel plate to allow the HF to distinguish between electromagnetic (which goes less deep) and hadronic (which goes deeper) showers. The fibers form towers of size $(\Delta\eta, \Delta\varphi) = (0.175, 0.175)$ [71].

However, quartz calorimetry comes with some challenges that makes the energy resolution of HF somewhat coarse:

- The energy loss due to Cherenkov radiation in quartz is about 10000 times smaller than that due to ionization. Thus, the number of photons is small.
- The light transfer is done by total internal reflection in the fiber at large angles. Therefore, only a few percents of the photons are transported to the photodetectors.
- Quartz only occupy around 1% of the HF volume (it is ~10% active volume in HBHE).
- The Cherenkov radiation is much more sensitive to the electromagnetic component than the hadronic component of the showers. Thus, changes in the shower composition have an even more effect on the resolution than in other calorimeters, resulting in poor pion energy resolution.

Despite the somewhat coarse resolution of HF, it provides vital for the physics analyses.

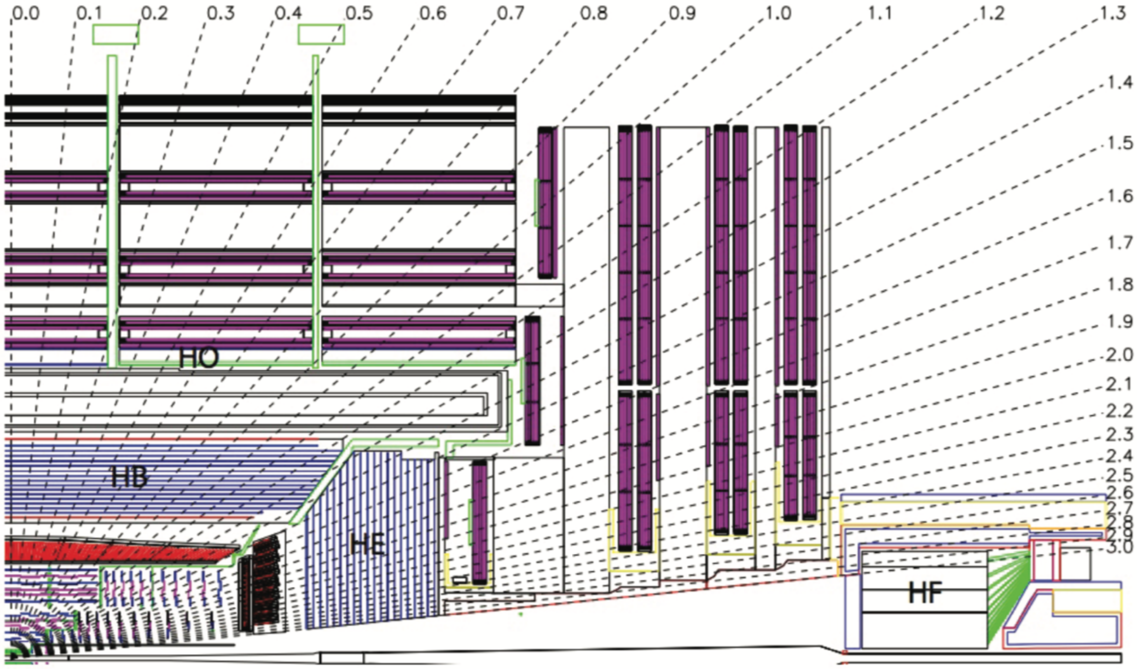


Figure 2-10: One-quarter (r,z) view showing the layout of the HCAL components within the CMS detector [72].

The light from the plastic scintillator tiles in the HB, HE, and HO is transferred by Kuraray Y-11 wavelength-shifting (WLS) fibers to the outside and then using clear optical fibers to hybrid photodiodes (HPD) for electronic processing.

2.2.6 Muon System

The muon system is the outermost component of the CMS detector. Muons are heavy and therefore have high penetrating power which allows them to reach beyond all other layers and even go through the magnet. The muon system detects muons as they fly away from the collision point and contains three different gas-and-electrode technologies to measure the trajectories of muons. These trajectories and tracker trajectory can be combined for greater precision.

A schematic view of the muon system is presented in Figure 2-11. Like most other subsystems, it is also divided into the barrel (cylindrical) and endcap (two discs) regions. Muons are detected using a combination of drift tubes (DTs) and resistive plate chambers (RPCs) in the barrel, whereas cathode strip chambers (CSC) and RPCs are used in the endcaps. All three of these types detect muons exploiting the trail of ionization that the muons leave by passing through a gas. The free electrons get attracted to positive electrodes and the ions to negative electrodes producing electric signals that are read out. Then using the time difference between signals, one can tell how far along the electrode the impact occurred.

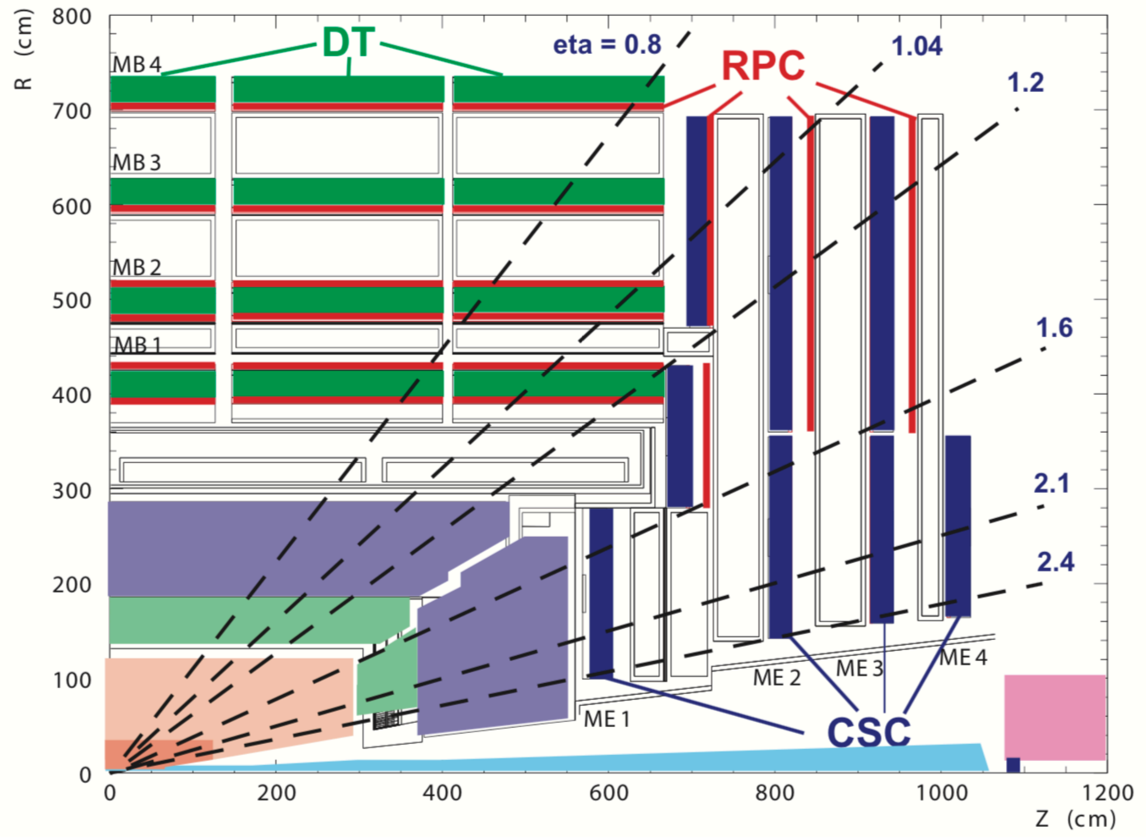


Figure 2-11: One-quarter (r, z) view of the layout of the CMS muon system [70].

Figure 2-12 shows a basic drift cell comprising the drift tubes (DT). DT is an approximately rectangular tube filled with a mixture of 85% Ar and 15% CO_2 gases, $42mm \times 13mm \times 2.4m$ with a 50μm positive wire running down its center. The electric field is ~2 kV/cm witch causes the electron drift velocity of 60 $\frac{\mu m}{ns}$ and the drift time of maximum 380 ns. DTs cannot measure the coordinate parallel to the wire and therefore, they provide 1D position measurements with a resolution of ~250 μm. 2D position measurement is possible by stacking multiple layers which are called a drift tube station. DTs are used because they are precise yet inexpensive and well-suited in the muon barrel, where the particle flux is low and magnetic field is both low and constant [71]. In the muon endcaps, both particle rate is higher and also the magnetic field is not uniform. That is why instead of DTs we use a faster detector called CSC (Cathode Strip Chambers).

The CSC are planar chambers filled with a mixture of 40% Ar, 50% CO_2, and 10% CF_4. They contain anode wires running in one direction, and cathode strips running perpendicularly, providing a 2D position measurement. CSCs are used in the endcap due to their precision, fastness (< 225 ns), and operation in the high magnetic field at the fringes of the superconducting solenoid [71].

Resistive plate chambers (RPC) are also planar chambers filled with a mix of 96.2% $C_2H_2F_4$, 3.5% iC_4H_{10}, and 0.3% SF_6. They have an anode plate on one face, and a cathode plate on the other face. In RPCs, electrons are detected by metal strips outside the chambers. RPCs are used to enhance the CSCs and DTs performance mainly due to their 1 ns timing resolution; although their spatial resolution is not as good as the DTs or the

CSCs [71]. The signals coming from these three subsystems are processed by electronics inside the detector volume as well as outside of it.

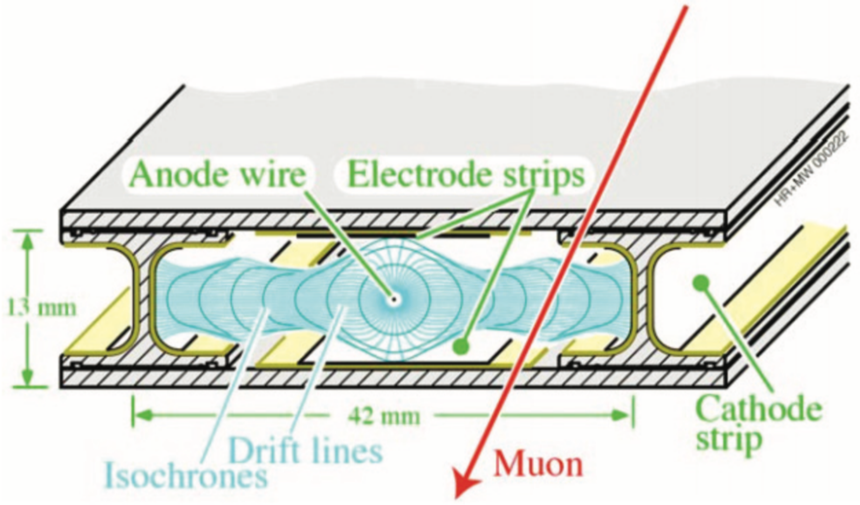


Figure 2-12: A schematic view of CMS muon drift chamber. This image is taken from [72].

2.2.7 Trigger System

During 2016 data taking, the peak instantaneous luminosity was $1.5 \times 10^{34}\ cm^{-2}s^{-1}$ and up to 52 collisions per bunch crossing (pileup) were recorded in 2016 [83]. It is not wise to keep so many events per 25 ns bunch crossing. Because on one hand it is not possible due to technology restrictions and on the other hand most of the events are not even interesting. Most interesting events are the ones that contain W or Z bosons, top quarks, or Higgs bosons. And these interesting events are 5 to 10 orders of magnitude rarer than proton-proton cross-section. Therefore, we need a mechanism to choose interesting events and only store them. This mechanism is called the trigger system and it has two levels jassem118:

- The Level-1 trigger (L1T) which applies a primitive selection that reduces the rate from ~40 MHz to ~100 kHz. The L1 trigger seeks signatures such as ionization tracks in the muon system, HCAL deposits consistent with jets, ECAL deposits consistent with electrons or photons, and so on
- The High-Level Trigger (HLT) which reduces the rate further to several hundred Hz. HLT receives events chosen by L1T and perform an analysis on its farm of thousands of CPUs that is close to offline reconstruction with some simplifications. After deciding which events contain physics processes of interest, those events are stored into different HLT paths.

Applying triggers can cause some difficulties that must be compensated during physics analysis:

- The trigger selections are not perfect, meaning that they can miss a portion of the events they are supposed to trigger on. For instance, a trigger that selects events containing muons with $p_T > 17$ GeV might only work 98% of the time and the efficiency may even depend on the muon p_T. This efficiency is measured and accounted for in the physics analyses.
- A trigger selection (e.g. a single electron) can occur so frequently that the triggers cannot handle it. To cope with this flood of desired events, we program the trigger software to keep only every n^{th} event (prescaling the trigger), and we weigh that event by n.

2.3 Reconstruction and Identification

What we have so far is just electronxcic outputs from the CMS detector. However, we need a meaningful picture of what particles and objects are present in the event. This process is called event reconstruction ("reco") and particle identification ("ID"). The CMS Collaboration uses a system called particle flow (PF) to do basic reconstruction [75].

The particle flow algorithm associates all of the detector readings with known particles. Its logic basically follows the logic of scientists who designed the CMS. It starts by reconstructing tracks. Tracks allow us to measure the momentum and charge of particles. If the particle goes all the way to the muon system, then it is a muon. PF identifies muon candidates using muon detector tracks, inner tracks, and calorimeter energy deposits. After PF reconstructs all muons, it removes all the corresponding detector signatures and starts with the next easiest identifications, and so on. Each successive ID step only uses the information that was not consumed by the previous step. Some postprocessing considerations are performed at the end [67].

2.3.1 Charged Tracks and Vertices

The PF algorithm starts by reconstructing tracks from the pixel and strip hits in the inner tracker. Track reconstruction is performed by a software named the Combinatorial Track Finder (CTF) based on the Kalman filter, which is described in References [84] and [85]. The main idea in Kalman filter is to connect hits in the successive layers of the tracker and fit them to a helical curve. The method puts adjustable constraints on the maximum number of missing hits, the minimum number of sequential hits, and also other quantities to ensure that tracks that are being reconstructed have a minimum level of quality.

Then it starts grouping all the tracks identified in the event into vertices. A vertex is a point in space where a collision or a decay took place and we observe some particles emitted from it. The method reconstructs vertices by extrapolating tracks backward in time to see where they intersect. Every bunch crossing can contain multiple proton-proton collisions as well as intermediate particle decays, and thus multiple vertices. The vertex that has the highest $\sum_{tracks} p_T^2$ is defined as the primary vertex, and all other vertices are secondary. Because the primary vertex has the highest energy, we mostly ignore secondary vertices when analyzing an event, although displaced secondary vertices are used in reconstructing for example b-jets.

2.3.2 Calorimeter Clusters

Before starting to reconstruct particles there is one more step to take and it is to identify clusters of energy deposits in the ECAL and HCAL. Clusters of energy can be used to differentiate between electrons and photons, and also between charged hadrons and neutral hadrons. We start from the calorimeter cell that its energy deposit is above some defined threshold (it is called the seed). From the seed, we recursively add any adjacent cell with sufficient energy forming a cluster, and then fitting a 2D Gaussian to the energy distribution. Due to different geometry and detector layout in different regions, separate calibrations are applied in the endcap and barrel of each calorimeter [67].

2.3.3 Muons

Now we have reconstructed the tracks, vertices, and clusters and it is the time to reconstruct the muons in the event. Among all particles, muons' signature is usually very clean and distinctive. Muons are reconstructed by lining up a track in the inner tracker with a track in the muon system. The muon that is reconstructed using both tracker and muon-system information is known as global muon. It is also possible to reconstruct tracker-only or standalone muons, which are reconstructed using the information solely from the inner tracker, or the muon system, respectively. Once a muon candidate is identified, its associated tracks are removed for future steps [67].

2.3.4 Electrons and Isolated Photons

In the next step, both electrons and isolated photons are identified since the reconstruction of both particles relies on ECAL energy deposits. An electron candidate is a charged track in the tracker that matches up with a comparable ECAL energy deposit while it is a photon candidate if there is an ECAL deposit with no corresponding track. To confirm the identification, a number of further criteria are checked. For example, the shape of the ECAL energy deposit is a little different for electrons and photons. The tracker bends the electrons which have negative charge and therefore, they produce bremsstrahlung (a shower of photons). These photons are still high-energy and tend to decay to electron-positron pairs, which in turn can produce further bremsstrahlung, and so on. This cascade which is emitted tangent to the curved trajectory of the electrons causes a long tail in the ϕ direction of their energy deposit in ECAL. To the contrary, photons do not bend and therefore create rounder energy deposit distributions in the ECAL [67]. Once an electron or photon candidate is identified, its associated signatures are removed for future steps.

2.3.5 Hadrons

We have already removed the isolated muon, electron, and isolated photon candidates. Therefore, what is left should be hadrons. However, we also want to reconstruct non-isolated photons and muons coming from hadron decays. Particle flow assigns any (non-isolated) ECAL deposit that is not linked to a track as a photon; similarly, any HCAL deposit not linked to any track is considered to be a neutral hadron. Finally, any remaining HCAL deposit has to be linked to a track and are therefore considered to be charged hadrons. When an ECAL deposit, an HCAL deposits, and/or a track overlap, particle flow reconstructs them as combinations of charged and neutral hadrons and photons [67].

Chapter 3 $H \rightarrow aa \rightarrow 4b$ Analysis

3.1 Introduction

In this dissertation, a search for an NMSSM physics signal in a four-bottom quark final state is presented. The signature process sought is the production of an SM-like Higgs boson H followed by its decay to a pair of lighter pseudo-scalar Higgs bosons a, each of which decays to a pair of bottom quarks. When there is substantial mass difference between H and a ($a < 30\ GeV$ [86]), the a's are produced with significant boost. Also, due to the high QCD background in the 4b final state, the cleanest channel for the SM Higgs production is the W associated production channel (Figure 3-1 - left) where a high-p_T isolated lepton from the vector boson decay provides a convenient trigger. [86]

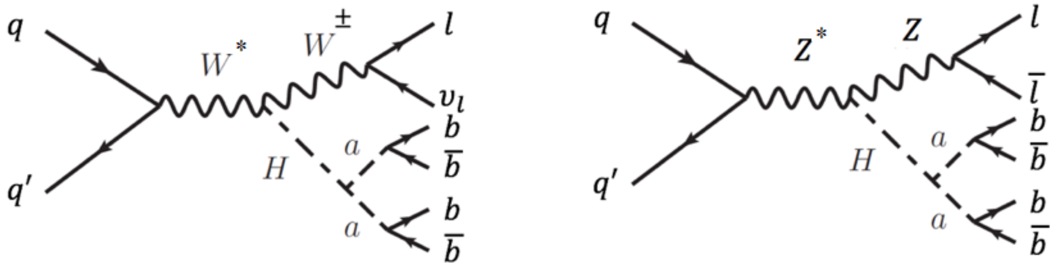


Figure 3-1: Feynman diagrams of W associated production (left) and Z associate production (right). Only W associated production channel is considered in this analysis

3.2 Motivation

After the discovery of a Higgs-like particle by the CMS and ATLAS experiments at the LHC in 2012 [87, 88], additional measurements of its properties using the full datasets at $\sqrt{s} = 7$ and 8 GeV revealed that the observed state with a mass near 125 GeV is consistent with the predicted SM Higgs [89-91]. The measurements of the Higgs production cross-sections and branching ratios illustrate this agreement (Figure 3-2)

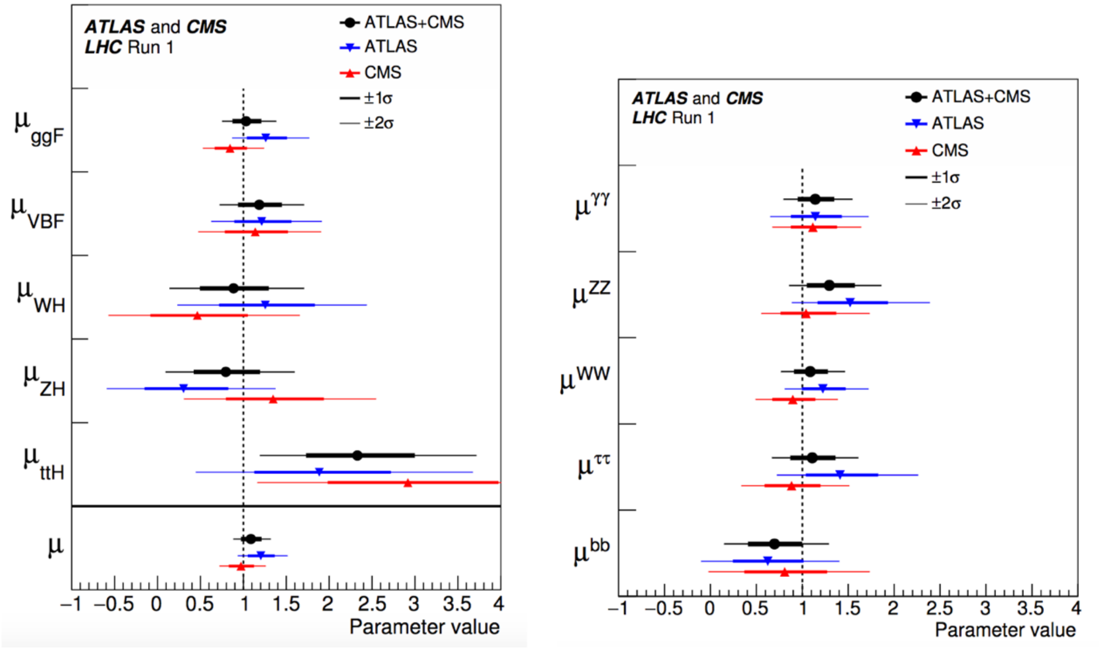


Figure 3-2: Best fit results for production signal strengths in different Higgs boson production processes' cross-sections (left) and branching ratios (right) for the combination of ATLAS and CMS data, normalized to Standard Model predictions. The error bars indicate the 1σ *(thick lines) and* 2σ *(thin lines) intervals. [92]*

After this exciting discovery, the experimental investigation of this new state began. This experimental investigation includes:

- Precise measurement of the Higgs couplings to SM particles
- Search for additional Higgs-like states
- Search for exotic decays of Higgs

This dissertation falls in the third category. Exotic or non-standard decays of Higgs are the ones that are not predicted by the SM. Higgs boson has a very narrow width (SM prediction: 4.2 MeV; LHC run 1: < 22 MeV at 95% CL, subject to various assumptions [93]) to mass ratio due to the suppression of tree-level Yukawa couplings. Therefore, even if the coupling to a non-SM light state is small, it can still open non-negligible decay modes. [94]

Among the topologies of exotic decays, the $h \rightarrow XX \rightarrow 2Y2Y'$ is naturally realized in many well-motivated BSM frameworks including R-symmetry limit of the NMSSM, two Higgs doublet models containing additional singlet Higgs fields (2HDM+S), extensions of the SM with hidden light gauge bosons, the Little Higgs model, extensions of the SM with hidden light gauge bosons, and commonly in the Hidden Valley scenario[18] [86, 95, 96]. The NMSSM is one of the best-studied examples of this topology in which an approximate R-symmetry yields a SM-like Higgs boson with considerable branching ratio into a pair of light pseudoscalars 'a' followed by 'a' decaying to 2X2Y with X, Y: fermions (μ, τ), b-quarks, gluons or γ's. Table 3-1 summarizes the studies of Higgs exotic decays to 2a to 2X2Y both in CMS and ATLAS.

Some of the studies that consider the possible decay of the SM-like Higgs to a pair of lighter Higgs bosons are [86, 97-104] in 2HDM context, [86, 105-108] in NMSSM and NMSSM-like context and [86, 109, 110] in the general case of adding a singlet field to the SM or the 2HDM.

It is worth mentioning that $h \rightarrow aZ$ can also happen in BSM physics if $m_a < m_h - m_Z$. However, this analysis only pertains to $h \rightarrow aa$ [96].

18- But not possible in the (CP-conserving) minimal supersymmetric standard model (MSSM) because of the tightly constrained nature of its Higgs sector.

Table 3-1: Overview of Run1 [8 TeV] and Run2 [13 TeV] $H \to aa$ searches in ATLAS and CMS. This table is taken from [96] and [111]

Final State	**CMS**			**Atlas**		
	Dataset	**ID**	**arXiv**	**Dataset**	**ID**	**arXiv**
$H \to aa \to 4l$ ($h \to ZZ_d/Z_dZ_d \to 4l$)	8 TeV + 13 TeV	Phys.Lett.B 752 (2016) 146, HIG-18-003 (PAS)	1506.00424	13 TeV	JHEP06(2018)166	1802.03388
$H \to aa \to 4\tau$	8 TeV	JHEP(2016)079	1510.06534 1701.02032	-	-	-
$H \to aa \to 2b2\mu$	8 TeV	JHEP(2017)076	1701.02032	13 TeV	Submitted to PLB	1807.00539
$H \to aa \to 2\mu2\tau$	8 TeV + 13 TeV	Submitted to JHEP	1701.02032 1805.04865	8 TeV	Phys.Rev.D92 (2015) 052002	1505.01609
$H \to aa \to 2b2\tau$	13 TeV	Submitted to PLB	1805.10191	-	-	-
$H \to aa \to 4b$	13 TeV	This analysis!		13TeV	Submitted to JHEP	1806.07355
$H \to aa \to 4\gamma$	-	-	-	8 TeV	EPJC 76(4) 2016	1509.05051
$H \to aa \to 2\gamma 2g$	-	-	-	13 TeV	Phys.Lett.B 782 (2018) 750	1803.11145

The more precise measurements of the Higgs Boson decay branching ratios constrain the extended Higgs sector more every day. However, the possibility of decays of the type $H \to hh$ (h is a lighter scalar) or $H \to aa$ (a is a lighter pseudoscalar) is still considerable. Using the observed decays of the Higgs Boson, we can only put a limit on unseen decays which we will show as $H \to U$. These unseen decays potentially contain $H \to aa$ and $H \to hh$. According to [112, 113]:

- $BR(H \to U) < 0.09$ at 95.4% C.L. for a Higgs with completely SM-like couplings
- $BR(H \to U) < 0.23$ for a SM Higgs if extra loop contributions to its $\gamma\gamma$ and gg couplings are allowed
- $BR(H \to U) < 0.22$ if the couplings to up quarks, down quarks and vector bosons can vary within a general model with only doublets and singlets in the Higgs sector (and no extra loop contributions to the gg and $\gamma\gamma$ couplings). If the up, down and

> vector boson couplings are allowed to vary completely freely (except for restricting vector boson couplings to be less than their SM values), then all LHC rates can be reproduced if all the couplings-squared are increased by a factor of $\frac{1}{1-\mathrm{BR}(\mathrm{H}\to\mathrm{U})}$. The only limit arises from direct limits on the observed Higgs total width. At the moment, this is at the level of $\Gamma_{\mathrm{tot}} \leq 4\Gamma_{\mathrm{tot}}^{\mathrm{SM}}$ [114].

If the couplings-squared are all increased by $\frac{1}{1-\mathrm{BR}(\mathrm{H}\to\mathrm{U})}$, the rates for $gg \to H \to U$ and other production mechanisms are all increased by a factor of $\frac{\mathrm{BR}(\mathrm{H}\to\mathrm{U})}{1-\mathrm{BR}(\mathrm{H}\to\mathrm{U})}$, making such modes even more accessible. However, even if one adopts the more conservative approach of only considering doublets + singlets models, there is still an excellent prospect for seeing Higgs pair modes if $\mathrm{BR}(\mathrm{H} \to \mathrm{U}) \leq 0.22$.

$\mathrm{BR}(\mathrm{H} \to \mathrm{aa}, \mathrm{hh})$ are constrained by the requirement that $\mathrm{BR}(\mathrm{H} \to \mathrm{U}) \leq 0.22$. In other words:

$$\mathrm{BR}(\mathrm{H} \to \mathrm{aa}) + \mathrm{BR}(\mathrm{H} \to \mathrm{hh}) < \mathrm{BR}(\mathrm{H} \to \mathrm{U}) \leq 0.22 \tag{30}$$

In the NMSSM, the pseudoscalar mass eigenstate is defined by

$$a = \cos\theta_A \, a_{MSSM} + \sin\theta_A \, a_S \tag{31}$$

Where a_{MSSM} is the MSSM-like pseudoscalar and a_S is the singlet pseudoscalar of the NMSSM. To keep the H_{aa} coupling constant small, we must have

$$\cos\theta_A \ll 1 \tag{32}$$

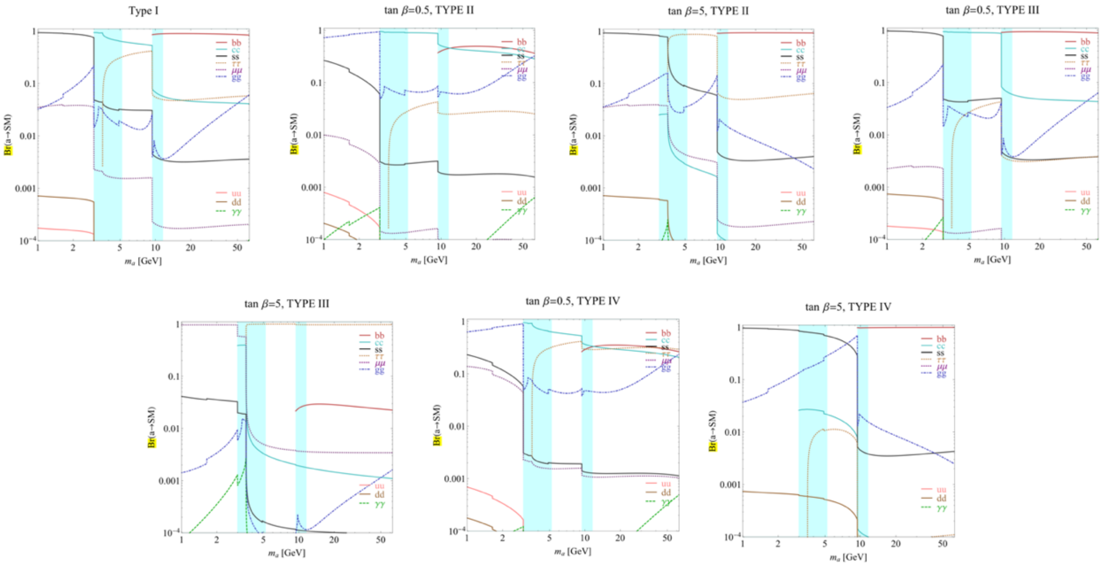
To suppress the doublet content of the a.

3.2.1 Motivation for $H \rightarrow aa \rightarrow 4b$ channel

Among the $h \rightarrow aa \rightarrow 2Y2Y'$ decays, $h \rightarrow aa \rightarrow 4b$ is theoretically very motivated mainly because $aa \rightarrow 4b$ can be the dominant decay in some regions of the parameter space (Figure 3-3)[96].

The 4b final state can be dominant in several classes of models:

- R-symmetry limit of the NMSSM: In this model, the coupling of a to fermions is proportional to the Yukawa matrices, which are intensified by sin β/sin α. This creates large decay branching ratios for $a \rightarrow 2b$ in large regions of the parameter space [96].
- 2HDM+S: Depending on tanβ, the decay $a \rightarrow 2b$ is generic in all four 2HDM Types if kinematically allowed, i.e., $m_a > 2m_b$ [96].
- Little Higgs models: Like R-symmetry limit of the NMSSM, the couplings of a to SM fermions are again proportional to the SM Yukawa matrices if Minimal Flavor Violation (MFV) is imposed. Therefore, the coupling to the b-quark is typically increased again [96].


Type I

tan β=0.5, TYPE II

tan β=5, TYPE II

tan β=0.5, TYPE III

tan β=5, TYPE III

tan β=0.5, TYPE IV

tan β=5, TYPE IV

Figure 3-3: Branching ratios of a singlet-like pseudoscalar in different types of the 2HDM+S model; Branching ratio of $a \to bb$ often dominates when kinematically allowed, i.e., $m_a > 2m_b$ [86]

In a more general view, if NMSSM is a true explanation of the nature, then the SM-like Higgs boson could be either the lightest CP-even scalar or the second lightest CP-even scalar [105]. Therefore, two scenarios are possible as [115] explains:

- "NMSSM1 scenario: the lightest Higgs boson acts as the SM-like Higgs boson h. In this scenario, the mixing effect[19] is to pull down m_h, and if the mixing effect is dominant, large radiative correction is needed to predict $m_h \simeq 125$ GeV.
- NMSSM2 scenario: the next-to-lightest Higgs boson acts as the SM-like Higgs boson h. In this scenario, the mixing effect is to push up m_h. Both the mixing effect

---

[19] - Meaning the stop mixing that contributes to the SM-like Higgs mass

and the additional tree level contribution make the large radiative correction unnecessary."

In other words, two kinds of analyses can be done with 4b final state:

- NMSSM1: A massive resonance decays to a pair of SM-like Higgs with 4b final state ($X \to hh \to 4b$). A few other models also predict this channel. Some examples of this kind of analysis both at CMS and ATLAS are [116-121].
- NMSSM2: The SM-like Higgs decays to a pair of lighter pseudo-scalar or pseudoscalar Higgs. This scenario is more motivated than NMSSM1 since it does not require large radiative correction for m_h. Although the $h \to aa \to 4b$ analysis is done in ATLAS ([118, 122]), it was never done in CMS. This analysis is done to fill this vacancy.

If the pair of pseudo-scalar Higgs (a) is boosted (Mass < 20-30 GeV), then $h \to aa \to 4b$ in principle exists in $h \to 2b$ search signal region. This is because when the boosted light Higgs decays to a $b\bar{b}$ pair, the pair is collimated, and the two b-jets will be reconstructed as one jet. The most recent $h \to 2b$ results from CMS and ATLAS confirm this decay with the significance of 5.6 and 5.4 standard deviations respectively. Signal strength[20] is 1.04 ± 0.2 (for CMS) and $1.01 \pm 0.12(stat.)^{+0.16}_{-0.15}(syst.)$ for ATLAS [124, 125]. This result can be used to limit the branching ratio of $h \to 2a \to 4b$.

20- Defined as the ratio of the best-fit value for the production cross-section for a 125 GeV Higgs boson, relative to the standard model cross-section. 123. Chatrchyan, S., et al., *Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks*. 2014. **89**(1): p. 012003.

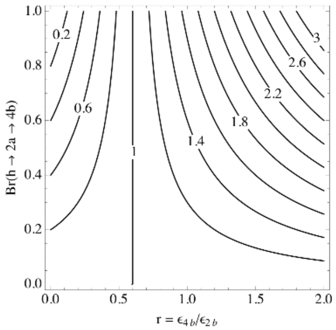


Figure 3-4: Simulated signal strength in $h \rightarrow 2b$ *search vs* $Br(h \rightarrow 2a \rightarrow 4b)$ *vs efficiency ratio (r). This plot is simulated with run 1 setup assuming* $Br(h \rightarrow 2b)^{SM} = 0.6$. *This plot is taken from [86]*

Figure 3-4 shows the simulated signal strength of $h \rightarrow 2b$ vs the branching ratio of $h \rightarrow 2a \rightarrow 4b$ vs the efficiency ratio r. The efficiency ratio (The x-axis of Figure 3-4) is defined as:

$$r_{4b}(m_a) = \frac{\epsilon_{h\rightarrow 2a\rightarrow 4b}}{\epsilon_{h\rightarrow 2b}} \tag{33}$$

Where $\epsilon_{h\rightarrow 2a\rightarrow 4b}$ is the probability that a $h \rightarrow 2a \rightarrow 4b$ decay is confused with $h \rightarrow 2b$ and ends up in its signal region. And likewise, $\epsilon_{h\rightarrow 2b}$ is the probability that a $h \rightarrow 2b$ decay is identified correctly and put in the signal region [86]. The simulation of [86] using run 1 configuration gives us the value of $r = \sim 1.5$ for $m_a = 15\ GeV$. Although this simulation is done with run 1 configuration, it can be used as an approximation.

In 2018 CMS $h \rightarrow 2b$ paper, we have $signal\ strength < 1.24$, and therefore from Figure 3-4 we can read:

$$Br(h \rightarrow 2a \rightarrow 4b) < 15\% \text{ for } m_a = 15\ GeV$$

In this analysis, we are only interested in the channel shown on the left side of Figure 3-1 which we call signal. All other processes that happen during proton-proton collisions is called background. Although in principle it is possible to have background from other BSM theories, we expect the effect to be negligible. Therefore, we only talk about SM background. The main goal of this analysis is to isolate signal events by rejecting SM background, then evaluate our performance and the results using statistical methods.

Background rejection starts while taking the data at CMS as explained in the previous chapter (Trigger primitives, Level 1 trigger, and HLT). In other words, the background of the data we start with is already strongly suppressed. The next step in our SM background suppression is to use our basic understanding of high energy physics and find the so-called pre-selection requirements that can distinguish between signal and SM background. This is done in section 3.6.1. Then we will mix our knowledge of physics with machine learning techniques in part 3.5.2 to reduce SM background even more. However, if we keep making more strict requirements to suppress the SM backgrounds, we will also lose signal events. Therefore, we have trade off signal acceptance against background rejection. As a result, instead of trying to eliminate background even further we continue by estimating the amount of remaining background events instead. This is done in part 3.

3.3 Datasets

3.3.1 Data

In this analysis, we used 35.9 fb^{-1} of proton-proton collision data taken in 2016 at CMS at $\sqrt{s} = 13\ GeV$. The collision data were analyzed using different global tags as shown in Table 3-2. Since we only considered the WH production that come with a high-p_T isolated lepton, only single lepton datasets (SingleElectron + SingleMu) are used both for the signal and control regions. Efficiencies are measured using dedicated trigger paths stored in the same datasets as the ones used for our search. All data are analyzed using "Cert 271036-284044 13TeV 23Sep2016ReReco Collisions16 JSON.txt" file to select the good luminosity sections.

Table 3-2: The MiniAODv2 datasets used for this analysis. The primary datasets correspond to SingleElectron and SingleMu. The integrated luminosity and the run-ranges are shown for each data period.

Dataset	**Global Tag**	$\int \mathcal{L}\,(fb^{-1})$	**Run range**
/Run2016B-03Feb2017 ver2-v2	80X dataRun2 2016SeptRepro v7	5.933	273150-275376
/Run2016C-03Feb2017-v1	80X dataRun2 2016SeptRepro v7	2.646	275656-276283
/Run2016D-03Feb2017-v1	80X dataRun2 2016SeptRepro v7	4.353	276315-276811
/Run2016E-03Feb2017-v1	80X dataRun2 2016SeptRepro v7	4.117	276831-277420
/Run2016F-03Feb2017-v1	80X dataRun2 2016SeptRepro v7	3.186	277932-278808
/Run2016G-03Feb2017-v1	80X dataRun2 2016SeptRepro v7	7.721	278820-280385
/Run2016H-03Feb2017 ver2-v1	80X dataRun2 Prompt v16	8.636	281207-284035
/Run2016H-03Feb2017 ver3-v1	80X dataRun2 Prompt v16	0.221	284036-284068
Total		**35.9**	

3.3.2 Background MC

The Monte Carlo (MC) datasets that we used in this analysis for relevant SM background processes are tabulated in Table 3-3 (The cross-sections are based on MCFM from [126]). We will briefly explain these SM backgrounds in section 3.6.1. All samples are generated using MadGraph5 at LO [127], PYTHIA [30, 128], and POWHEG[21] [130] generators and a full simulation of the CMS detector based on GEANT4 [131].

- $t\bar{t}$+jets is the dominant irreducible[22] SM background in this analysis and it is simulated with POWHEG.
- $W \to l\nu$ and DY ($W \to ll$) are simulated with Madgraph at LO. However, we applied the k-factors[23] for both processes. For $W \to l\nu$, the k-factor is 1.21, and for DY it is 1.21 and 1.17 in the dilepton mass range of 10-50 and above 50 respectively.
- The single top and di-boson processes are simulated with POWHEG with NLO cross-sections
- And aMC@NLO is used for the $t\bar{t}$+X processes with NLO cross-sections

---

21- POWHEG is a method for mixing the accuracy of NLO calculations with speed of parton shower programs. It generates the hardest emission with NLO accuracy first, which creates events with positive weights. Then it uses the LO Shower Monte Carlo programs like Pythia to perform the rest of the shower.
129. Nason, P., B.J.A.R.o.N. Webber, and P. Science, *Next-to-leading-order event generators*. 2012. **62**: p. 187-213.
22- Irreducible background is defined in section 3.
23- Ratio of NLO to LO.

3.3.3 Signal MC

The Monte Carlo signal samples for associated WH production were generated at LO with MadGraph5 [132] requiring that W decays leptonically (~11% per lepton flavor from the inclusive W decays) and $H \rightarrow aa \rightarrow 4b$ was also forced. The NMSSM model used in MadGraph is taken from [133][24] with ickkw=1 (MLM matching) and xqcut = 15. The SM Higgs mass in NMSSM was set to 125 GeV which is the mass found at LHC in 2012. Also, a mass scan is performed over m_a from 12 (A little higher than $2m_b$) up to 60 GeV (just below $\frac{m_{Higgs}}{2}$) in increments of 12, 15, 20, 25, 30, 40, 50, and 60 GeV. We took the SM WH cross-section from the Yellow Report 4 of the LHC Higgs Cross-Section Working Group (LHC Higgs XS WG)[134-136]

We did a study comparing the properties of particles at MC level with and without ISR and we observed no noticeable difference. Therefore, we did not consider ISR in our final version of MadGraph MC production. The output of the MadGraph was given to Pythia 8.1 for further decays, hadronization and detector simulation as instructed by the MC experts. The final result was ~900k MiniAOD events for each m_a mass point.

24- There is a more specific NMSSM model only for $H \rightarrow aa$ decays on the “Exotic Higgs Decays” website (http://exotichiggs.physics.sunysb.edu/web/wopr/). However, this model does not contain the whole NMSSM, and we could not model $W^{\pm} \rightarrow l^{\pm}\upsilon$ that is needed in our analysis with it!

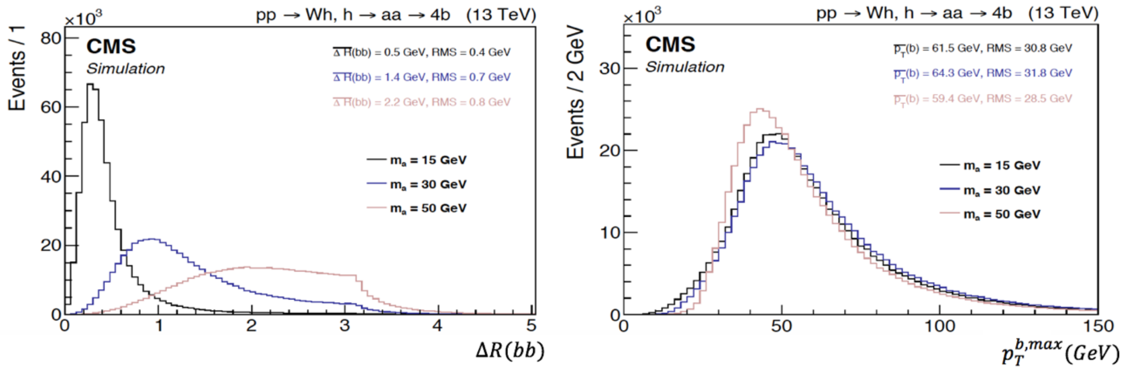


Figure 3-5: Two plots using MC level information. (left) As a mass decreases, the b quarks from the same parent tend to be collimated; (right) b-quarks from the a-decay tend to have very low p_T in general.

Table 3-3: List of the SM MC samples used. S16 is an abbreviation for RunIISummer16MiniAODv2-PUMoriond17 80X mcRun2 asymptotic 2016 TrancheIV v6(ext)-v1.*

Process	**Dataset**	$\boldsymbol{\sigma . k(pb}$
W → lυ	/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	61526.7
	/W1JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	9493
	/W2JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	3120
	/W3JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	942.3
	/W4JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	524.2
Z → l l	/DYJetsToLL M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	18610
	/DY1JetsToLL M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	725
	/DY2JetsToLL M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	394.5
	/DY3JetsToLL M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	96.47
	/DY4JetsToLL M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	34.84
	/DYJetsToLL M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	5765.4
	/DY1JetsToLL M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	1016
	/DY2JetsToLL M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	331.4
	/DY3JetsToLL M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	96.36
	/DY4JetsToLL M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/S16	51.4
$t\bar{t}$ + X	/TT TuneCUETP8M2T4 13TeV-powheg-pythia8/S16	831.76
	/TTGJets_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/S16	3.697
	/TGJets_TuneCUETP8M1_13TeV amcatnlo madspin pythia8/S16	2.967
	/TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/S16	0.2043
	/TTZToLLNuNu M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8/S16	0.2529
Single Top	/ST s-channel 4f leptonDecays 13TeV-amcatnlo-pythia8 TuneCUETP8M1/S16	3.362
	/ST t-channel top 4f inclusiveDecays 13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1/S16	136.02
	/ST t-channel antitop 4f inclusiveDecays 13TeV-powhegV2-madspin-pythia8 TuneCUETP8M1/S16	80.95
	/ST tW antitop 5f inclusiveDecays 13TeV-powheg-pythia8 TuneCUETP8M1/S16	35.6
Dibosons	/ZZ_TuneCUETP8M1_13TeV-pythia8/S16	16.523
	/WWTo2L2Nu 13TeV-powheg/S16	12.178
	/WWToLNuQQ 13TeV-powheg/S16	49.997
	/WZ_TuneCUETP8M1_13TeV-pythia8/S16	47.13
Triboson	/ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/S16	0.01398
	/WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/S16	0.05565
	/WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/S16	0.1651
	/WWW 4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8/S16	0.2086
QCD	/QCD Pt-*to*_EMEnriched_TuneCUETP8M1_13TeV_pythia8/S16	
	/QCD Pt-20toInf MuEnrichedPt15_TuneCUETP8M1_13TeV_pythia8/S16	

3.3.3.1 Pile-up reweighting

As defined in the previous chapter, pile-up is the number of interactions per bunch crossing [137]. The MC generation at CMS uses a generic pile-up distribution centered at the expected number of pile-up interactions in data. This allows researchers to reweight the MC pile-up distributions to what they get from data as data taking can happen or continue after the MC is generated. In this analysis, we reweighed the in-time pileup distribution of MC to the estimated value in data. We chose the minimum-bias cross-section value to be 69.2 mb to have the best possible agreement with the distribution of the reconstructed vertices simulation. The uncertainty of the minimum-bias cross-section value is chosen to be 5% to evaluate the sensitivity of the analysis on pileup. Figure 3-6 shows the vertex multiplicity distribution before and after reweighting using the one lepton datasets (before any cut or requirement).

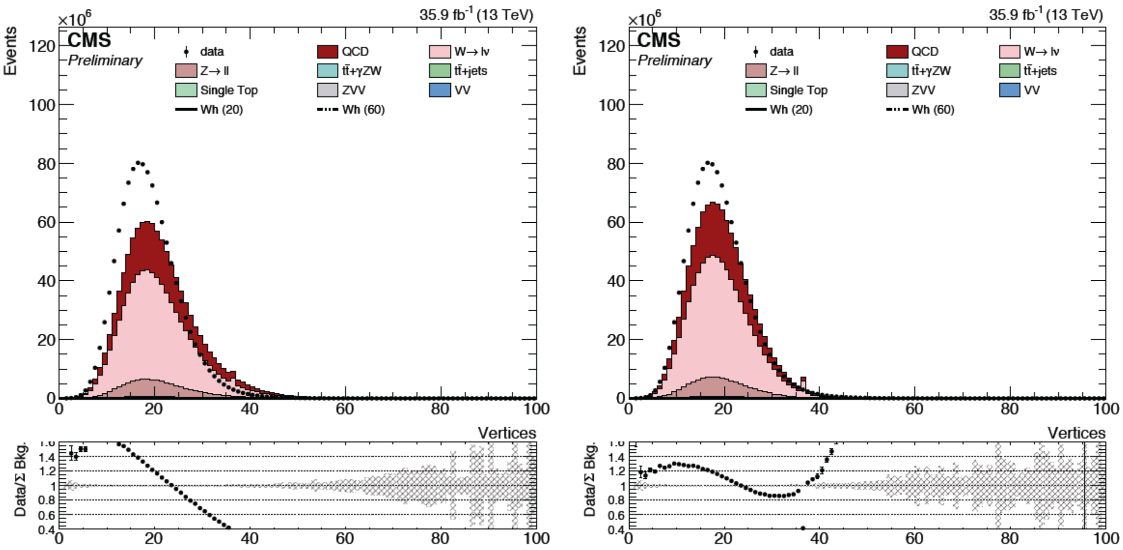


Figure 3-6: Vertex multiplicity distribution in data and MC before (left) and after (right) the pileup reweighting based on the estimated distribution of the number of interactions in each event.

3.4 Event Selection

3.4.1 Introduction

In this part we explain the selections and corrections that are used to:

- Reduce malfunctions of the detector
- “Account for known differences between the actual detector and the simulated detector. Because the trigger conditions typically vary over the data-taking period, it is difficult to model them correctly in simulated event samples. To simplify the determination of the trigger efficiency, one typically applies an offline selection requirement that is somewhat more stringent than the most stringent trigger requirement, establishing a uniform condition over the full running period.“ [138]
- Isolate the signal process and reject background events

Before we start. Let’s define Primary vertices. Primary Vertices (PV) are vertices which are required to be reconstructed inside a cylindrical region of length 24 cm and radius 2 cm surrounding the nominal interaction point. They are in addition required to have a minimum of 4 degrees of freedom regarding the vertex fit. When multiple PV’s are present in an event, the vertex with the highest $\sum_{tracks} p_T^2$ is chosen.

The triggers used online as the first step of background suppression are listed in Table 3-4 with the terminology in Table 3-5. The same trigger paths are also used for the control regions defined and used for the background estimation.

Table 3-4: The online trigger paths used in the analysis to suppress the SM background

Dataset	Trigger path	Used for
SingleMu	HLT_IsoMu24 HLT_IsoTkMu24	Signal and control regions
SingleElectron	HLT_Ele27_WPTight_Gsf	Signal and control regions

Table 3-5: Meaning of the terms in the HLT paths. Isolation is defined in part 3.4.2.5

Term	Meaning
IsoMuXX	Existence of at least one isolated muon with $p_T > XX$
IsoTkMuXX	Existence of at least one isolated muon with tracker $p_T > XX$
EleXX	Existence of at least one electron with $p_T > XX$
Eta2p1	Within coverage of $\|\eta\| < 2.1$
WPTight/WPLoose	Use of the tight/loose working point
Gsf	Tracks fit with Gsf filter [139]

3.4.2 Preselection

3.4.2.1 Sample cleanup filters

Sample clean up filters are applied to minimize [140]:

- The contamination from beam-gas interactions: As described in the previous chapter, the interaction of the beam with the residual gas in the vacuum chamber tube can cause machine-induced secondary particles. [141]

- Beam-halo: Particles moving far from the beam center. How far may be somewhat arbitrary [142, 143]. Beam-halo can create machine-induced secondary particles as well by integrating with limiting apertures. [141]
- Calorimeter noise.

3.4.2.2 The identification ID requirements

3.4.2.2.1 Electrons

We are interested in triggering on the lepton from the WH associated Higgs production. We don't expect this lepton to be low energy, therefore we consider both ECAL and tracker electrons with $p_T > 25\ GeV$ and $|\eta| < 2.5$ (tracker range for electrons) in our analysis, except for those whose superclusters falling in the barrel-endcap gap region (i.e., $1.4442 < |\eta| < 1.5660$). As described in the previous chapter, CMS can easily confuse particles due to the uncertainties in the measurement of particle properties. Therefore, the requirements in Table 3-7 (description in Table 3-6) are imposed to minimize electron mis-identification in CMS.

It is worth mentioning that the cuts on the impact parameter are no longer included in the tuned identification criteria (Table 3-7). The main reasons are

- The cut efficiency depends significantly on the physics of the event
- And the discrimination power of the variables suffers from the significant pile-up increase in 2016.

Table 3-6: Description of the variables in Table 3-7. [144-147]

Requirement	Description
$\sigma_{i\eta i\eta}$	A measure of the spread in η in units of crystals of the electron’s energy in a 5×5 block centered on the seed crystal.
$\Delta\eta_{in}$	Difference in eta between track and supercluster[25]
$\Delta\phi_{\mathrm{in}}$	Difference in phi between track and supercluster
H/E	Ratio of the energy deposited in HCAL over energy deposited in ECAL within $\Delta R < 0.1$
$\|1/\mathrm{E} - 1/\mathrm{p}\|$	$\frac{1}{energy} - \frac{1}{momentum}$. Motivation: Many fakes from jets have $\frac{E}{p} < 1$ partly because of the low response of ECAL to charged pions. However, $\frac{E}{p}$ is not often measured less than 1 for electrinos.
Missing hits	Number of missing hits along the track trajectory in the inner tracker. This is required to further suppress electrons from decays in flight.
Pass conversion veto	A flag to identify electrons that originated from photons (photon conversions). A real electron should have a compatible track in the tracker.
PF relative isolation	Explained in part 3.4.2.5

---

[25] Superclusters are clusters of BasicClusters to collect radiated energy (look at the description of ECAL mechanism in the previous chapter).

Table 3-7: Electron id requirements[148]

Requirement	**Barrel**		**Endcap**	
	Loose	**Tight**	**Loose**	**Tight**
$\sigma_{i\eta i\eta}$	0.011	0.00998	0.0314	0.0292
$\Delta\eta_{in}$	0.00477	0.00308	0.00868	0.00605
$\Delta\phi_{\mathrm{in}}$	0.222	0.0816	0.213	0.0394
H/E	0.298	0.0414	0.101	0.0641
$\|1/\mathrm{E}-1/\mathrm{p}\|$	0.241	0.0129	0.14	0.0129
Missing hits	1	1	1	1
Pass conversion veto	yes	Yes	yes	yes
PF relative isolation	0.0994	0.0588	0.107	0.0571

3.4.2.2.2 Muons

Both tracker and standalone muons with $p_T > 25\ GeV$ and $|\eta| < 2.4$ (tracker range for muons) are considered. In order to minimize muon mis-identification, the requirements in Table 3-9 (description in Table 3-6) are imposed.

Table 3-8: Description of the variables in Table 3-7. [145, 149]

Requirement	Description
Global	The muon candidate has to be reconstructed as a global muon. For each standalone muon track, a search for tracks matching it among those reconstructed in the inner tracking system is performed, and the best-matching tracker track is selected. The result is referred to as "global muons".
PF	Particle flow reconstruction
χ^2/DOF (Global)	If the muon p_T is less than 200 GeV, χ^2 value of the track fit normalized to the number of degrees of freedom, χ^2/DOF must be less than 10. If p_T is greater than 200 GeV, the uncertainty on p_T must be less than 30% of the measured p_T value.
Valid pixel hits	Number of pixel hits. This is required to further suppress muons from decays in flight.
Pixel layers with measurement	Number of hits in the inner pixel system
Tracker layers with measurement	Number of tracker layers with hits. This is required to guarantee a good p_T measurement and also suppress muons from decays in flight.
Valid muon hits	The number of muon-chamber hits included in the global muon track fit. This is required to suppress hadronic punch-through and muons from decays in flight.
Matched muon stations	Muon segments in the muon stations. This is required to suppress punch-through and accidental track-to-segment matches.
$d_0(vtx)$ (cm)	The transverse impact parameter of the tracker track with respect to the PV. This is required to suppress cosmic muons and further suppress muons from decays in flight.
$d_z(vtx)$ (cm)	The longitudinal distance of the tracker track with respect to the PV. This is required to suppress cosmic muons, muons from decays in flight and tracks from PU.
PF rel iso **track rel iso**	Explained in part 3.4.2.5

Table 3-9: Muon ID requirements

Requirement	Soft	Loose	Tight
Global		Or tracker	True
PF		True	True
χ^2/DOF (Global)			< 10
Valid pixel hits			> 0
Pixel layers with measurement	> 0		
Tracker layers with measurement	> 5		> 5
Valid muon hits			> 0
Matched muon stations			> 1
$d_0(vtx)$ (cm)	0.3		0.2
$d_z(vtx)$ (cm)	20		0.5
PF rel iso		< 0.2	< 0.15
track rel iso		< 0.1	0.1

We apply all electron and muon corrections to account for differences in the energy scale, trigger, reconstruction/identification and isolation efficiencies between data and MC.

It is interesting to know that there is an asymmetry in the production of electrons and muons. In other words, e^+ and μ^+ are generated more often than e^- and μ^- from the W decay! This is because W^+ is most often made by an up and an anti-down quark while W^+ is made by an anti-up and a down quark. Since protons have two valence up-quarks and one valence down-quark, W^+ has a higher chance of being produced!

3.4.2.2.3 Jets

The AK4 algorithm [150, 151] reconstructs jets from all the PF candidates. The reconstruction may be seeded using all reconstructed particle flow candidates after having removed the charged hadron candidates, which are not associated to the primary vertex of the event (charge hadron subtracted ak4PFJetsCHS). The energy of the reconstructed jets is corrected in 3 steps:

- L1 FastJet (for pileup/underlying event subtraction
- L2 (for relative corrections)
- L3 for absolute scale corrections.

For data an extra residual correction is included in the absolute scale correction. The median r of the energy density of the reconstructed jets with $p_T > 10$ GeV is used as an estimator for the contamination of the jet energy scale (or the lepton or photon isolation). An extra correction is applied for the simulated jets in order to reproduce the measured jet energy resolution. For each jet, the transverse momentum is smeared using the data-MC scale factors for the jet energy resolution as provided by the JetMET POG[26]. Using the transverse momentum of the generator level matched jet and the measured data-MC resolution ratio, the corrected transverse momentum is given by:

$$p_T \rightarrow \max\left\{0, p_T^{gen} + g_{aus}(\bar{c}, \sigma_c) \cdot \left(p_T - p_T^{gen}\right)\right\} \qquad (34)$$

---

[26] - Physics object Group

where c is the core resolution scaling factor, i.e., the data to MC scale factor for the jet energy resolution, and σ_c is the corresponding uncertainty. In the case where the reconstructed jet has no match in MC, no smearing is applied.

We consider for the analysis all jets with $p_T > 20$ GeV and $|\eta| < 4.7$ passing the PF-loose requirements [152] summarized in Table 3-10. No PU jet id requirement is currently imposed.

Table 3-10: Jet ID requirements

ID	Variable	Region	Comment
PF-loose	$n \quad > 1$	$\|\eta\| < 2.7$	Number of constituents
	$nhf < 0.99$	$\|\eta\| < 2.7$	Neutral hadron energy fraction
	$nef < 0.99$	$\|\eta\| < 2.7$	Neutral electromagnetic energy fraction
	$cef < 0.99$	$\|\eta\| < 2.4$	Charged electromagnetic energy fraction
	$chf > 0$	$\|\eta\| < 2.4$	Charged hadron energy fraction
	$nch > 0$	$\|\eta\| < 2.4$	Charged multiplicity
	$nef > 0.01$	$2.7 < \|\eta\| \leq 3.0$	Neutral electromagnetic energy fraction
	$nhf < 0.98$	$2.7 < \|\eta\| \leq 3.0$	Neutral hadron energy fraction
	$nnp > 2$	$2.7 < \|\eta\| \leq 3.0$	Number of neutral particles
	$nef < 0.9$	$\|\eta\| > 3.0$	Neutral electromagnetic energy fraction
	$nnp > 10$	$\|\eta\| > 3.0$	Number of neutral particles

3.4.2.3 Relative Isolation criteria

The leptons that we are interested in to trigger on are leptons produced from the WH associated production. Other leptons generated from hadronization or other decays are expected to be collimated with a jet. Therefore, the isolation of a lepton (that there is no jet near it) is a good way to distinguish between primary and secondary leptons although it does not provide a perfect separation. Appling an isolated lepton requirement significantly reduces the QCD background.

The isolation of a lepton is computed as the scalar sum of the p_T of all PF candidates, excluding the lepton itself, within a cone of $\Delta R = 0.3$ around the lepton direction [153] where:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{35}$$

Then the relative isolation is defined as:

$$\mathrm{I_{rel}} = \frac{I}{p_{lep}} \tag{36}$$

Transverse momentum of particles is computed independently for the charged hadrons, neutral hadrons and photon candidates. When dealing with electron candidates, the neutral flux is corrected by using the average energy density due to pileup and underlying event in the central region of the detector (r) and an effective area (A_{eff}) correction which normalizes this estimator in such a way that the isolation is independent of the number of pileup interactions. The electron isolation is therefore defined as:

$$I_{rel} = \frac{1}{p_T}\left[I_{ch} + \max\left\{I_{nh} + I_g - A_{eff} \cdot \rho, 0\right\}\right] \tag{37}$$

And for the muons:

$$I_{rel} = \frac{1}{p_T}\left[I_{ch} + \max\left\{I_{nh} + I_g - 0.5 \cdot I_{chPU}, 0\right\}\right] \tag{38}$$

Where I_{chPU} is the "$\Delta\beta$ correction" to correct for pileup contamination. It is calculated as the sum of transverse energy of all charged particles not coming from the primary vertex in the cone of interest and the factor 0.5 corresponds to a naive average of neutral to charged particles and has been measured in jets in [154, 155].

3.4.2.4 Missing Transverse Energy (E_T^{miss})

Missing energy exists in our signal region from the neutrino coming from W decay where W is coming from the WH production. Therefore, missing energy can be used as a discriminator between background and signal. We require than $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV to suppress background. This requirement especially affects QCD background. [138]

$\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ is estimated from the imbalance of the transverse momentum of all the reconstructed particle flow candidates as:

$$E_T^{miss} = -\sum_{i=1}^{N_{PF}} \vec{p}_{T,i} \tag{39}$$

The jet energy scale corrections may be reflected in the estimate of the E_T^{miss} by adding an extra term, which corresponds to the net balance of the clustered flux after correcting individually each jet's energy. This estimator is usually referred to as Type-I E_T^{miss} estimator and can be written as:

$$typeI - E_T^{miss} = E_T^{miss} - \sum_{i=1}^{N_{jets}} \delta\vec{p}_{T,j} \tag{40}$$

The Type-I correction is computed from CHS (ak4PFJetsCHS) [15], as prescribed by the JetMET group.

3.4.2.5 Transverse Mass

Transverse mass is defined as (in natural units) [156]:

$$M_T = \sqrt{m^2 + p_x^2 + p_y^2} = \sqrt{E^2 - p_z^2} \tag{41}$$

And it is useful because it is invariant under the Lorentz boost along the beam direction (z-direction). If particle a decays to particles b and c, i.e. $a \to bc$, then we will have:

$$M_{T,a} = \sqrt{E_a^2 \ - p_{z,a}^2} = \sqrt{(E_b + E_c)^2 \ - (p_{z,b} + p_{z,c})^2}$$

$$= \sqrt{E_b^2 + E_c^2 + 2E_bE_c \ - p_{z,b}^2 - p_{z,c}^2} - 2p_{z,b} \cdot p_{z,c} \tag{42}$$

$$= \sqrt{M_{T,b}^2 + M_{T,c}^2 \ + 2(E_bE_c - 2p_{z,b} \cdot p_{z,c})}$$

Therefore, the transverse mass for particle a is:

$$M_{T,a} = \sqrt{M_{T,b}^2 + M_{T,c}^2 \ + 2E_bE_c(1 - \tfrac{p_{z,b}p_{z,c}}{E_bE_c} \cos\phi_{b,c})} \tag{43}$$

In the special case of $W \to l\nu$ which is more relevant to this analysis, we can ignore the mass of the lepton (electron or muon) and the neutrino[27] at the energy level of LHC. Therefore, $p_{z,l} \approx E_l$ and $p_{z,\upsilon} \approx E_\upsilon$. And since neutrinos escape CMS undetected, they are a source of the missing energy (E_T^{miss}). Provided that $W \to l\nu$ is the main source of E_T^{miss} in an event, $p_{z,\upsilon} \approx E_\upsilon \approx E_T^{miss}$ and we can try to reconstruct the transverse mass of the W using p_T^l and E_T^{miss}:

---

[27] According to the SM, neutrinos are massless. However, now we know that it is not true at least for two of the neutrino families.

$$M_T = \sqrt{2p_T^l \cdot E_T^{miss}\left(1 - \cos\phi_{l,\,E_T^{miss}}\right)} \tag{44}$$

The neutrino from the WH production is the only original source of missing energy in our topology. Therefore, if there is no other source of missing energy in an event, if CMS measured energy and angle flawlessly, and if the isolated lepton is identified correctly, then M_T will be equal to the transverse mass of W which should be distributed around $m_W = 80.385\,GeV$. By requiring $50\,GeV < M_T < 250\,GeV$, we suppress outliers which are mostly background and rarely signal. This requirement mostly affects QCD background.

3.4.2.6 B-tagged jets

B-jets are jets arising from the hadronization of b quarks. Some analyses, like the present one, heavily depend on b-jet identification, otherwise known as b-tagging. What helps us to distinguish b-jets from jets initiated by gluons or light flavor quarks is the long half-life of B hadrons (about 1.5 ps). This causes B hadrons to decay at a measurable distance from the primary vertex creating a secondary vertex. These secondary vertices are displaced hundreds of micrometers away from the primary vertex and therefore, the high resolution pixel detector that CMS has is capable of identifying b quarks efficiently. [157]

Many b-tagging algorithms were used during run 1 in CMS [158]. However, CSV gained popularity over time. This b-tagging method uses the following set of variables with high discriminating power and low correlation:

- The vertex category (“real”, “pseudo,” or “no vertex”);
- The flight distance significance in the transverse plane (“2D”);
- The vertex mass;

- The number of tracks at the vertex;
- The ratio of the energy carried by tracks at the vertex with respect to all tracks in the jet
- The pseudorapidities of the tracks at the vertex with respect to the jet axis;
- The 2DIP significance of the first track that raises the invariant mass above the charm threshold of 1.5GeV/c2 (tracks are ordered by decreasing IP significance and the mass of the system is recalculated after adding each track);
- The number of tracks in the jet;
- The 3D IP significances for each track in the jet.[158]

The DeepCSV approach starts from the same jet features as CSVv2 but extends the range of the maximum considered tracks per jets (up to 25) and exploits a more modern deep neural network architecture. [159]

We select events with at least two b-tagged jets identified with the DeepCVS b-tagging algorithm. This requirement especially reduces the W+jets and Z+jets backgrounds [138].

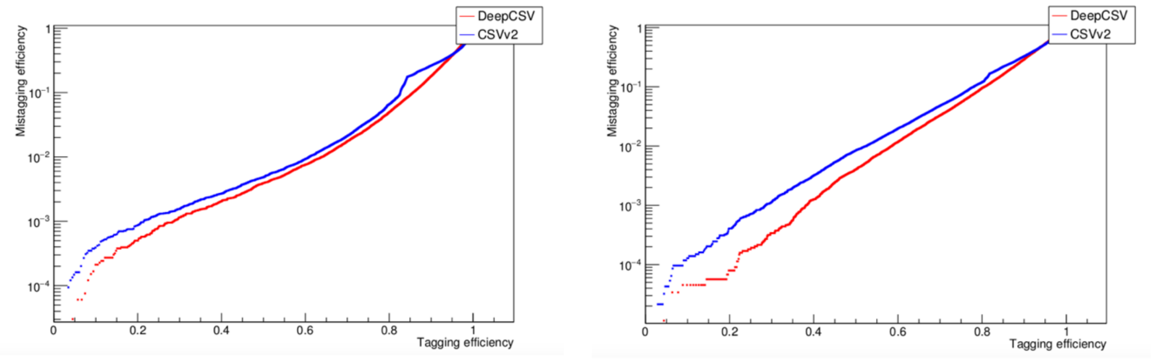


Figure 3-7: Performance of the DeepCSV b-jet identification algorithm on (left) $H \rightarrow aa \rightarrow 4b$ MC simulation with m_a = 20 GeV and (right) a $t\bar{t}$ background sample.

According to the BTV POG group [13], the b-tagging efficiency scale factors is calculated as:

$$SF = \frac{\varepsilon_{Data}}{\varepsilon_{MC}} \tag{45}$$

for b, c and light flavor jets. The scale factors in general depend on the jet flavor, jet p_T, and jet h. Usually, the b-tagging methods use machine learning techniques and they are optimized using MC simulation. However, due to inaccuracies in simulation, the efficiencies measured in data are different than those in the MC simulation. Therefore, the scale factors need to be applied to simulated events to take this difference into account. We use the period-dependent SFs from Moriond17 eras B-H. The scale factors are used to update the b-tagging status on a jet-by-jet basis. In order to upgrade or downgrade the b-tagging status of each individual jet, a random number generator is utilized. In its simplest form with a single operating point being used, the logic of this method is the following. If SF < 1, all that needs to be done is to downgrade fraction

$$f = 1 - SF \tag{46}$$

of the b-tagged jets from the “tagged” status to the “not tagged” status, and in this case, it is not necessary to know the MC b-tagging efficiency. If, on the other hand, SF > 1, it is necessary to upgrade the b-tagging status of some of the untagged jets and the fraction of such jets that needs to be upgraded is

$$f = \frac{1 - SF}{1 - 1/\varepsilon_{MC}} \tag{47}$$

This result follows from a simple requirement that $f(1 - \varepsilon_{MC}) + \varepsilon_{MC} = SF \times \varepsilon_{MC}$

3.4.2.6.1 Soft b-tagging

Soft b-tags are identified using Secondary Vertices (SVs) reconstructed with the Inclusive Vertex Finder algorithm. Soft b-tagging from SVs has been previously used and commissioned successfully in [160], in order to identify very soft bottom or charm quarks. We apply the same selection criteria in SV observables which are the following:

- The distance in the transverse plane between the SV and the PV is less than 3 cm.
- The significance of the distance, SIP3D, between the SV and the PV is greater than 4.
- The pointing angle, defined as $\cos((PV, SV), \vec{P}_{SV})$ is greater than 0.98, where $\vec{P}_{SV}$ is the total four-momentum of the tracks associated to the SV.
- The number of tracks associated to the SV is greater or equal to 3.
- The p_T of the SV is less than 20 GeV.
- The distance to any jet, $\Delta R(jet, \vec{P}_{SV})$, is greater than 0.4 to achieve the orthogonality to the jets and b-tagged jets.

Soft b-tags are selected with p_T < 20 GeV and cross-cleaned from any b-tagged jet with $\Delta R(jet, \vec{P}_{SV}), = 0.4$.

We categorize the selected events based on the number of b-tags identified either with the DeepCSV or the soft-b tagging. Only events which have at least two jets and two b-tagged jets are kept at the preselection level. The events are then split in the “three b-tag” and the “four b-tag” categories.

3.4.2.7 Event pre-selection

To summarize, the pre-selection of the events in our analysis consists of the following requirements:

1- The event has passed at least one of the online triggers summarized in Table 3-4

2- Exactly one lepton reconstructed in the tracker acceptance (i.e., < 2.5 for electrons and < 2.4 for muons) with $p_T^l > 25\ GeV$. The identification and isolation criteria of the leptons can be found in sections 3.4.2.2 and 3.4.2.3 .

3- $E_T^{miss} > 25\ GeV$ and $50\ GeV < M_T < 250\ GeV$ (sections 3.4.2.4 and 3.4.2.5)

4- The event has at least two jets and two b-tagged jets identified with the DeepCSV algorithm or soft b-tagging satisfying the looseWP (deepFlavourJetTags > 0.2219); one of the two b-tagged jets should further satisfy the MediumWP (deepFlavourJetTags > 0.6324). The requirement for the jets is considered. (sections 3.3.2.5 and 3.3.2.6)

Table 3-11: Event yields expected for background and signal after the analysis pre-selection. The events are normalized to 35.9fb^{-1} of integrated luminosity.

Process	Raw	1 lepton	Trigger	$E_T^{miss} > 25$	$50 < M_T^W < 250$	>=2-jets	$(N_b = 3, N_j = 3,4)$	$(N_b = 4, N_j = 4)$
(V)VV	$(943.1 \pm 0.6) \times 10^3$	$(702.2 \pm 0.5) \times 10^3$	$(676.5 \pm 0.5) \times 10^3$	$(547.0 \pm 0.4) \times 10^3$	$(439.7 \pm 0.4) \times 10^3$	$(251.1 \pm 0.3) \times 10^3$	736.5 ± 16.1	37.3 ± 3.8
Single Top	$(165.53 \pm 0.05) \times 10^4$	$(126.82 \pm 0.04) \times 10^4$	$(122.66 \pm 0.04) \times 10^4$	$(104.13 \pm 0.04) \times 10^4$	$(783.9 \pm 0.3) \times 10^3$	$(590.3 \pm 0.3) \times 10^3$	$(243.9 \pm 0.6) \times 10^2$	$(125.6 \pm 1.4) \times 10^1$
$t\bar{t} + b\bar{b}$	$(753.1 \pm 0.2) \times 10^4$	$(403.4 \pm 0.4) \times 10^3$	$(387.5 \pm 0.4) \times 10^3$	$(341.0 \pm 0.4) \times 10^3$	$(242.8 \pm 0.3) \times 10^3$	$(241.5 \pm 0.3) \times 10^3$	$(164.7 \pm 0.8) \times 10^2$	$(190.0 \pm 2.8) \times 10^1$
$t\bar{t} + c\bar{c}$	$(753.1 \pm 0.2) \times 10^4$	$(171.25 \pm 0.08) \times 10^4$	$(165.14 \pm 0.08) \times 10^4$	$(143.48 \pm 0.08) \times 10^4$	$(100.20 \pm 0.06) \times 10^4$	$(998.6 \pm 0.6) \times 10^3$	$(103.7 \pm 0.2) \times 10^3$	$(874.8 \pm 6.0) \times 10^1$
$t\bar{t} + light$	$(753.1 \pm 0.2) \times 10^4$	$(326.0 \pm 0.1) \times 10^4$	$(312.7 \pm 0.1) \times 10^4$	$(276.0 \pm 0.1) \times 10^4$	$(200.52 \pm 0.09) \times 10^4$	$(192.03 \pm 0.09) \times 10^4$	$(121.5 \pm 0.2) \times 10^3$	$(623.2 \pm 5.0) \times 10^1$
$t\bar{t}+$	$(644.0 \pm 2.0) \times 10^2$	$(405.7 \pm 1.7) \times 10^2$	$(386.3 \pm 1.6) \times 10^2$	$(339.0 \pm 1.5) \times 10^2$	$(240.4 \pm 1.3) \times 10^2$	$(230.1 \pm 1.2) \times 10^2$	884.9 ± 23.0	61.5 ± 5.6
$Z \to \ell\ell$	$(925.3 \pm 0.1) \times 10^5$	$(419.89 \pm 0.08) \times 10^5$	$(390.87 \pm 0.07) \times 10^5$	$(215.46 \pm 0.05) \times 10^5$	$(144.06 \pm 0.04) \times 10^5$	$(330.3 \pm 0.2) \times 10^4$	$(127.7 \pm 0.8) \times 10^2$	538.4 ± 16.8
$W \to \ell\nu$	$(498.90 \pm 0.10) \times 10^6$	$(402.47 \pm 0.09) \times 10^6$	$(390.98 \pm 0.09) \times 10^6$	$(312.21 \pm 0.08) \times 10^6$	$(291.73 \pm 0.08) \times 10^6$	$(278.5 \pm 0.1) \times 10^5$	$(362.4 \pm 4.1) \times 10^2$	$(146.1 \pm 7.8) \times 10^1$
QCD (MC)	$(468.6 \pm 1.1) \times 10^6$	$(607.4 \pm 4.1) \times 10^5$	$(580.1 \pm 4.1) \times 10^5$	$(306.2 \pm 2.8) \times 10^5$	$(100.7 \pm 1.7) \times 10^5$	$(320.2 \pm 5.8) \times 10^4$	$(495.0 \pm 73.1) \times 10^2$	827.8 ± 375.6
Total expected	$(108.5 \pm 0.1) \times 10^7$	$(512.6 \pm 0.4) \times 10^6$	$(495.2 \pm 0.4) \times 10^6$	$(370.5 \pm 0.3) \times 10^6$	$(320.7 \pm 0.2) \times 10^6$	$(383.8 \pm 0.6) \times 10^5$	$(366.2 \pm 7.3) \times 10^3$	$(210.6 \pm 3.9) \times 10^2$
Wh (20)	$(260.7 \pm 0.6) \times 10^2$	$(215.4 \pm 0.5) \times 10^2$	$(207.7 \pm 0.5) \times 10^2$	$(172.3 \pm 0.5) \times 10^2$	$(124.9 \pm 0.4) \times 10^2$	$(935.0 \pm 3.6) \times 10^1$	$(147.0 \pm 1.4) \times 10^1$	113.7 ± 4.0
Wh (60)	$(259.6 \pm 0.6) \times 10^2$	$(214.9 \pm 0.5) \times 10^2$	$(207.3 \pm 0.5) \times 10^2$	$(171.3 \pm 0.5) \times 10^2$	$(124.3 \pm 0.4) \times 10^2$	$(971.9 \pm 3.5) \times 10^1$	$(209.8 \pm 1.7) \times 10^1$	261.6 ± 5.8

A baseline analysis code is developed to implement the pre-selection as described above and the analysis that will be described below. The code can be found in [16].

3.5 Binned BDT shape analysis

3.5.1 Introduction

We can use machine learning methods in experimental high energy physics to help us distinguish between signal and background. This is done by training the algorithms using MC simulation (where we know which event is signal and which one is background) and extrapolating the result to the data taken at LHC. In high energy physics, the two popular machine learning methods are boosted decision tree (BDT) and artificial neural network (ANN). In this analysis, we decided to use BDT.

Decision tree is a sequence of yes/no questions. An example is Figure 3-8 showing the probability that a person plays video games on a regular basis based on age and gender. At each node of a decision tree, the corresponding variable is compared to a cutoff value, and depending on the result of the comparison, the decision “yes” or “no” is made, and the event is passed to the corresponding next node, which performs the same procedure using another variable. The final output of a decision tree is the probability of an event being signal. [161]

If we use a single decision tree to distinguish between signal and background, the discriminatory power is ‘weak’ or ‘shallow’. The main idea in boosting is to use many decision trees (an ensemble or jungle) and take a weighted average over all of them. In order to do so, we have to first train the BDT, meaning:

- Reweight misclassified events (i.e. signal events on a background leaf and background events on a signal leaf)
- Build & optimize new trees with reweighted events
- Score each tree
- Average over all trees, using the tree-scores as weights [161, 162]

There are several algorithms that train a BDT sample. In this analysis, we used the AdaBoost algorithm in the Toolkit for Multivariable Analysis (TMVA). [163, 164]

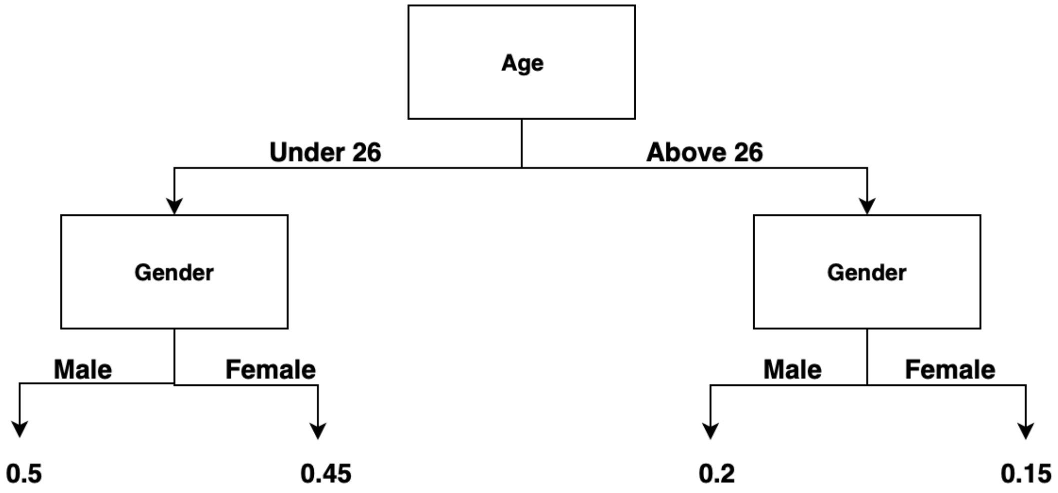


Figure 3-8: A made up example of a decision tree that gives the probability that a person plays computer games on a regular basis based on that person's age and gender

TMVA calculates a quantity of every node known as the separation to build and weigh decision trees. There are five ways to define the separation within TMVA [163, 165]:

- Gini index, defined as $p(1-p)$
- Gini index with Laplace's correction, defined as $\frac{(SB+S+B+1)}{(S+B+2)^2}$

- Cross entropy, defined as $-2p\ \log_2 p - (1-p)\log_2(1-p)$
- Misclassification error, defined by $1 - \max(p, 1-p)$
- Statistical significance, defined by $\frac{S}{S+B}$

Where S (B) is the number of signal (background) events in the node, and p is the purity of the node defined as $\frac{S}{S+B}$. In this analysis, we used the default type of separation which is the Gini index.

Another parameter is a measure of the importance of each discriminating variable. The absolute importance of a BDT-classifier X is defined as:

$$\text{Absolute importance of variable X} \quad = \sum_{m=1}^{M} \ln(\omega_m)\, I_{m,X} \qquad (48)$$

Where the sum is over the whole ensemble of trees (M being the total number of trees), ω_m is the weight of each tree, and $I_{m,X}$ is:

$$I_{m,X} = \sum_{i=1}^{n} \delta_i N_i \qquad (49)$$

Where the sum is over all nodes in a decision tree (n being the number of nodes in tree m), δ_i is the separation for each node and N_i is the number of events in that node.

Finally, importance is defined as the absolute importance of a variable divided by the sum of all absolute importances[28]. As one can tell from the name, importance shows how helpful a variable is to the BDT classifier and discrimination between signal and background. Variables with low importance can be removed for simplicity. [161, 165]

28- Importance is shown as percentage in this dissertation.

3.5.2 BDT input variables

BDT is a powerful tool for classification, however, it is blind. In other words, BDT performs significantly better when physicists cherry-pick training variables that have high discriminating power between signal and background [166].

The current analysis is done for eight different ‘a’ masses from 12 GeV to 60 GeV. In a perfect world, we would need to do this analysis eight times separately for each a mass. However, we prefer a single analysis that can be used for the whole 12-60 GeV range. Therefore, all MC samples (pseudo-scalar Higgs mass of 12-60 GeV) are added up and fed to TMVA. On the other hand, all background MC samples are also added up. Namely, $t\bar{t}$, W/DY + jets, diboson/triboson, single top, and QCD.

The variables that show high discriminating power between signal and background are listed in Table 3-12 and their shapes for signal (a mass = 60 GeV) and background are shown in Figure 3-9 and Figure 3-10. Note that $\Delta \mathrm{m}_{\mathrm{bb}}^{\mathrm{min}}$ is only defined in the 4b-tag category and is excluded in 3b-tag one.

Table 3-12: Input variables of the BDT training

	BDT Input Variable	Description
1	$\Delta R(b\bar{b})^{ave}$	Average ΔR (equation 34) of the two $b\bar{b}$ pairs from scalar Higgs decay
2	$\lvert\Delta\phi(j, E_T^{miss})\rvert^{min}$	Minimum of the phi difference of each jet and E_T^{miss} (equation 39)
3	$M_T(W)$	Transverse mass of the W (Equation 44)
4	$\lvert\Delta\phi(W, h)\rvert$	Absolute value of the phi angle difference between W and SM Higgs
5	$m_{bbb(b)}^{h}$	Reconstructed Higgs mass from 4b or 3b jets (see section 3.4.2.5
6	E_T^{miss}	Equation 39
7	p_T^{lep}	Transverse momentum of the isolated lepton
8	$p_T(h)$	Transverse momentum of the SM Higgs
9	$p_T(W)$	Transverse momentum of the W
10	H_T	$H_T = \sum_{jets} p_T$ for jets passing the selection
11	Δm_{bb}^{min}	Minimum difference among all reconstructed scalar Higgs masses from all permutations of b-jet pairs

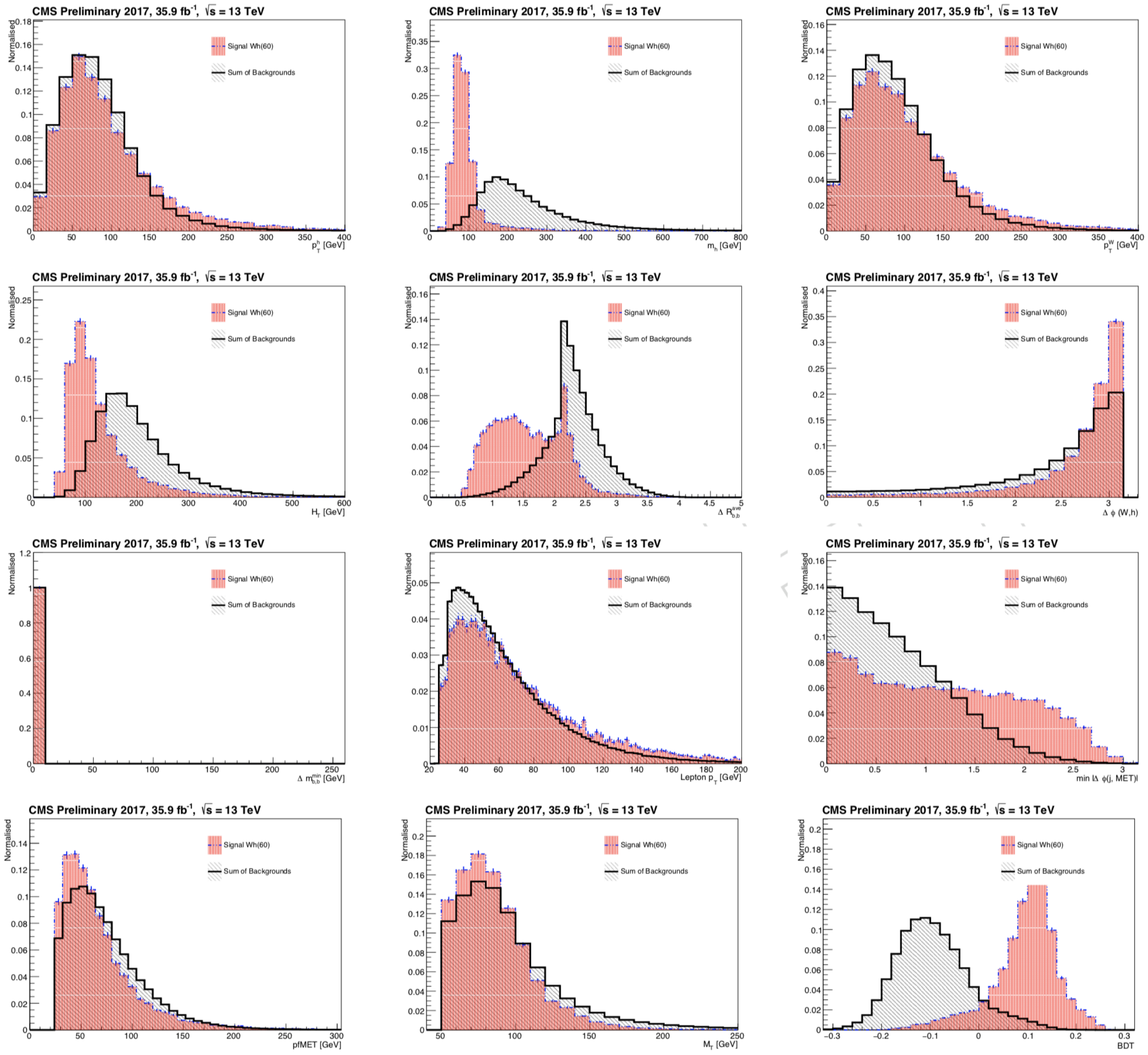


Figure 3-9: Distribution of BDT input variables for the 3b-tag category for background (black line with gray area) and signal with 60 GeV pseudo-scalar Higgs mass (blue line with red area). Note that Δm_{bb}^{min} *is not defined in the 3b-tag category.*

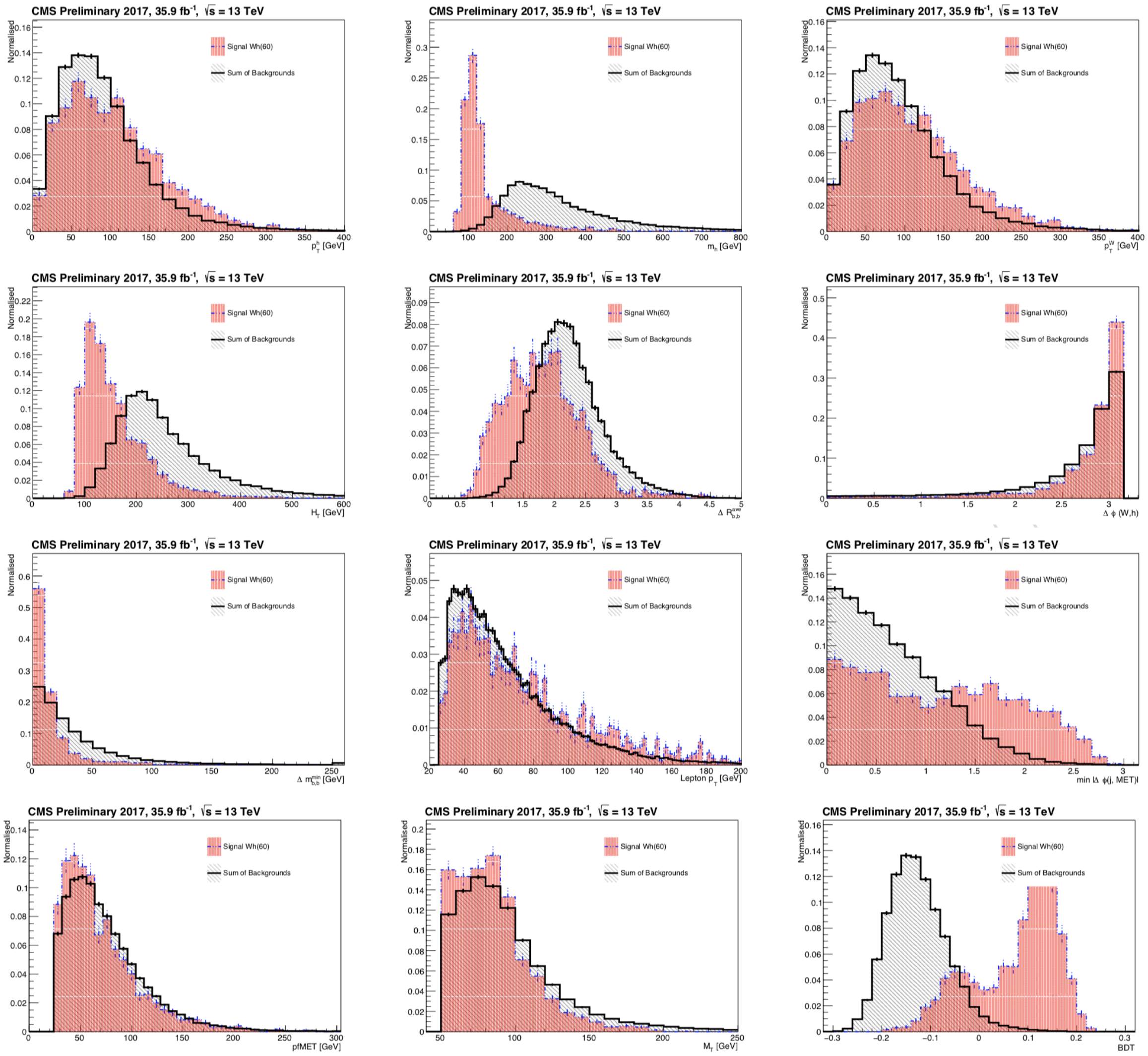


Figure 3-10: *Distribution of BDT input variables for the 4b-tag category for background (black line with gray area) and signal with 60 GeV pseudo-scalar Higgs mass (blue line with red area)*

Using these variables, the BDT is trained for 3b-tag and 4b-tag categories and the importance of each variable is shown in Table 3-13

Table 3-13: Importance of discriminating variables after BDT training for 3b-tag category (left) and 4b-tag category (right)

	BDT Input Variable	Importance (%)
1	$\Delta R(b\bar{b})^{ave}$	12.9
2	$\lvert\Delta\phi(W,h)\rvert$	12.0
3	$\lvert\Delta\phi(j,E_T^{miss})\rvert^{min}$	11.6
4	$M_T(W)$	11.4
5	m_{bbb}^{h}	9.6
6	$p_T(h)$	9.6
7	$p_T(W)$	8.4
8	E_T^{miss}	8.3
9	p_T^{lep}	8.1
10	H_T	8.0

	BDT Input Variable	Importance (%)
1	$\Delta R(b\bar{b})^{ave}$	11.0
2	$\lvert\Delta\phi(j,E_T^{miss})\rvert^{min}$	10.9
3	$M_T(W)$	10.5
4	$\lvert\Delta\phi(W,h)\rvert$	10.0
5	m_{bbbb}^{h}	9.6
6	E_T^{miss}	8.7
7	p_T^{lep}	8.5
8	$p_T(h)$	8.2
9	$p_T(W)$	7.8
10	H_T	7.6
11	Δm_{bb}^{min}	6.9

3.5.3 Overtraining check and efficiency plots

In general, a machine learning algorithm is overstrained if it tries to exploit artifacts of the training samples (Artifacts always exist due to the finite size of training samples). An overstrained algorithm works extraordinarily good on the original training sample that it has picked up its fluctuations, however, the fluctuations are not representative of the actually expected distributions and the algorithm underperforms on other similar samples [161]. In order to check for overtraining, we also used test samples in addition to training samples. If the performance of an algorithm is significantly better for the training sample compared to the test sample, then overtraining has occurred. [163, 164] We see a very good agreement between training and test samples and therefore no sign of overtraining as shown in Figure 3-12, Figure 3-13, and Figure 3-14.

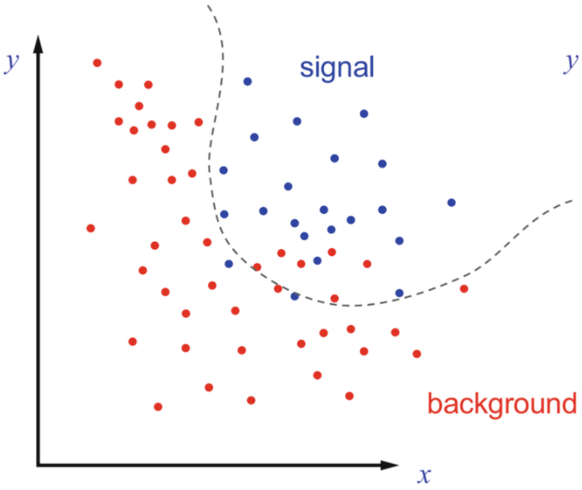


Figure 3-11: Illustrating of overtraining in two dimensions. Left: smooth selection of a regular algorithm output; right: selection based on an overtrained algorithm that picks fluctuations of the finite-size training samples. This image is taken from [161]

The output of the BDT algorithm is the probability of an event being signal. Figure 3-12 shows the distribution of the BDT output for both training and test samples.

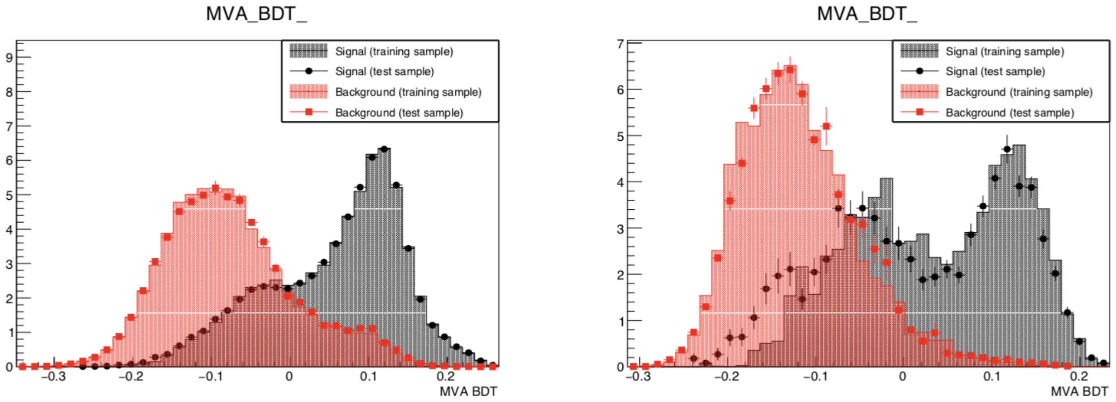


Figure 3-12: Distribution of BDT output for signal (black) and background (red). Superimposed on the training samples (histograms) are test samples (error bars) for 3b-jet category (left) and 4b-jet category (right)

A very basic question that is not answered yet is how well the BDT classifier algorithm is performing. To answer this question, we employ two techniques: Cross-validation (CV) and Receiver Operating Characteristics (ROC).

The ROC curve shows the famous type-I (rejecting the hypothesis H_0 if H_0 is true) and type-II (erroneously failing to reject H_0, if H_0 is false) errors. It is drawn by varying the cut on BDT output and measuring the signal efficiency and background rejection (Figure 3-13) [161]. Where:

$$\text{signal efficiency} = \frac{number\ of\ correctly\ identified\ signal\ events}{Total\ number\ of\ events}$$

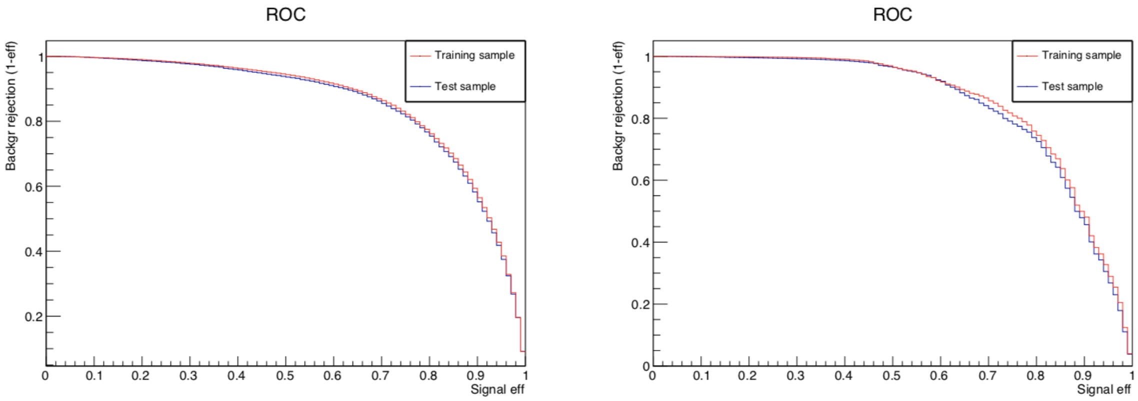


Figure 3-13: background rejection vs signal efficiency (ROC curve) for various cut values on the BDT classifier output for 3b-jet category (left) and 4b-jet category (right)

The closer the ROC curve is to the upper right corner; the more background events are rejected while signal events being correctly recognized. Meaning the better the algorithm is. Therefore, the area under the ROC curve can be used as a measure of how well the algorithm is performing [167].

Cross-validation in its basic form consists of dividing the data sample into n different subsets of roughly equal size. Then training BDT on n-1 subsets and using the last one as the test sample for verification. Then retaining the evaluation score and discarding the model. This is done n times and every time a different subset is chosen to be the test sample. And finally, an average is taken over the evaluation scores and the standard deviation is calculated [167]. Figure 3-14 shows the result of the cross-validation technique on our BDT model. The evaluation score used is the area under the ROC curve (Figure 3-13). Therefore, y-axis is score of our BDT performance (the area under the ROC curve) and x-axis is the number of fold (how many times we divided the sample into subsets). The red line shows the training samples and the blue line is the test subset.

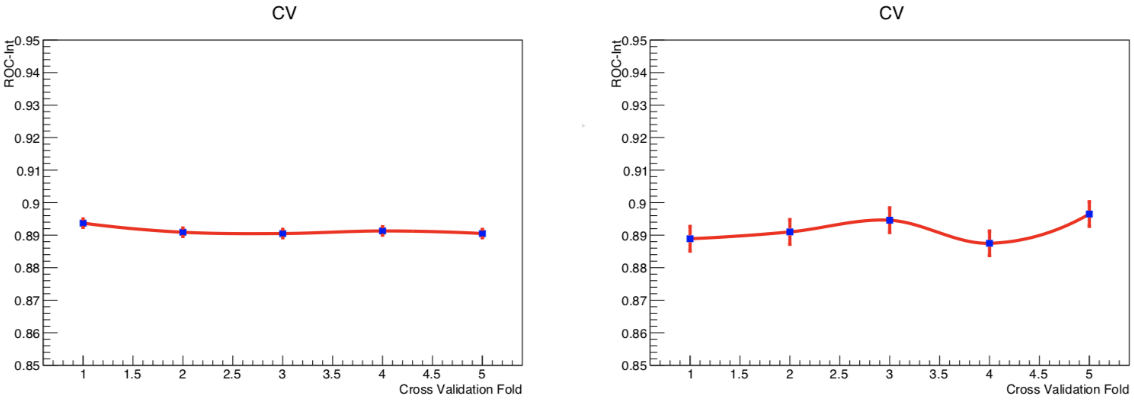


Figure 3-14: The result of the cross-validation technique on BDT classifier for 3b-jet category (left) and 4b-jet category (right). The red line shows the training samples and the blue line is the test subset.

In Figure 3-14 the average score of training samples is equal to the score of the test sample which clearly shows that the classifier is not overtrained. Also, the short uncertainty bars show consistency of the classifier performance. The right plot has a little more uncertainty which comes from having a smaller sample size (there are fewer events passing the 4b-jet requirement than 3b-jet requirement).

Looking back at Figure 3-12, most of the SM background is below BDT value of ~0.2. Therefore, applying a BDT cut around 0.2 looks interesting. We performed a sensitivity scan in the a-boson mass parameter space (12-60 GeV) with BDT cut values of cut 0-7: $0.18, 0.19, 0.20, 0.21, 0.22, 0.23, 0.24, 0.25,$ and 0.26.

In Figure 3-15 and Figure 3-16:

- The two upper plots show the number of events in signal (left) and background (right) after applying each BDT cut.
- The two middle plots show the statistical uncertainties
- And the two bottom plots show signal significance ($\frac{S}{\sqrt{S+B}}$) (left) and signal significance with 10% and 20% systematic uncertainties for signal and background respectively (right).

As expected, the signal significance rises with BDT. At low a masses (m_a < 30 GeV), the a's become boosted and cause collimated b-pairs. That in-turn causes lower signal significance for lower a-boson masses as seen in Figure 3-15 and Figure 3-16.

We define control regions for the major backgrounds of this analysis ($t\bar{t} + Jets$ and W + Jets) in order to validate the modeling of the input shapes used to train the BDT. Background control regions as well as the procedure to commission the BDT output, are described in the next section.

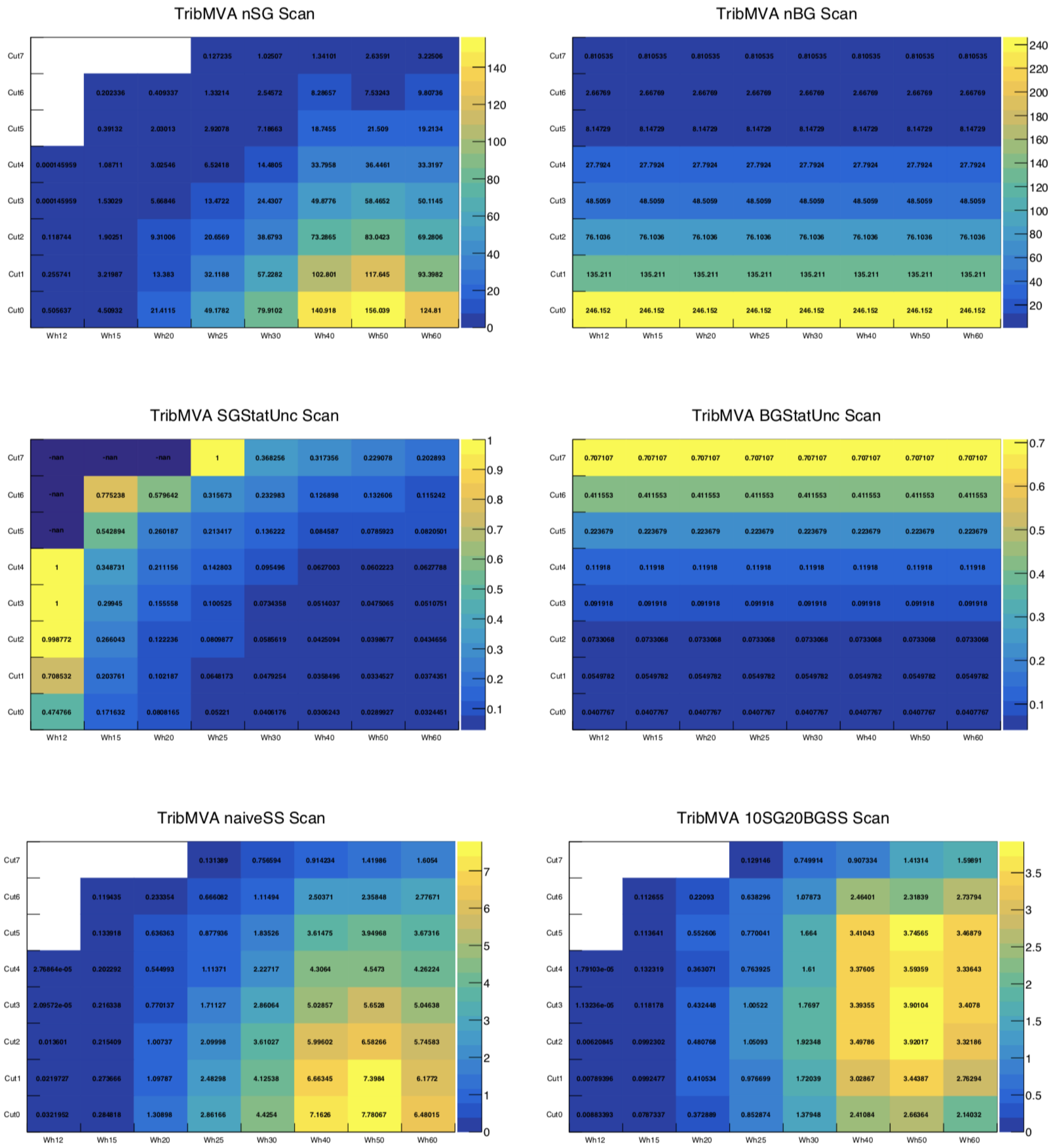
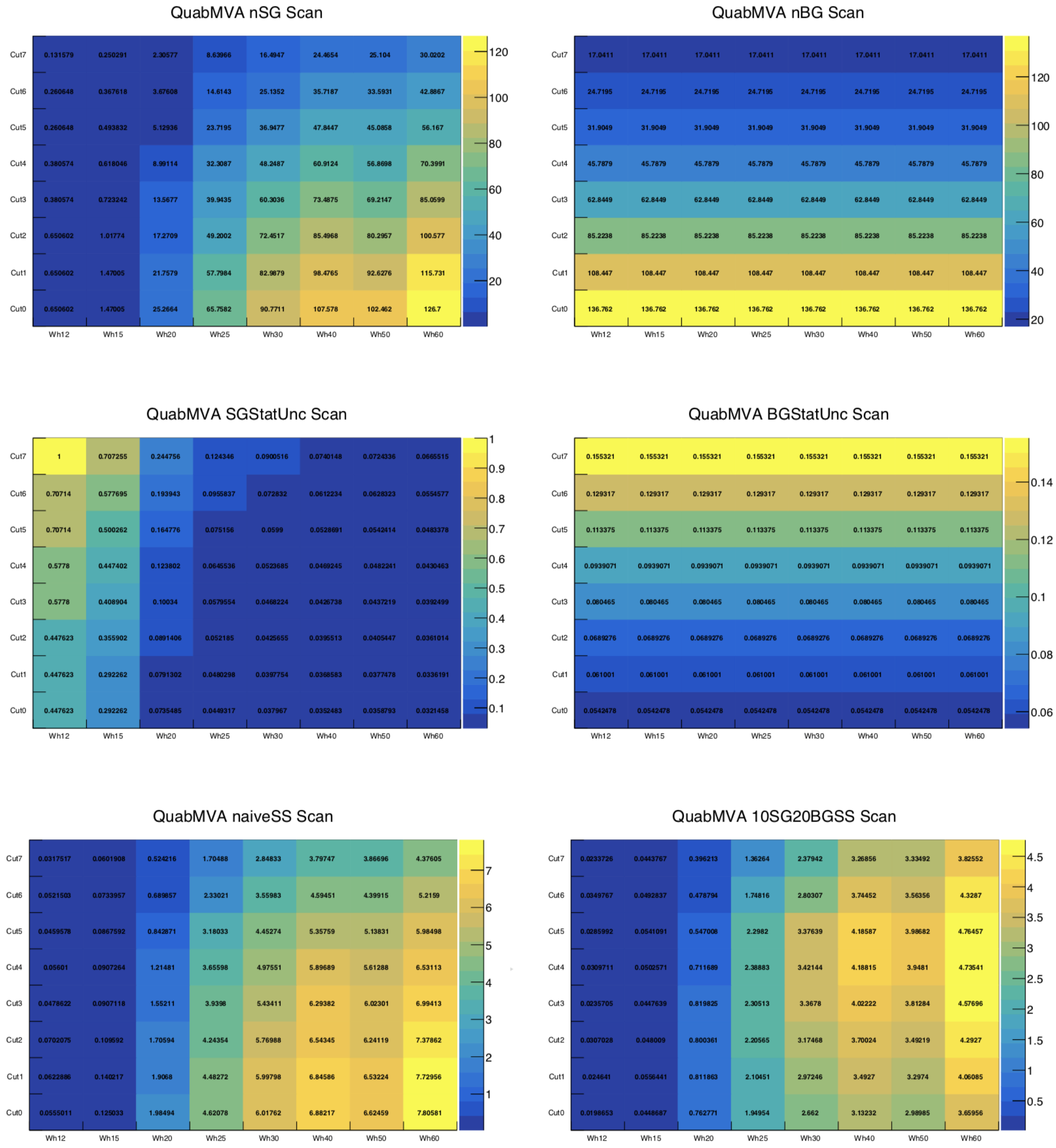


Figure 3-15: Sensitivity scan over the full a-boson mass space using the MVA BDT discriminant shown as a function of the BDT cut index (Cut0-7:0.18, 0.19, 0.20, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26). Upper row shows the number of events for signal (left) and background (right). Middle plots show the statistical uncertainty on the same events. Bottom plots show the signal significance as a function of the BDT cut index (Y-axis), calculated as $\frac{S}{\sqrt{S+B}}$ *(left) and assuming a 10% systematic uncertainty on the signal and a 20% syst. uncertainty on the background (right). (3b-category)*



QuabMVA nSG Scan

	Wh12	Wh15	Wh20	Wh25	Wh30	Wh40	Wh50	Wh60
Cut7	0.131579	0.250291	2.30577	8.63966	16.4947	24.4654	25.104	30.0202
Cut6	0.260648	0.367618	3.67608	14.6143	25.1352	35.7187	33.5931	42.8867
Cut5	0.260648	0.493832	5.12936	23.7195	36.9477	47.8447	45.0858	56.167
Cut4	0.380574	0.618046	8.99114	32.3087	48.2487	60.9124	56.8698	70.3991
Cut3	0.380574	0.723242	13.5677	39.9435	60.3036	73.4875	69.2147	85.0599
Cut2	0.650602	1.01774	17.2709	49.2002	72.4517	85.4968	80.2957	100.577
Cut1	0.650602	1.47005	21.7579	57.7984	82.9879	98.4765	92.6276	115.731
Cut0	0.650602	1.47005	25.2664	65.7582	90.7711	107.578	102.462	126.7

QuabMVA nBG Scan

	Wh12	Wh15	Wh20	Wh25	Wh30	Wh40	Wh50	Wh60
Cut7	17.0411	17.0411	17.0411	17.0411	17.0411	17.0411	17.0411	17.0411
Cut6	24.7195	24.7195	24.7195	24.7195	24.7195	24.7195	24.7195	24.7195
Cut5	31.9049	31.9049	31.9049	31.9049	31.9049	31.9049	31.9049	31.9049
Cut4	45.7879	45.7879	45.7879	45.7879	45.7879	45.7879	45.7879	45.7879
Cut3	62.8449	62.8449	62.8449	62.8449	62.8449	62.8449	62.8449	62.8449
Cut2	85.2238	85.2238	85.2238	85.2238	85.2238	85.2238	85.2238	85.2238
Cut1	108.447	108.447	108.447	108.447	108.447	108.447	108.447	108.447
Cut0	136.762	136.762	136.762	136.762	136.762	136.762	136.762	136.762

QuabMVA SGStatUnc Scan

	Wh12	Wh15	Wh20	Wh25	Wh30	Wh40	Wh50	Wh60
Cut7	1	0.707255	0.244756	0.124346	0.0900516	0.0740148	0.0724336	0.0665515
Cut6	0.70714	0.577695	0.193943	0.0955837	0.072832	0.0612234	0.0628323	0.0554577
Cut5	0.70714	0.500262	0.164776	0.075156	0.0599	0.0528691	0.0542414	0.0483378
Cut4	0.5778	0.447402	0.123802	0.0645536	0.0523685	0.0469245	0.0482241	0.0430463
Cut3	0.5778	0.408904	0.10034	0.0579554	0.0468224	0.0426738	0.0437219	0.0392499
Cut2	0.447623	0.355902	0.0891406	0.052185	0.0425655	0.0395513	0.0405447	0.0361014
Cut1	0.447623	0.292262	0.0791302	0.0480298	0.0397754	0.0368583	0.0377478	0.0336191
Cut0	0.447623	0.292262	0.0735485	0.0449317	0.037967	0.0352483	0.0358793	0.0321458

QuabMVA BGStatUnc Scan

	Wh12	Wh15	Wh20	Wh25	Wh30	Wh40	Wh50	Wh60
Cut7	0.155321	0.155321	0.155321	0.155321	0.155321	0.155321	0.155321	0.155321
Cut6	0.129317	0.129317	0.129317	0.129317	0.129317	0.129317	0.129317	0.129317
Cut5	0.113375	0.113375	0.113375	0.113375	0.113375	0.113375	0.113375	0.113375
Cut4	0.0939071	0.0939071	0.0939071	0.0939071	0.0939071	0.0939071	0.0939071	0.0939071
Cut3	0.080465	0.080465	0.080465	0.080465	0.080465	0.080465	0.080465	0.080465
Cut2	0.0689276	0.0689276	0.0689276	0.0689276	0.0689276	0.0689276	0.0689276	0.0689276
Cut1	0.061001	0.061001	0.061001	0.061001	0.061001	0.061001	0.061001	0.061001
Cut0	0.0542478	0.0542478	0.0542478	0.0542478	0.0542478	0.0542478	0.0542478	0.0542478

QuabMVA naiveSS Scan

	Wh12	Wh15	Wh20	Wh25	Wh30	Wh40	Wh50	Wh60
Cut7	0.0317517	0.0601908	0.524216	1.70488	2.84833	3.79747	3.86696	4.37605
Cut6	0.0521503	0.0733957	0.689857	2.33021	3.55983	4.59451	4.39915	5.2159
Cut5	0.0459578	0.0867592	0.842871	3.18033	4.45274	5.35759	5.13831	5.98498
Cut4	0.05601	0.0907264	1.21481	3.65598	4.97551	5.89689	5.61288	6.53113
Cut3	0.0478622	0.0907118	1.55211	3.9398	5.43411	6.29382	6.02301	6.99413
Cut2	0.0702075	0.109592	1.70594	4.24354	5.76988	6.54345	6.24119	7.37862
Cut1	0.0622886	0.140217	1.9068	4.48272	5.99798	6.84586	6.53224	7.72956
Cut0	0.0555011	0.125033	1.98494	4.62078	6.01762	6.88217	6.62459	7.80581

QuabMVA 10SG20BGSS Scan

	Wh12	Wh15	Wh20	Wh25	Wh30	Wh40	Wh50	Wh60
Cut7	0.0233726	0.0443767	0.396213	1.36264	2.37942	3.26856	3.33492	3.82552
Cut6	0.0349767	0.0492837	0.478794	1.74816	2.80307	3.74452	3.56356	4.3287
Cut5	0.0285992	0.0541091	0.547008	2.2982	3.37639	4.18587	3.98682	4.76457
Cut4	0.0309711	0.0502571	0.711689	2.38883	3.42144	4.18815	3.9481	4.73541
Cut3	0.0235705	0.0447639	0.819825	2.30513	3.3678	4.02222	3.81284	4.57696
Cut2	0.0307028	0.048009	0.800361	2.20565	3.17468	3.70024	3.49219	4.2927
Cut1	0.024641	0.0556441	0.811863	2.10451	2.97246	3.4927	3.2974	4.06085
Cut0	0.0198653	0.0448687	0.762771	1.94954	2.662	3.13232	2.98985	3.65956

Figure 3-16: Sensitivity scan over the full a-boson mass space using the MVA BDT discriminant shown as a function of the BDT cut index (Cut0-7:0.18, 0.19, 0.20, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26). Upper row shows the number of events for signal (left) and background (right). Middle plots show the statistical uncertainty on the same events. Bottom plots show the signal significance as a function of the BDT cut index (Y-axis), calculated as $\frac{S}{\sqrt{S+B}}$ *(left) and assuming a 10% systematic uncertainty on the signal and a 20% syst. uncertainty on the background (right). (4b-category)*

3.6 Background Estimation

3.6.1 Introduction

The backgrounds considered in this analysis are $W \to l\nu$, $Z \to ll$, $t\bar{t} + X$, single top, dibosons, tribosons, and QCD (Table 3-3). Let's briefly talk about every background. However, before that, it is useful to classify backgrounds in three categories: irreducible, reducible, and instrumental. As [168] explains:

- "Irreducible backgrounds are those that, on an event by event basis, cannot be distinguished from the signal, even in presence of a perfect detector.
- Reducible backgrounds include processes that share the main features of the signal, however have in addition some extra elements that would make them in principle distinguishable from it.
- Instrumental backgrounds arise when the characteristic features of the signals are due to the inaccuracy of the detector or of the measurement. The most important example in this category is QCD multijets, namely final states with only jets".

3.6.1.1 $t\bar{t} + X$

$t\bar{t} + X$ is the most important background in this analysis because it is dominant and irreducible. Figure 3-17 shows two example Feynman diagrams in which two top-quarks are produced in a collision. Top-quarks almost always decay as $t \to Wb$ and if one or both of W's decay as $W \to l\nu$, we will have:

- One or two bottom quarks and also high jet multiplicity due to large top-quark mass. Therefore, a $t\bar{t} + X$ event can pass our requirement of at least two jets and two b-tagged jets. (section 3.4.2.) In addition, other jets from QCD radiation can be real b-jets (for example from gluon splitting $g \rightarrow b\bar{b}$) or they can be mis-identified as b-jets. Therefore, a $t\bar{t} + X$ can be in 3b-jet or 4b-jet categories as well.

- Isolated leptons. If one W decays leptonically (semileptonic channel), we will have a single isolated lepton. However, even if both W's decay as $W \rightarrow l\nu$ (dilepton channel), one of the leptons can still be in the blind spot of the trigger[29] and does not get reconstructed. Therefore, the event is categorized as a single lepton event. Hence, if one or both W's decay leptonically, the $t\bar{t} + X$ can pass the isolated lepton requirement.
- Missing energy. Neutrino(s) from the W decay(s) can create enough missing energy that the $t\bar{t} + X$ event passes the $E_T^{miss} > 25\ GeV$ requirement (section 3.4.2.). The missing energy can also come from other sources like mismeasurements of the detector.

Therefore, the $t\bar{t} + X$ is an irreducible background because even if CMS had a perfect response, some $t\bar{t} + X$ events would not still be distinguishable from the signal events.

We will define control regions for the $t\bar{t} + X$ background in the next section (the reason will be explained there). Therefore, we have to find regions of the phase space which are rich in $t\bar{t} + X$ events. We know that all $t\bar{t} + X$ events have at least one (semileptonic decay) or two (dilepton decay) real b-jets. Therefore, 1b-jet and especially 2b-jet categories are richer than 3b-jet or 4b-jet categories in $t\bar{t} + light$ background (light means not bb and

---

29- $|\eta| > 2.5$ for electrons and $|\eta| > 2.4$ for muons.

cc). However, if X in $t\bar{t} + X$ is actually bb or cc, this is not true anymore. For $t\bar{t} + heavy$ (i.e. $t\bar{t} + bb$ and $t\bar{t} + cc$), we already have two extra potential b-jets. Therefore, the areas that are rich in $\overline{tt} + heavy$ background are 3-bjet and 4-jet categories which are the same area as signal. In order to distinguish between signal and $t\bar{t} + heavy$ CR, we can use the higher jet multiplicity due to high mass of the top-quark. To summarize:

- There is rich $\overline{tt} + light$ background when number of b-tagged jets = 2 and number of jets = 3 or 4
- And there is rich $\overline{tt} + heavy$ background when number of b-tagged jets = 3 or 4 and number of jets ≥ 5

We will use these control regions for the $t\bar{t} + X$ in the next section.

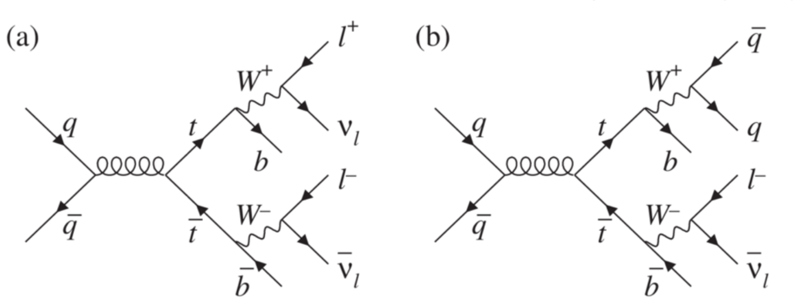


Figure 3-17: Feynman diagram of $tt \rightarrow (W^+b)(W^-\bar{b})$ in semileptonic channel (a) and diboson channel(b). This image is taken from [169].

3.6.1.2 W + Jets

The Leading order diagrams describing this process are shown in Figure 3-18 with process (b) (quark–gluon interaction) occurring predominately over processes (a) and (c) (quark–antiquark contribution) because of higher gluon luminosity in the protons. The lepton that passes the pre-selection comes from the W decay. Also, b-quarks can come from gluon splitting. Therefore, it is much more common that a W + jets event has two b-quarks rather than 3 or 4. In other words, a bin of (number of b-tagged jets = 2, number of jets = 3,4) is rich in W + jets background and can be used as a control region. We will use this information to define the W + Jets control region in the next section.

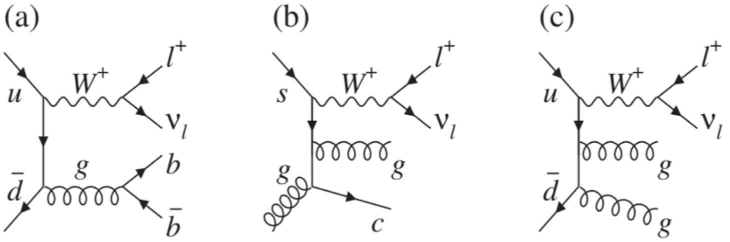


Figure 3-18: Some representative diagrams for W + jets production. This image is taken from [169]

3.6.1.3 Drell-Yan process

The DY process is an effect in which a quark-antiquark pair annihilate to create lepton pair through Z or γ [170]. If one of the leptons is missed, this event can pass the pre-selection requirements (Figure 3-19). This background is small compared to $t\bar{t} + X$ and W + Jets.

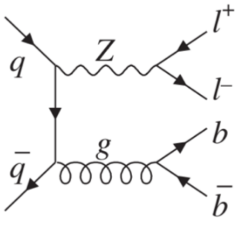


Figure 3-19: Representative Feynman diagram for the DY process. This image is taken from [169]

3.6.1.4 Single Top

Figure 3-20 shows the processes that create a single top. Single top background passes our pre-selection more or less for the same reasons that $t\bar{t} + X$ with only one t decaying leptonically can pass the pre-selection requirements. We will not discuss more about this background as it only constitutes a small portion of the background.

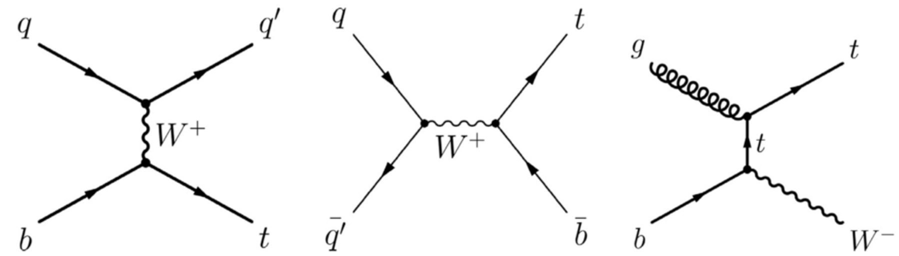


Figure 3-20: Feynman diagrams of the single top production: t-channel (left), s-channel (center), and W-associated production or tW (right) This image is taken from [169]

3.6.1.5 Dibosons and Tribosons

Some diboson (WW, ZZ/γ^*, and WZ/γ^*) and triboson (WWW, ZWW, and ZZZ) events can also pass the pre-selection and contribute to the background. They also constitute only a small portion of the background. The mechanism that they can pass the pre-selection is shown in Figure 3-21 and Figure 3-22.

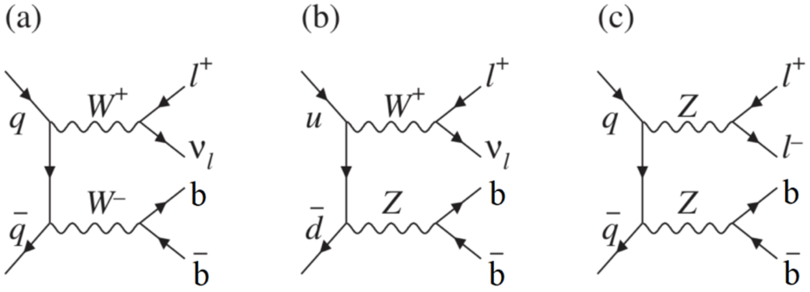


Figure 3-21: Some Feynman diagrams of the diboson production which provide a small contribution to the background. This image is taken from [171]

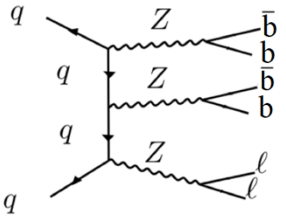


Figure 3-22: An example of a triboson production that can contribute to the background if one of the leptons is missed. This image is taken from [172]

3.6.1.6 QCD

This background is all hadronic events in which a jet is misidentified as a lepton. Therefore, QCD is an instrumental background and we also call it the fake lepton background in this analysis. The probability of misidentifying a jet as a lepton is not very high, however, QCD background is so enormous that even with a very small chance of misidentification, this background is considerable. Most requirements applied in the pre-selection also reduce this background further.

Simulating this background is very challenging. This is mainly due to the extremely large number of events in this background. The cross-section of this background is so high that it is computationally impossible to simulate all of QCD events in a dataset. Therefore, some representative events are simulated and then weighted to account for a large number of QCD background events. However, after applying the cuts and killing most of the QCD background, we do not get smooth and rich QCD distributions. This is a problem that we will address in section 3.6.4.

3.6.2 Background estimation methods

Both $t\bar{t}+X$ and W + Jets backgrounds which are the dominant backgrounds in this analysis have high jet multiplicities. This means that many QCD events should be simulated in the MC simulation of these backgrounds. As we discussed before, QCD simulation is not very accurate. Therefore, we will use control regions to have a data-driven method for $t\bar{t}+X$ and W + Jets backgrounds estimation and to not rely solely on MC simulation. This mis-modeling of the additional jets in $t\bar{t}+X$ is also seen for example in [173, 174]

The inaccuracy in $t\bar{t}+X$ simulation shows itself more when there is top + heavy flavor as there is more QCD events from heavy flavor quarks. Therefore, we separate the $t\bar{t}+X$ background to heavy flavor (HF) content ($t\bar{t}+c\bar{c}$ and $t\bar{t}+b\bar{b}$) and light flavor content (everything else). The $t\bar{t}+b\bar{b}$ background will be normalized using the data, however, the shape of $t\bar{t}+c\bar{c}$ will be derived from MC in the control region.

The smaller backgrounds of this analysis are DY + jet, diboson, tribosons, and single top processes, which will be taken from MC simulation. Finally, a data-driven method will be used to model the BDT shape of the QCD background.

3.6.3 Background Control samples

As described in the previous section, we will use control regions for the $t\bar{t} + heavy$, $t\bar{t} + light$, and W + Jets backgrounds. Therefore, we have to find (number of b-tags, number of jets) bins that are rich in either $t\bar{t} + heavy/light$ or W + Jets events after imposing pre-selection requirements (Figure 3-23).

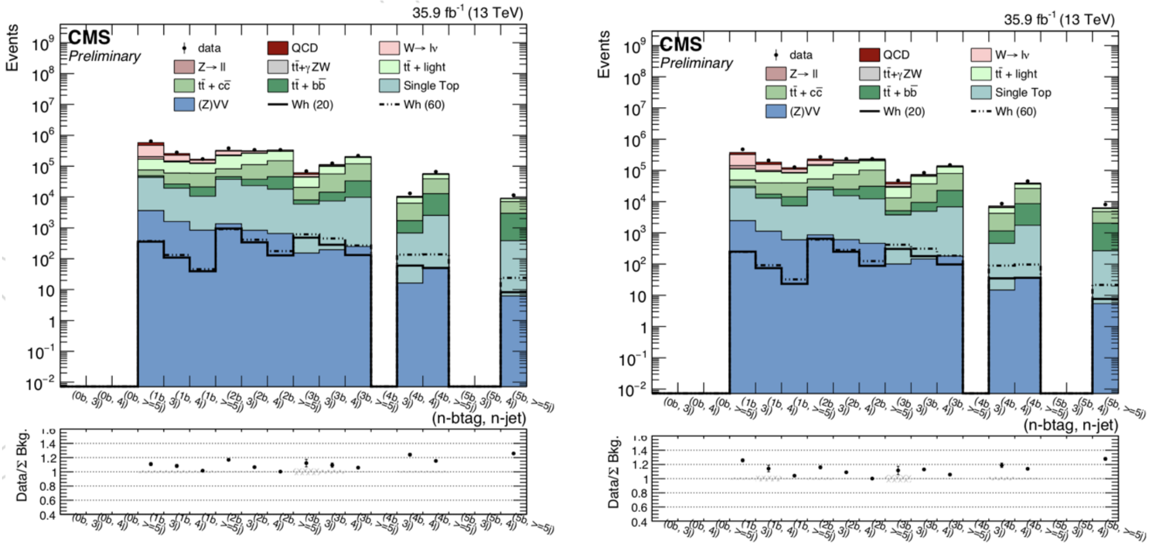


Figure 3-23: Number of events in every (number of b-tags, number of jets) bin for electrons (left) and muons (right). We define the (2b, 3j) and (2b, 4j) bins as $t\bar{t} + light$ and W + Jets CRs. Also, the (3b, 5j) and (4b, 5j) bins are $t\bar{t} + heavy$ control regions. The (3b, 3jets), (3b, 4jets) and (4b, 4jets) bins are SR (signal regions).

We want the CRs to have negligible signal content and be as close as possible to SRs to have close BDT shape. Using Figure 3-23:

- (2b, 3j) and (2b, 4j) bins are chosen to be both the W + Jets CR and $t\bar{t} + light$ CR. The two b-jets should be identified either with the DeepCSV discriminator or the soft b-tagging with at least one DeepCSV b-jet satisfying the medium DeepCSV working point as well. In other words, CRs are defined as the bin with one less b-tag multiplicity with respect to the SR. The reason for this choice is explained in section 3.6.1.1 using the $\overline{tt} + jets$ and W + jets topology. The purity of the (2b, 4j) control region for $t\bar{t} + light$ is approximately 65%. Also, the W + bb process is approximately 40% of the total background in (2b, 3j).

- (3b, ≥ 5j) and (4b, ≥ 5j) bins are defined as the $t\bar{t} + heavy$ control region. Which are the bins with the same b-tag multiplicity as in the SRs and the number of AK4 jets >= 5. The purity of $t\bar{t} + heavy$ processes in the (3b, 5j) and (4b, 5j) bins is approximately 65% and 70% respectively.

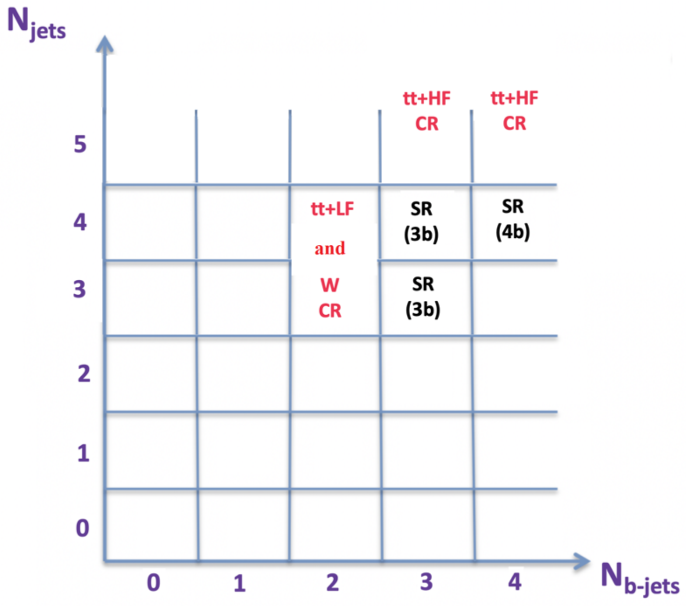


Figure 3-24: Signal and control regions in (N_{jet}, N_{bjet}) bins.

3.6.4 QCD Background Estimation

There are numerous methods for estimating the amount of the fake-lepton background originating from QCD processes, and also the shape of it as a function of variables like the transverse momentum, pseudorapidity, and in some cases more complicated variables.

Since the QCD simulation is challenging and not reliable, all of the QCD background estimation methods use “control regions”. Control regions are areas of the phase space with rich QCD statistics which have the same QCD shape as in the signal region. In this analysis, we choose the so-called “ABCD method”, where we define four regions (one signal region and three control regions) as shown in Figure 3-25.

The main idea in the ABCD method is to have two uncorrelated variables α and β and their respective cuts. Uncorrelation means that the distribution of QCD vs α doesn’t change when applying the β cut and vice versa. This will divide the phase space into four regions like Figure 3-25. By definition, the uncorrelation of α and β means:

$$\frac{N_{QCD}^{A}}{N_{QCD}^{B}} = \frac{N_{QCD}^{C}}{N_{QCD}^{D}} \tag{50}$$

Where N_{QCD}^{X} is the number of QCD events in region X. Therefore:

$$N_{QCD}^{A} = \frac{N_{QCD}^{C}}{N_{QCD}^{D}} . N_{QCD}^{B} \tag{51}$$

In other words, by knowing the number (shape) of QCD events in control regions B, C, and D, we can extrapolate the number (shape) of QCD events in region A. Region A is our signal region where we are looking for new physics. If we define a scale factor as:

$$SF_{QCD} = \frac{N_{QCD}^{C}}{N_{QCD}^{D}} \tag{52}$$

We will have:

$$N_{QCD}^{A} = SF_{QCD} . N_{QCD}^{B} \tag{53}$$

And we can use region B which is richer in QCD events to estimate the number or shape of QCD events in region A.

In practice, due to difficulties in QCD background simulation, we like to calculate the scale factor from data rather than MC, therefore, we change formula 52 and use (Data) – (non - QCD background estimation of data) instead of QCD MC.

$$SF_{QCD} = \frac{N^{C}_{Data-non\_QCD}}{N^{D}_{Data-non\_QCD}} \tag{54}$$

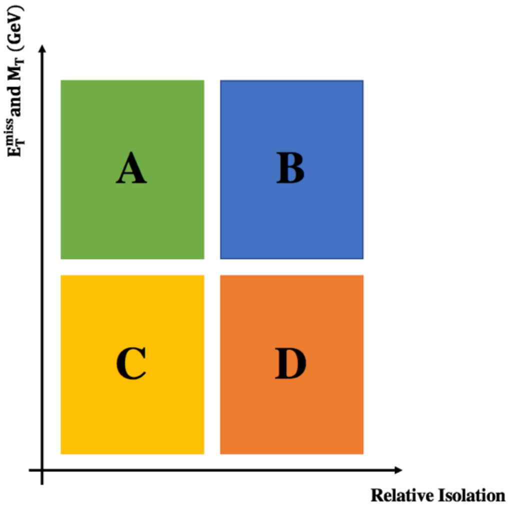


Figure 3-25: Schematic description of the signal and control regions in the ABCD method. The cuts distinguishing the regions are not easy to show as described in the text.

In this analysis, we choose lepton relative isolation and E_T^{miss} as uncorrelated variables as their uncorrelation has been shown before. More precisely speaking, we combine E_T^{miss} and M_T in one cut ($E_T^{miss} > 25$ GeV, $M_T > 50\ GeV$) and use the combined cut vs lepton relative isolation cut. Therefore, the four regions are:

- Region A: relIso < relIso cut, $E_T^{miss} > 25$ GeV, $M_T > 50$ GeV
- Region B: relIso > relIso cut, $E_T^{miss} > 25$ GeV, $M_T > 50$ GeV
- Region C: relIso < relIso cut, $E_T^{miss} < 25$ GeV, $M_T < 50$ GeV
- Region D: relIso > relIso cut, $E_T^{miss} < 25$ GeV, $M_T < 50$ GeV

Where the relIso cut values are 0.1 for Muons, 0.0588 for electrons in the barrel, and 0.0571 for electrons in the endcap regions. These cuts are recommendations of [148]

In order to have a better estimation of the QCD background, in this section, we used the whole QCD background datasets as tabulated in Table 3-14. However, even after using the whole QCD MC datasets, we still suffer from the lack of enough QCD MC statistics. Therefore, we look at the QCD MC shapes at four levels of event selection: right after single lepton cut, after 2-jet and 2-bjet cut, after 3b-tag cut, and finally after 4b-tag cut.

Table 3-14: List of the SM QCD MC samples used to perform the MC closure tests in QCD background estimation section. The last column shows the total cross-section of the MC sample. S16 is used as an abbreviation for RunIISummer16MiniAODv2- PUMoriond17 80X mcRun2 asymptotic 2016 TrancheIV v6(ext)-v1.*

Type	Dataset	$\sigma.k(pb)$
EMEnriched	/QCD Pt-20to30_EMEnriched_TuneCUETP8M1_13TeV_pythia8/S16	5352960
	/QCD Pt-30to50_EMEnriched_TuneCUETP8M1_13TeV_pythia8/S16	9928000
	/QCD Pt-50to80_EMEnriched_TuneCUETP8M1_13TeV_pythia8/S16	2890800
	/QCD Pt-80to120_EMEnriched_TuneCUETP8M1_13TeV_pythia8/S16	350000
	/QCD Pt-120to170_EMEnriched_TuneCUETP8M1_13TeV_pythia8/S16	62964
	/QCD Pt-170to300_EMEnriched_TuneCUETP8M1_13TeV_pythia8/S16	18810
	/QCD Pt-300toInf_EMEnriched_TuneCUETP8M1_13TeV_pythia8/S16	1350
MuEnrichedPt5	/QCD Pt-15to20_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/S16	3819570
	/QCD Pt-20to30_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/S16	2960198.4
	/QCD Pt-30to50_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/S16	1652471.46
	/QCD Pt-50to80_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/S16	437504.1
	/QCD Pt-80to120_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/S16	106033.6648
	/QCD Pt-120to170_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/S16	25190.51514
	/QCD Pt-170to300_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/S16	8654.49315
	/QCD Pt-300to470_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/S16	797.35269
	/QCD Pt-470to600_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/S16	79.02553776
	/QCD Pt-600to800_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/S16	25.09505908
	/QCD Pt-800to1000_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/S16	4.707368272
	/QCD Pt-1000toInf_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/S16	1.62131692
bcToE QCD	/QCD Pt 15to20 bcToE_TuneCUETP8M1_13TeV_pythia8/S16	1272980000
	/QCD Pt 20to30 bcToE_TuneCUETP8M1_13TeV_pythia8/S16	557627000
	/QCD Pt 30to80 bcToE_TuneCUETP8M1_13TeV_pythia8/S16	159068000
	/QCD Pt 80to170 bcToE_TuneCUETP8M1_13TeV_pythia8/S1	3221000
	/QCD Pt 170to250 bcToE_TuneCUETP8M1_13TeV_pythia8/S16	105771
	/QCD Pt 250toInf bcToE_TuneCUETP8M1_13TeV_pythia8/S16	21094.1

3.6.4.1 Validation of similarity of QCD shapes in ABCD regions

As explained in the last part, a critical step in the ABCD method is the demonstration that the inaccuracy in extrapolation from the control regions to the signal region is small and can be estimated reliably. It starts by showing the uncorrelation of the chosen parameters which in our case are: 1- lepton relative isolation and 2- missing energy-transverse mass. Figure 3-26 to Figure 3-31 show this correlation.

In all the of these figures, we have a very good ratio of the red line to the blue line in 1lep plot. Which clearly shows the uncorrelation of relative isolation and E_T^{miss}- MT cut. However, as we apply more cuts on data, the ratio deteriorates. This is coming from having a poor QCD MC statistics because we also see that the uncertainties get bigger. We do not expect that applying cuts on the number of b-jets in an event has any effect on the uncorrelation of relative isolation and E_T^{miss}- MT. Therefore, it is enough to show the uncorrelation in 1lep plot and extrapolate the result to 3b-jet and 4b-jet categories.

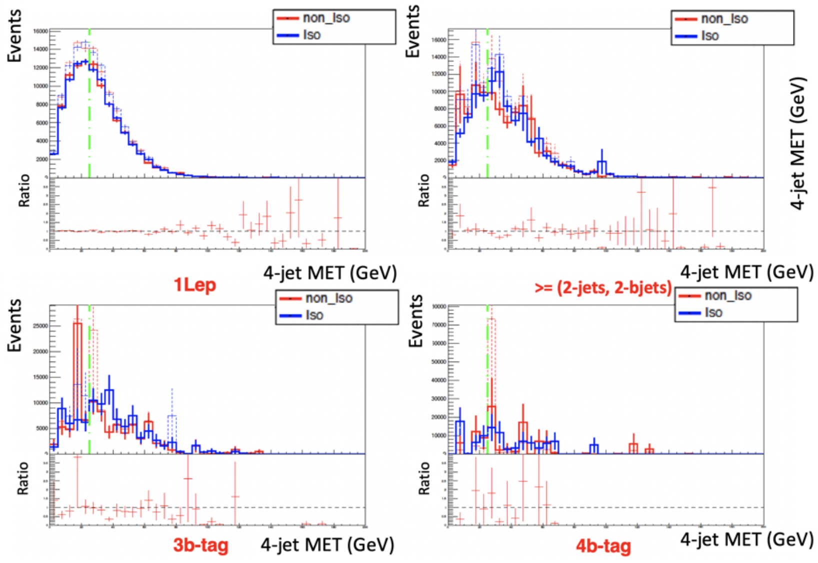


Figure 3-26: Comparison of MET distributions between non-isolated electrons and isolated electrons in Barrel with only 1 lepton cut (top left), 2-jet 2-bjet cut (top right), 3b cut (bottom left), and 4b cut (bottom right). The solid lines are after excluding all QCD datasets with $p_T < 80\ GeV$. The dashed lines show the plots if we kept all datasets. The ratio is calculated excluding datasets with $p_T < 80\ GeV$.

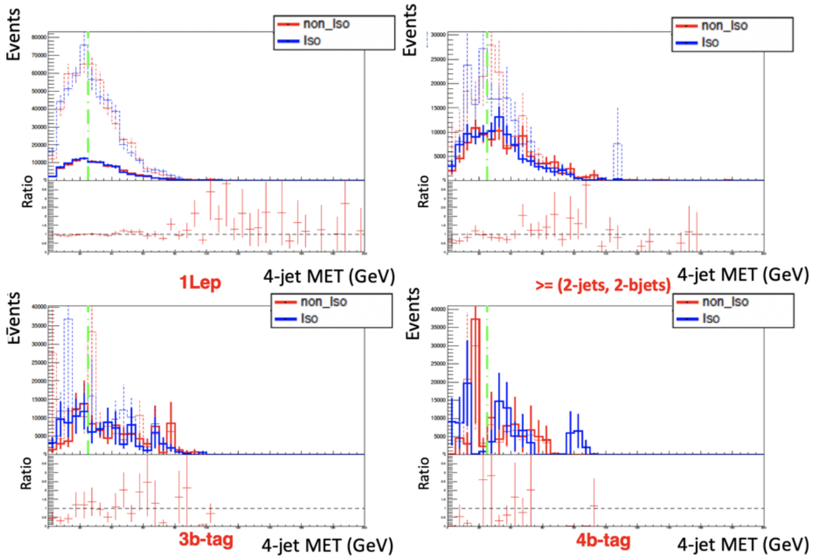


Figure 3-27: Comparison of MET distributions between non-isolated electrons and isolated electrons in Endcap with only 1 lepton cut (top left), 2-jet 2-bjet cut (top right), 3b cut (bottom left), and 4b cut (bottom right). The solid lines are after excluding all QCD datasets with $p_T < 80\ GeV$. The dashed lines show the plots if we kept all datasets. The ratio is calculated excluding datasets with $p_T < 80\ GeV$..

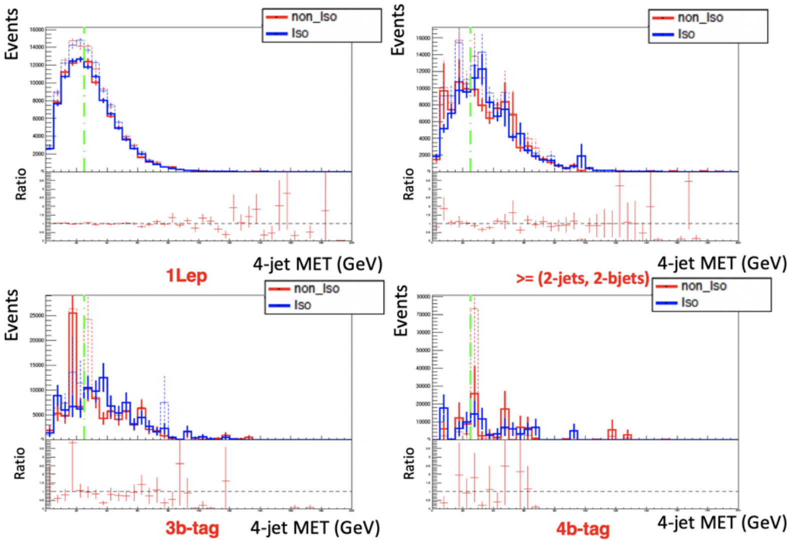


Figure 3-28: Comparison of MET distributions between non-isolated muons and isolated muons with only 1 lepton cut (top left), 2-jet 2-bjet cut (top right), 3b cut (bottom left), and 4b cut (bottom right). The solid lines are after excluding all QCD datasets with $p_T < 80\ GeV$. The dashed lines show the plots if we kept all datasets. The ratio is calculated excluding datasets with $p_T < 80\ GeV$.

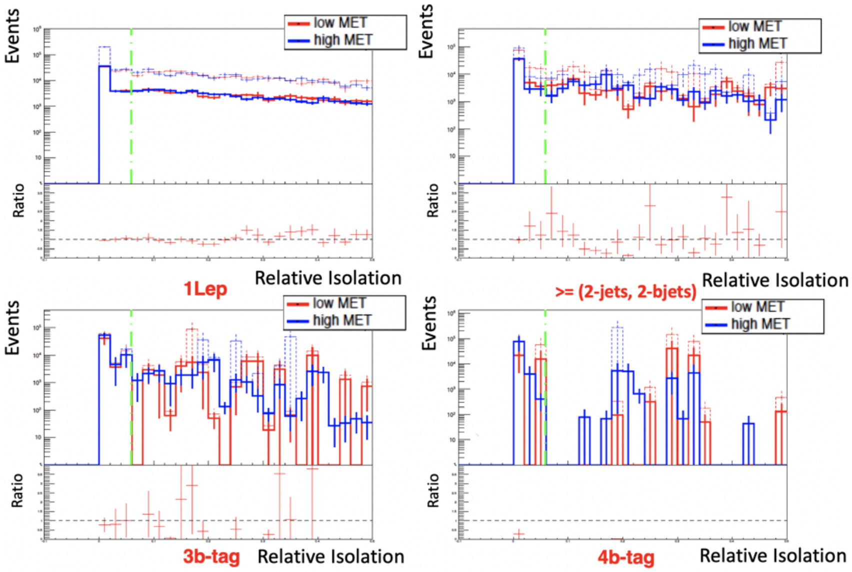


Figure 3-29: Comparison of Relative Isolation distributions between low-MET electrons and high-MET electrons in Barrel with only 1 lepton cut (top left), 2-jet 2-bjet cut (top right), 3b cut (bottom left), and 4b cut (bottom right). The solid lines are after excluding all QCD datasets with $p_T < 80\ GeV$. The dashed lines show the plots if we kept all datasets. The ratio is calculated excluding datasets with $p_T < 80\ GeV$.

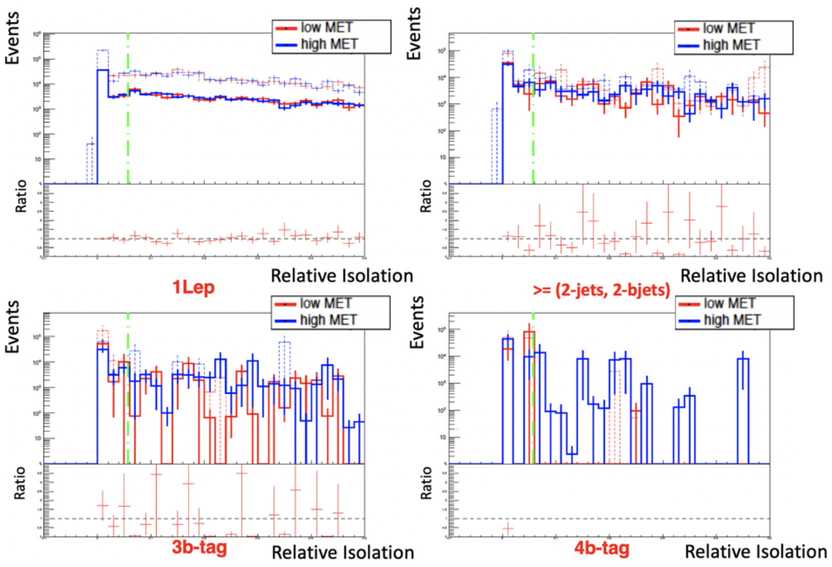


Figure 3-30: Comparison of Relative Isolation distributions between low-MET electrons and high-MET electrons in Endcap with only 1 lepton cut (top left), 2-jet 2-bjet cut (top right), 3b cut (bottom left), and 4b cut (bottom right). The solid lines are after excluding all QCD datasets with $p_T < 80\ GeV$*. The dashed lines show the plots if we kept all datasets. The ratio is calculated excluding datasets with* $p_T < 80\ GeV$*.*

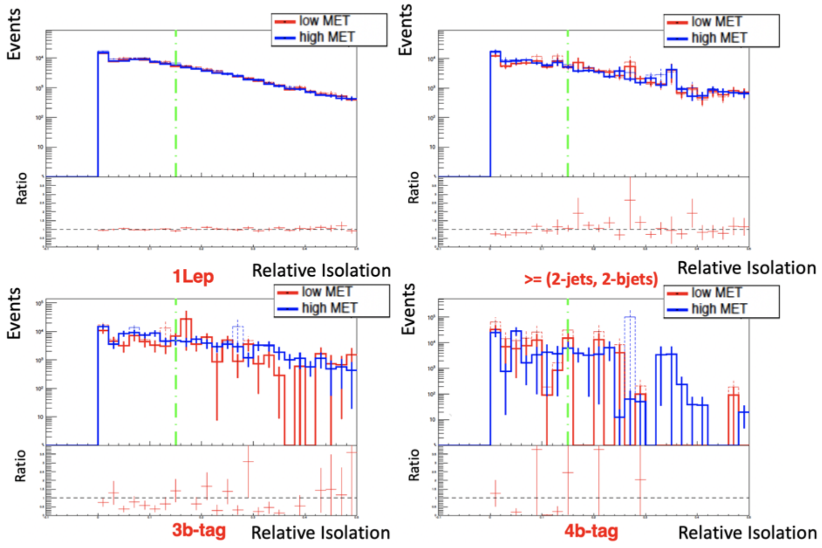


Figure 3-31: Comparison of Relative Isolation distributions between low-MET muons and high-MET muons with only 1 lepton cut (top left), 2-jet 2-bjet cut (top right), 3b cut (bottom left), and 4b cut (bottom right). The solid lines are after excluding all QCD datasets with $p_T < 80\ GeV$*. The dashed lines show the plots if we kept all datasets. The ratio is calculated excluding datasets with* $p_T < 80\ GeV$*.*

3.6.4.2 Similarity of BDT distributions of QCD MC in regions A, B, C, and D

As explained before, we estimate the number of $t\bar{t} + HF$ and W + jets background events in the signal region from the control regions. However, it is not only in the signal region that we suffer from poor QCD MC statistics, we also have poor QCD MC in the Top and W control regions. Therefore, in order to improve the Data/MC agreement in the control regions we will apply the ABCD method independently in control regions as well to make up for the poor QCD statistics. The Data/MC agreement is especially important in the BDT shape because we will derive the normalization of the $t\bar{t} + HF$ and W + jets backgrounds using the BDT shape as we will see later.

To justify the utilization of the ABCD method, it is important to show that the BDT shape of QCD MC does not change in different regions of the ABCD method. Therefore, we compare

- Regions A vs B, and regions D vs C which belong to the same E_T^{miss}-M_T region but different reliso region.
- Regions A vs C, and regions B vs D which belong to different E_T^{miss}-M_T region but the same reliso region.

The comparison is done independently for 3b-tag and 4b-tag cuts as shown in Figure 3-32 to Figure 3-39. As we can see from these plots, the better the statistics the closer the ratio is to one. The number of events in every ABCD region is mentioned in the description of every photo. Therefore, we can deduce the similarity of BDT distributions in different regions of ABCD method because when there is enough statistics, we see the similarity of shape.

Table 3-15 and Table 3-16 show the results of applying the ABCD method on both Data and MC. In other words, they show the number of predicted and observed events in the signal and control regions. "SF_{QCD} MC" and "SF_{QCD} measured" are scale factors calculated with eq.52 and eq.54 respectively. The last row is the ratio of predicted yield to observed yield in every region which should be close to one. Although this ratio is consistent with one in most regions, we consider a conservative 50% systematic uncertainty to account for uncertainty in prediction coming from poor QCD MC statistics.

3.6.4.2.1 Electrons in $t\bar{t}$ + HF control region

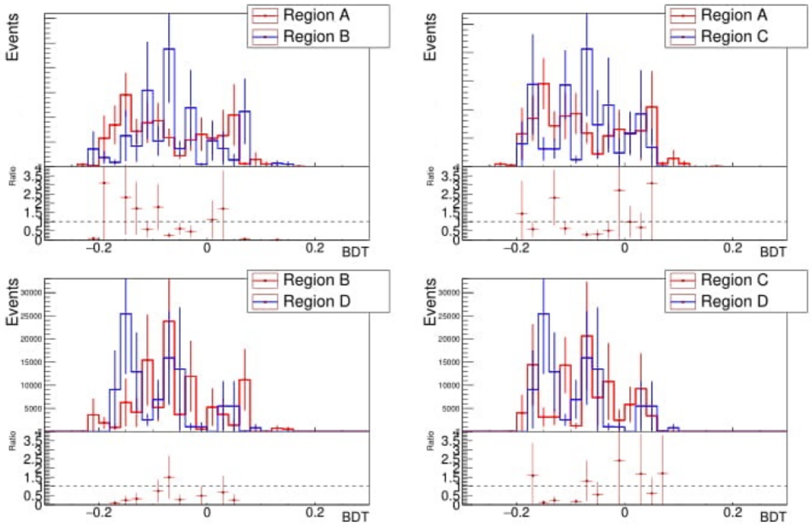


Figure 3-32: Comparison between BDT distributions of different regions of the ABCD method for electrons in the tt+HF control region (3b category). The plots are scaled to have the same area. The numbers of events are A:4009.27, B:1950, C:2062.27, and D:1188.1. As we can see, the more events we have the better is the agreement. (QCD datasets with $p_T < 80\ GeV$ are removed.)

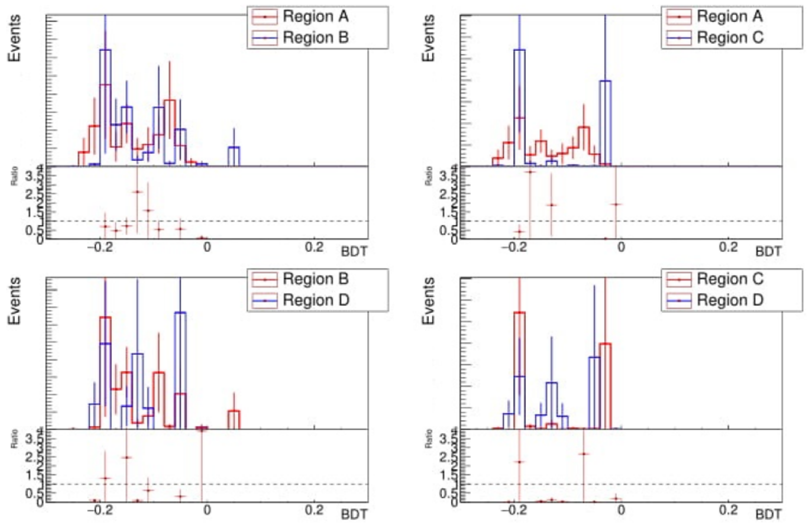


Figure 3-33: Comparison between BDT distributions of different regions of the ABCD method for electrons in the tt+HF control region (4b category). The plots are scaled to have the same area. The numbers of events are A:912.858, B:338.608, C:397.4, and D:310.078. As we can see, the more events we have the better is the agreement. (QCD datasets with $p_T < 80\ GeV$ are removed.)

3.6.4.2.2 Muons in $t\bar{t}\ +\ HF$ control region

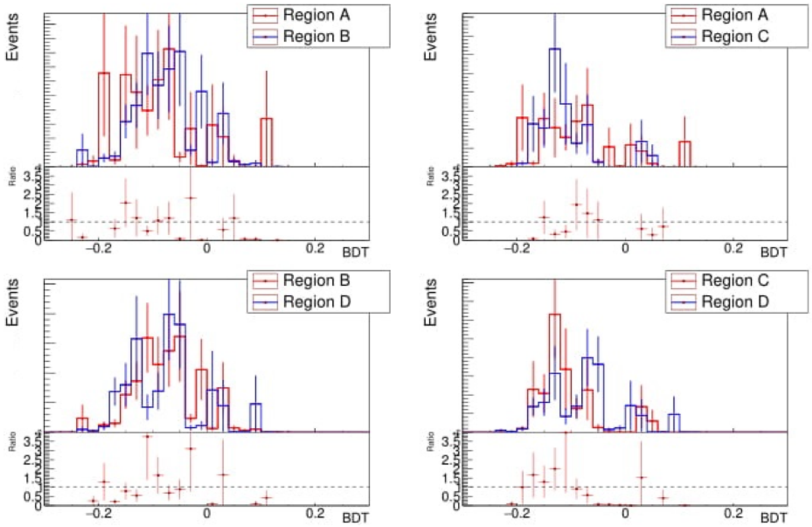


Figure 3-34: Comparison between BDT distributions of different regions of the ABCD method for muons in the tt+HF control region (3b category). The plots are scaled to have the same area. The numbers of events are A:2353.99, B:4098.94, C:2472.6, and D:3429.97. As we can see, the more events we have the better is the agreement. (QCD datasets with $p_T < 80\ GeV$ are removed.)

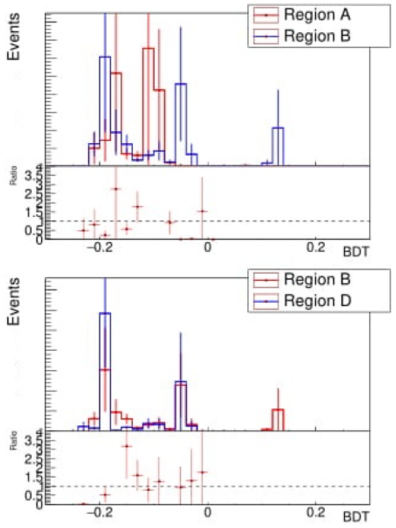


Figure 3-35: Comparison between BDT distributions of different regions of the ABCD method for muons in the tt+HF control region (4b category). The plots are scaled to have the same area. The numbers of events are A:301.635, B:1230.79, C:468.53, and D:635.34. As we can see, the more events we have the better is the agreement. (QCD datasets with $p_T < 80\ GeV$ are removed.)

3.6.4.2.3 Electrons in W + Jets / $t\bar{t}\ +\ LF$ control region

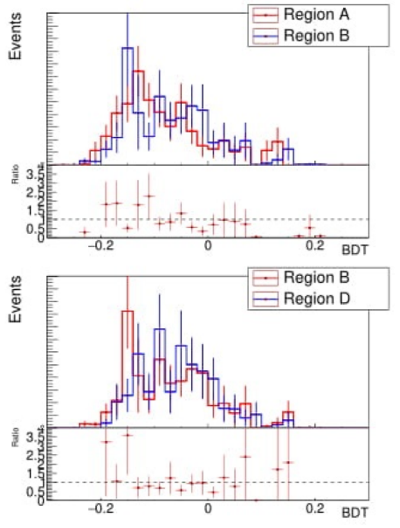


Figure 3-36: Comparison between BDT distributions of different regions of the ABCD method for electrons in the W/tt+LF control region (3b category). The plots are scaled to have the same area. The numbers of events are A:12393.6, B:6783.45, C:7951.47, and D:5715.64. As we can see, the more events we have the better is the agreement. (QCD datasets with $p_T < 80\ GeV$ are removed.)

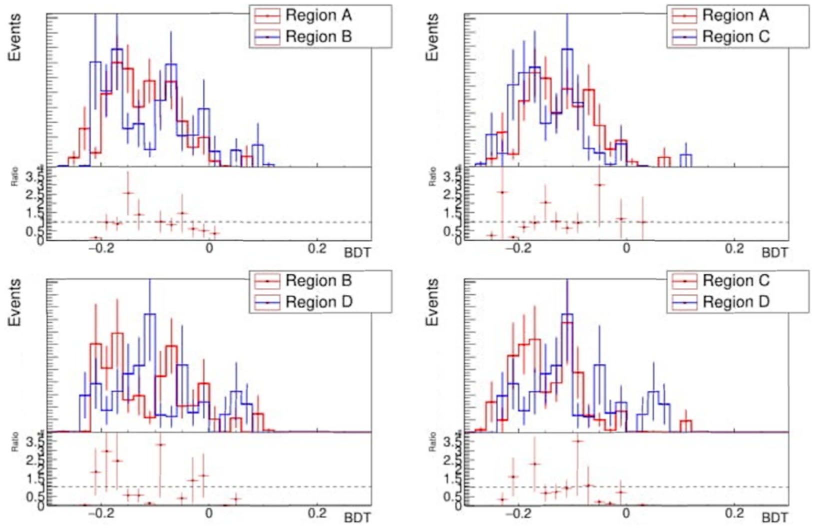


Figure 3-37: Comparison between BDT distributions of different regions of the ABCD method for electrons in the W/tt+LF control region (4b category). The plots are scaled to have the same area. The numbers of events are A:9352.94, B:7362.77, C:6020.9, and D:4852.61. As we can see, the more events we have the better is the agreement. (QCD datasets with $p_T < 80\ GeV$ are removed.)

3.6.4.2.4 Muons in W + Jets / $t\bar{t}$ + LF control region

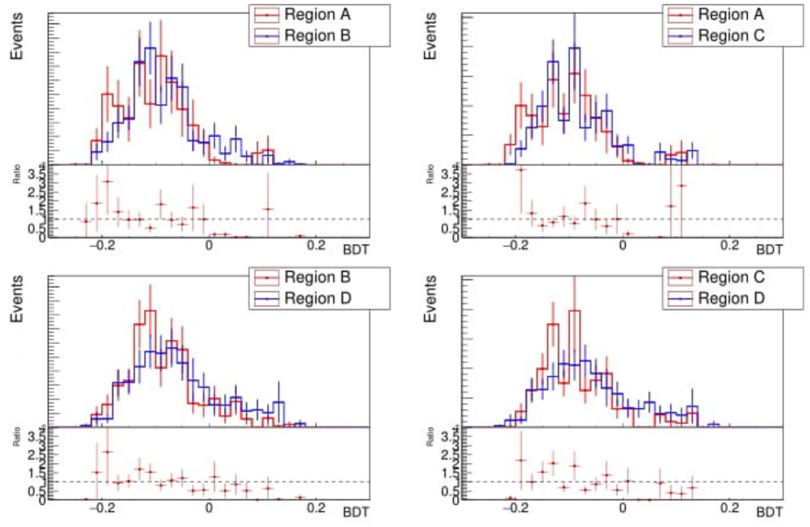


Figure 3-38: Comparison between BDT distributions of different regions of the ABCD method for muons in the W/tt+LF control region (3b category). The plots are scaled to have the same area. The numbers of events are A:8508.14, B:14422.3, C:18194, and D:22220.7. As we can see, the more events we have the better is the agreement. (QCD datasets with $p_T < 80\ GeV$ are removed.)

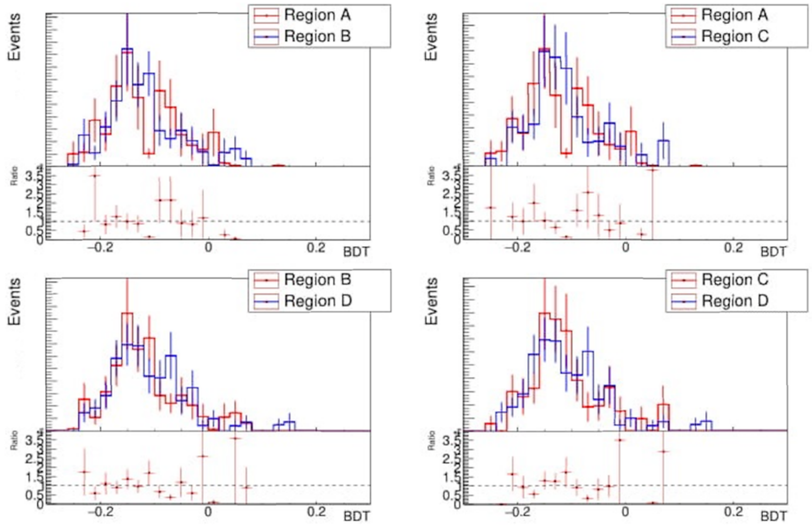


Figure 3-39: Comparison between BDT distributions of different regions of the ABCD method for muons in the W/tt+LF control region (4b category). The plots are scaled to have the same area. The numbers of events are A:6593.12, B:10620.8, C:9605.35, and D:13916. As we can see, the more events we have the better is the agreement. (QCD datasets with $p_T < 80\ GeV$ *are removed.)*

Table 3-15: Data-driven QCD background estimation in the 3b-tag category.

channel	e tt+HF CR - (3b, 5j)	e W CR - (2b, 3j)	e SR - (3b, 3j), (3b, 4j)	mu tt+HF CR - (3b, 5j)	mu W CR - (2b, 3j)	mu SR - (3b, 3j), (3b, 4j)
SF_{qcd} measured	1.36 ± 0.03	1.287 ± 0.008	1.41 ± 0.02	0.573 ± 0.007	0.542 ± 0.002	0.571 ± 0.003
QCD yield predicted (data)	$(123.2 \pm 2.0 \pm 61.6) \times 10^2$	$(477.3 \pm 3.5 \pm 238.6) \times 10^2$	$(168.5 \pm 2.2 \pm 84.2) \times 10^2$	$(855.7 \pm 12.2 \pm 427.9) \times 10^1$	$(356.5 \pm 1.9 \pm 178.3) \times 10^2$	$(121.2 \pm 1.2 \pm 60.6) \times 10^2$
QCD yield observed (MC)	$(525.0 \pm 101.5) \times 10^1$	$(269.9 \pm 40.4) \times 10^2$	$(820.3 \pm 226.5) \times 10^1$	$(496.6 \pm 182.6) \times 10^1$	$(194.1 \pm 34.3) \times 10^2$	$(123.2 \pm 42.0) \times 10^2$
SF_{qcd} (MC)	1.0 ± 0.6	0.8 ± 0.2	0.8 ± 0.4	0.8 ± 0.2	0.47 ± 0.05	0.5 ± 0.1
QCD yield predicted (MC)	$(524.0 \pm 101.3) \times 10^1$	$(216.5 \pm 32.4) \times 10^2$	$(688.1 \pm 190.0) \times 10^1$	$(383.9 \pm 141.2) \times 10^1$	$(921.8 \pm 162.7) \times 10^1$	$(674.9 \pm 229.9) \times 10^1$
QCD yield observed (MC)	$(525.0 \pm 101.5) \times 10^1$	$(269.9 \pm 40.4) \times 10^2$	$(820.3 \pm 226.5) \times 10^1$	$(496.6 \pm 182.6) \times 10^1$	$(194.1 \pm 34.3) \times 10^2$	$(123.2 \pm 42.0) \times 10^2$
ratio MC (syst)	1.0 ± 0.2	0.8 ± 0.1	0.8 ± 0.3	0.8 ± 0.4	0.47 ± 0.10	0.5 ± 0.3

Table 3-16: Data-driven QCD background estimation in the 4b-tag category.

channel	e tt+HF CR - (4b, 5j)	e W CR - (2b, 4j)	e SR - (4b, 4j)	mu tt+HF CR - (4b, 5j)	mu W CR - (2b, 4j)	mu SR - (4b, 4j)
SF_{qcd} measured	1.74 ± 0.08	1.34 ± 0.01	1.56 ± 0.09	0.77 ± 0.02	0.537 ± 0.003	0.77 ± 0.02
QCD yield predicted (data)	$(471.9 \pm 13.2 \pm 235.9) \times 10^1$	$(305.7 \pm 2.9 \pm 152.8) \times 10^2$	$(107.2 \pm 5.2 \pm 53.6) \times 10^1$	$(371.5 \pm 8.8 \pm 185.8) \times 10^1$	$(195.0 \pm 1.5 \pm 97.5) \times 10^2$	$951.4 \pm 37.2 \pm 475.7$
QCD yield observed (MC)	$(156.1 \pm 51.0) \times 10^1$	$(119.9 \pm 15.4) \times 10^2$	227.4 ± 154.9	$(114.2 \pm 78.7) \times 10^1$	$(144.2 \pm 32.7) \times 10^2$	128.3 ± 115.6
SF_{qcd} (MC)	1.4 ± 1.0	1.0 ± 0.3	0.005 ± 0.006	0.1 ± 0.1	0.52 ± 0.07	1.4 ± 1.4
QCD yield predicted (MC)	$(215.4 \pm 70.4) \times 10^1$	$(114.9 \pm 14.8) \times 10^2$	1.1 ± 0.7	161.4 ± 111.2	$(753.8 \pm 171.0) \times 10^1$	182.8 ± 164.7
QCD yield observed (MC)	$(156.1 \pm 51.0) \times 10^1$	$(119.9 \pm 15.4) \times 10^2$	227.4 ± 154.9	$(114.2 \pm 78.7) \times 10^1$	$(144.2 \pm 32.7) \times 10^2$	128.3 ± 115.6
ratio MC (syst)	1.4 ± 0.6	1.0 ± 0.1	0.005 ± 0.005	0.1 ± 0.2	0.5 ± 0.1	1.4 ± 2.4

3.6.4.3 Basic kinematic distributions after applying the ABCD method

The basic kinematic distributions are shown in Figure 3-40 to Figure 3-47 for CRs of W + Jets, $t\bar{t} + LF$, and $t\bar{t}\,HF$. In W + Jets and $t\bar{t}\ +\ LF$[30] control regions, there are only two reconstructed b-jets. Therefore, it is not possible to calculate variable like Δm_{bb}^{min} that are only defined with 3 or 4 b-jets. In this case, we add "fake" b-tags by adding the jet(s) that are not tagged as b-jets but have the highest DeepCSV discriminator value. In all plots, the fake-lepton background has been replaced by the data-driven estimate as described in section 3.6.4.

---

[30] LF stands for light flavor.

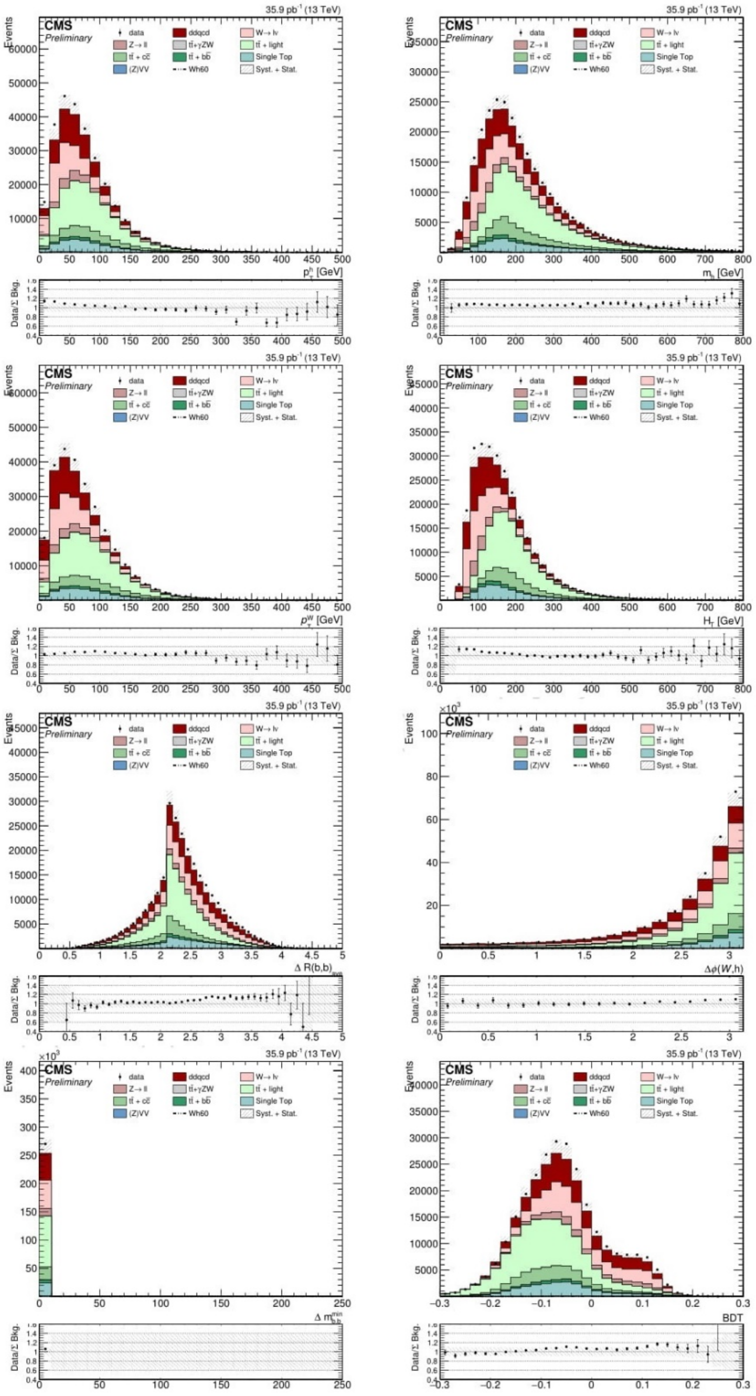


Figure 3-40: Analysis variable distributions in the (2b, 3jets) bin (W Control Region) for electrons for 3b-jet category. Note that Δm_{bb}^{min} is not defined in the 3b-tag category.

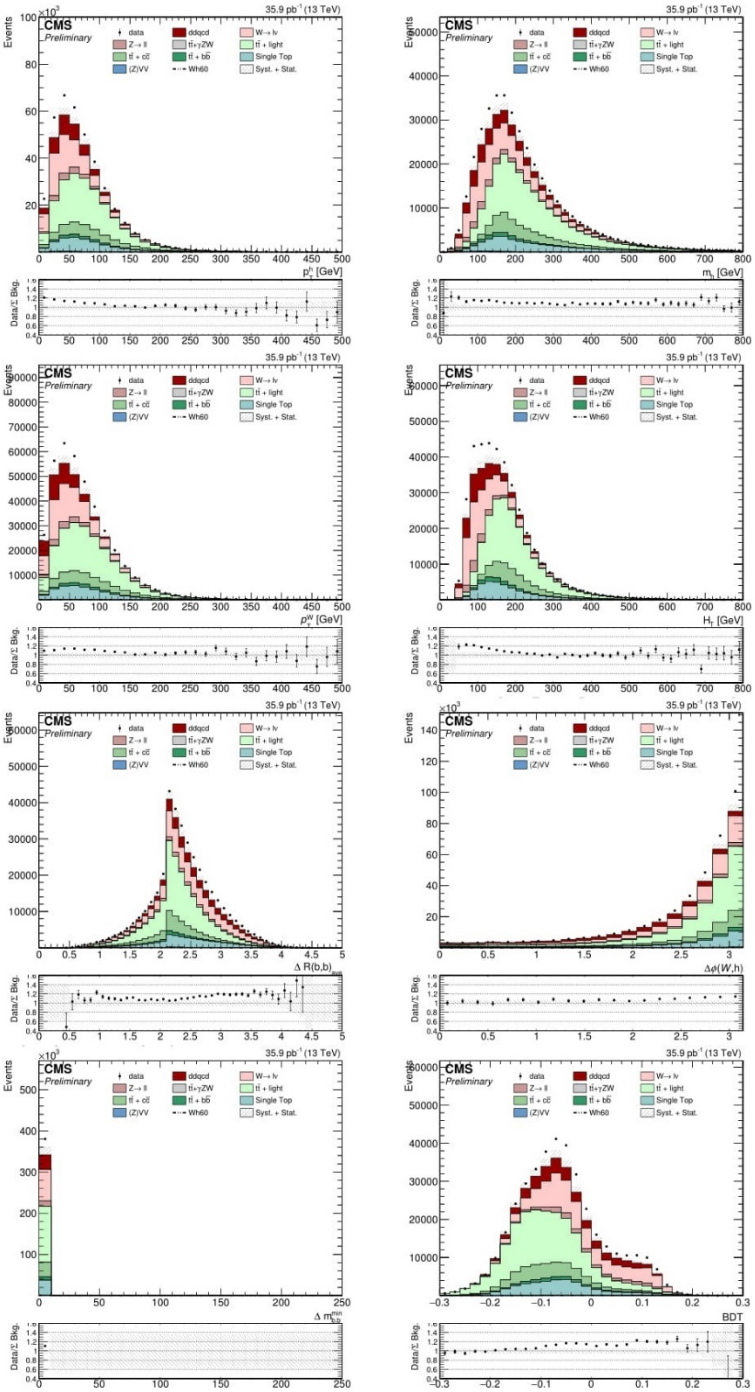


Figure 3-41: Analysis variable distributions in the (2b, 3jets) bin (W Control Region) for muons for 3b-jet category. Note that Δm_{bb}^{min} is not defined in the 3b-tag category.

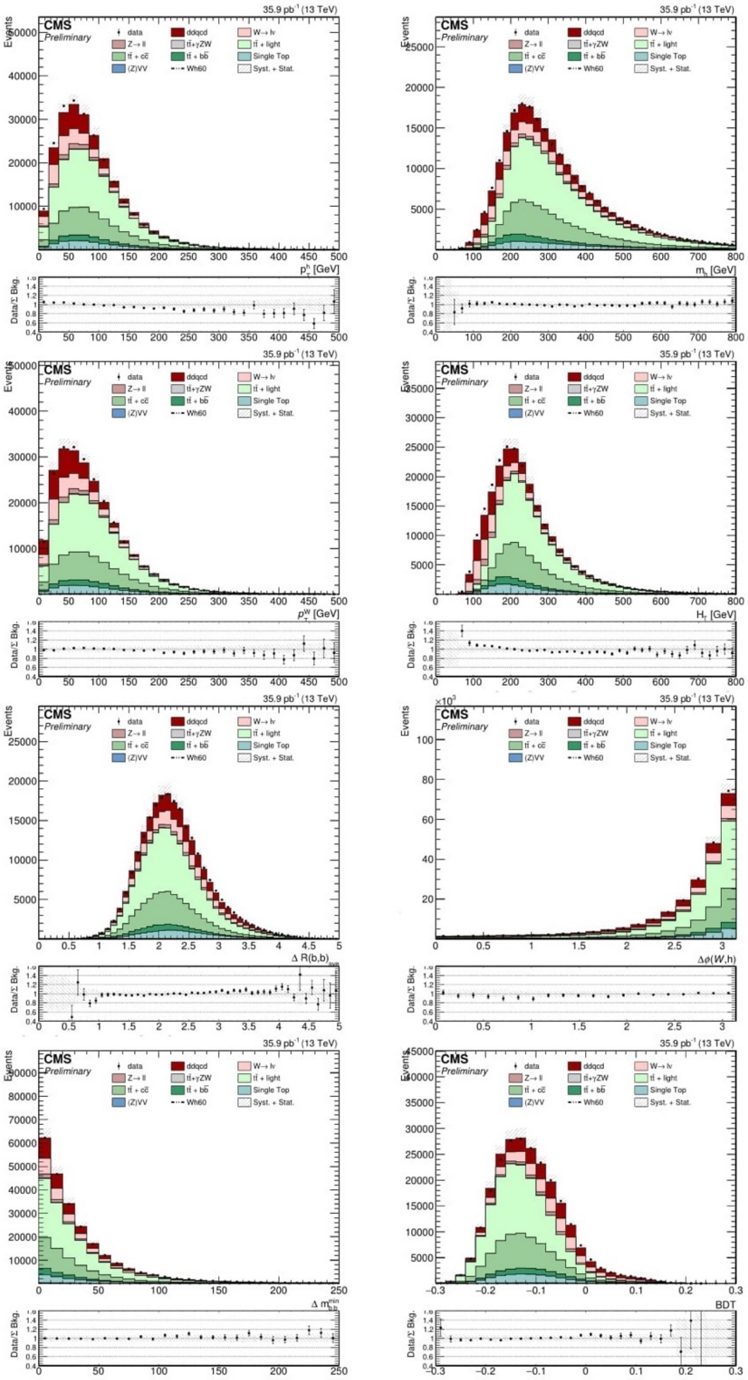


Figure 3-42: Analysis variable distributions in the (2b, 4jets) bin ($t\bar{t}+light$ Control Region) for electrons for 4b-jet category.

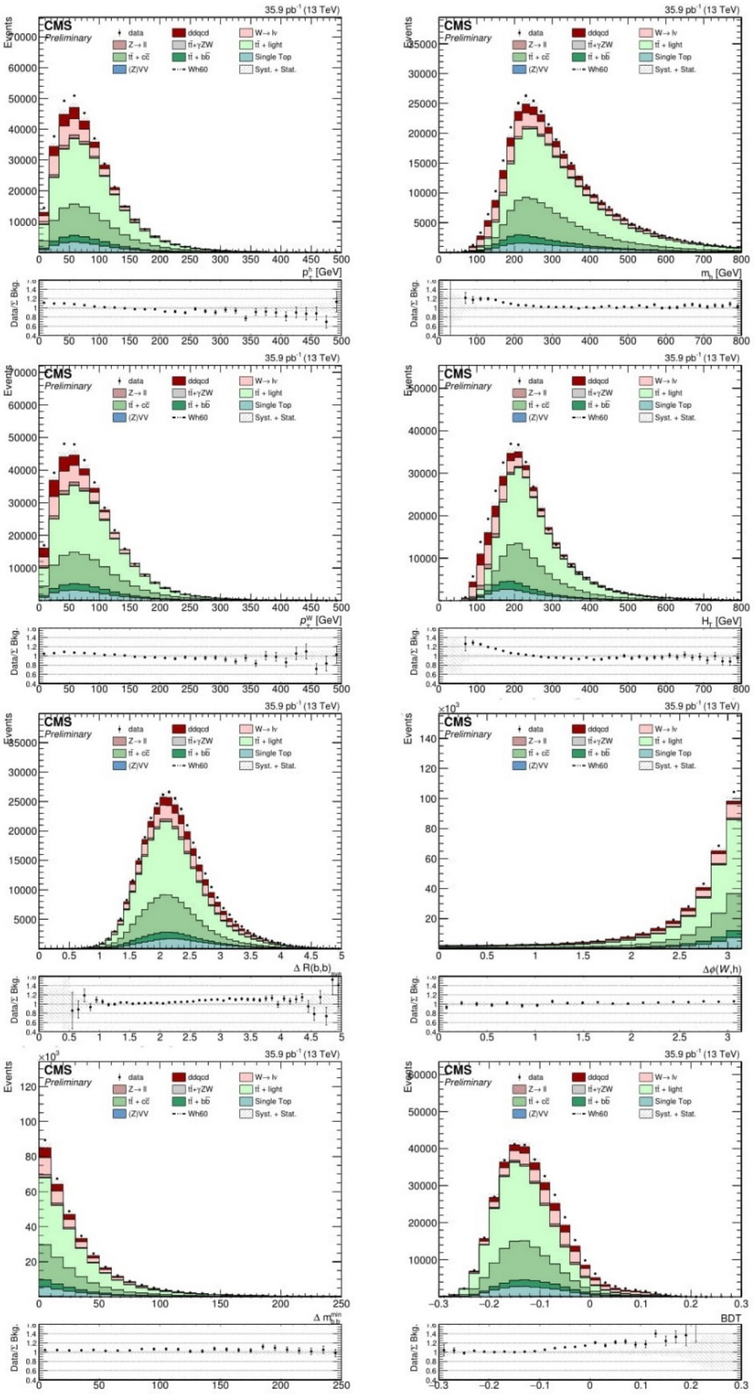


Figure 3-43: Analysis variable distributions in the (2b, 4jets) bin ($t\bar{t}$ + $light$ Control Region) for muons for 4b-jet category.

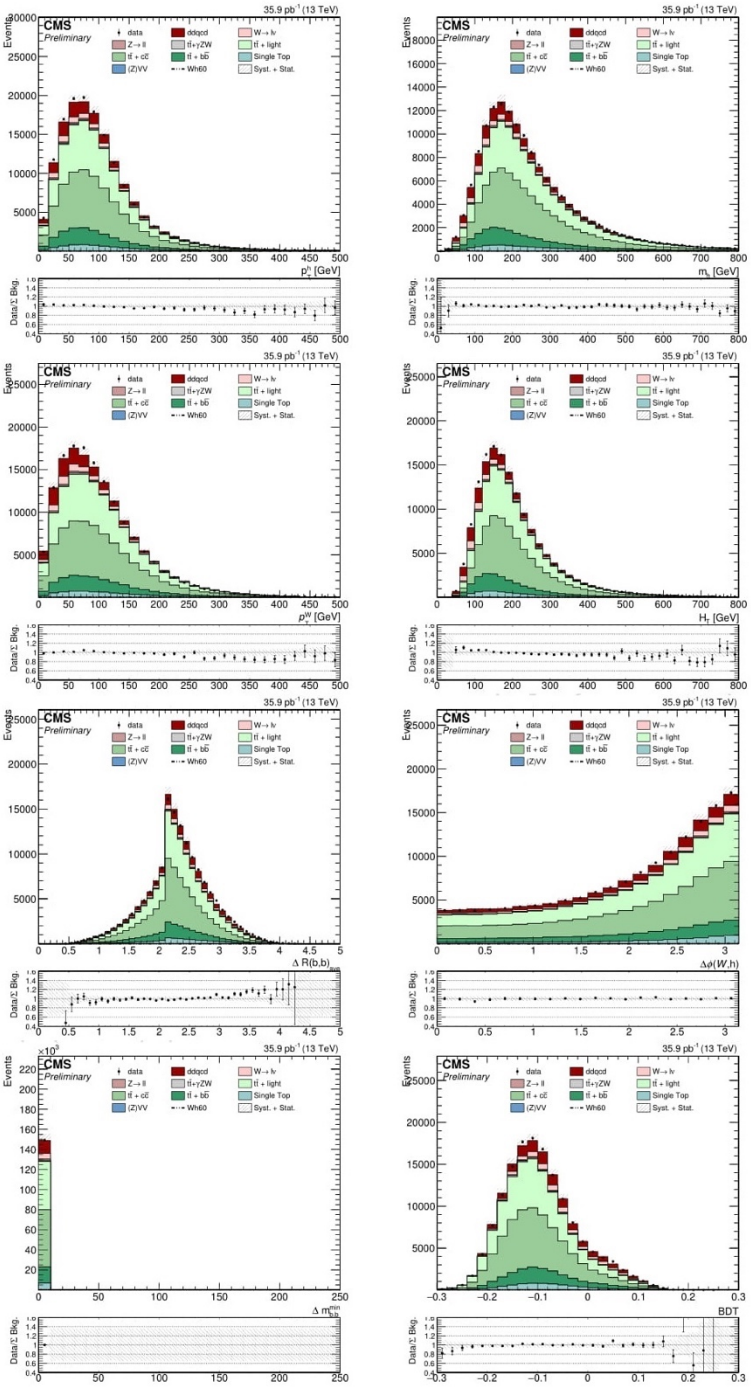


Figure 3-44: Analysis variable distributions in the (3b, 5jets) bin ($t\bar{t}$ + $heavy$ Control Region) for electrons for 3b-jet category. Note that Δm_{bb}^{min} is not defined in the 3b-tag category.

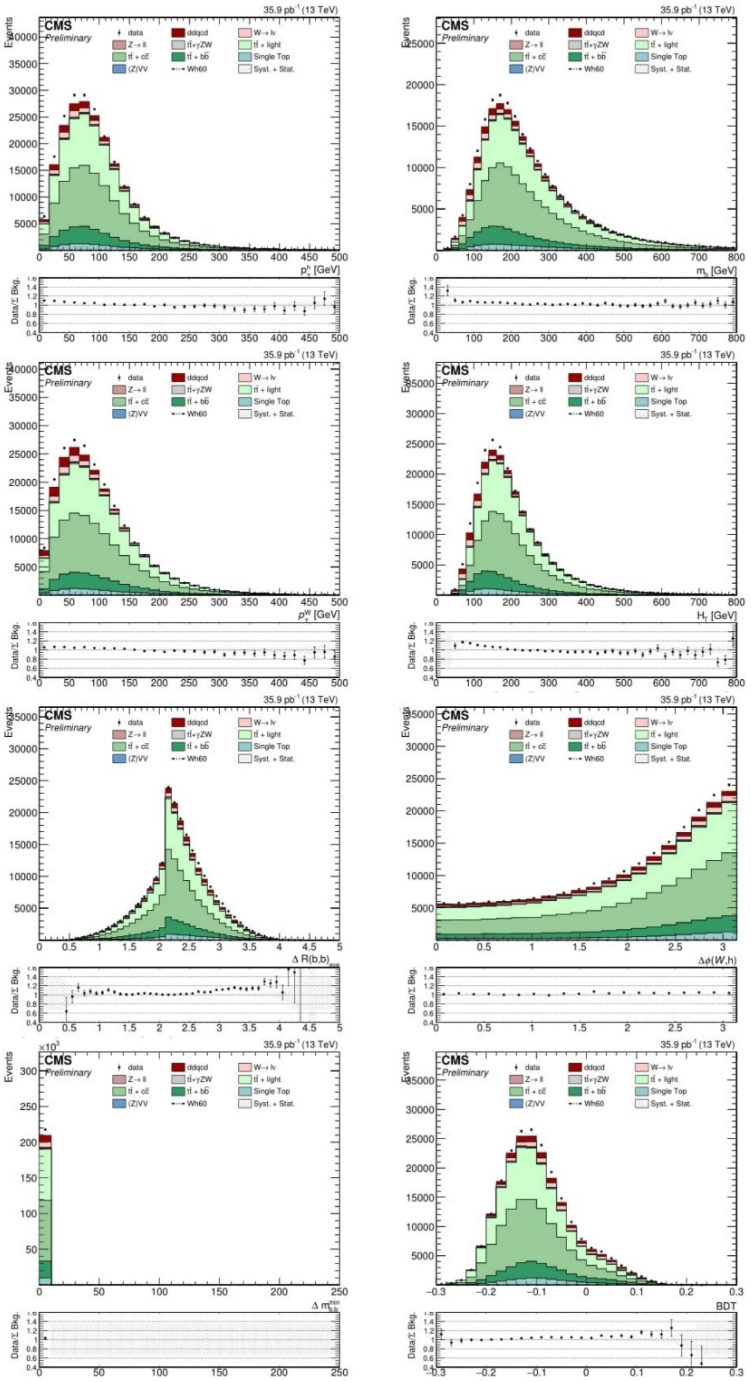


Figure 3-45: Analysis variable distributions in the (3b, 5jets) bin ($t\bar{t} + heavy$ Control Region) for muons for 3b-jet category. Note that Δm_{bb}^{min} is not defined in the 3b-tag category.

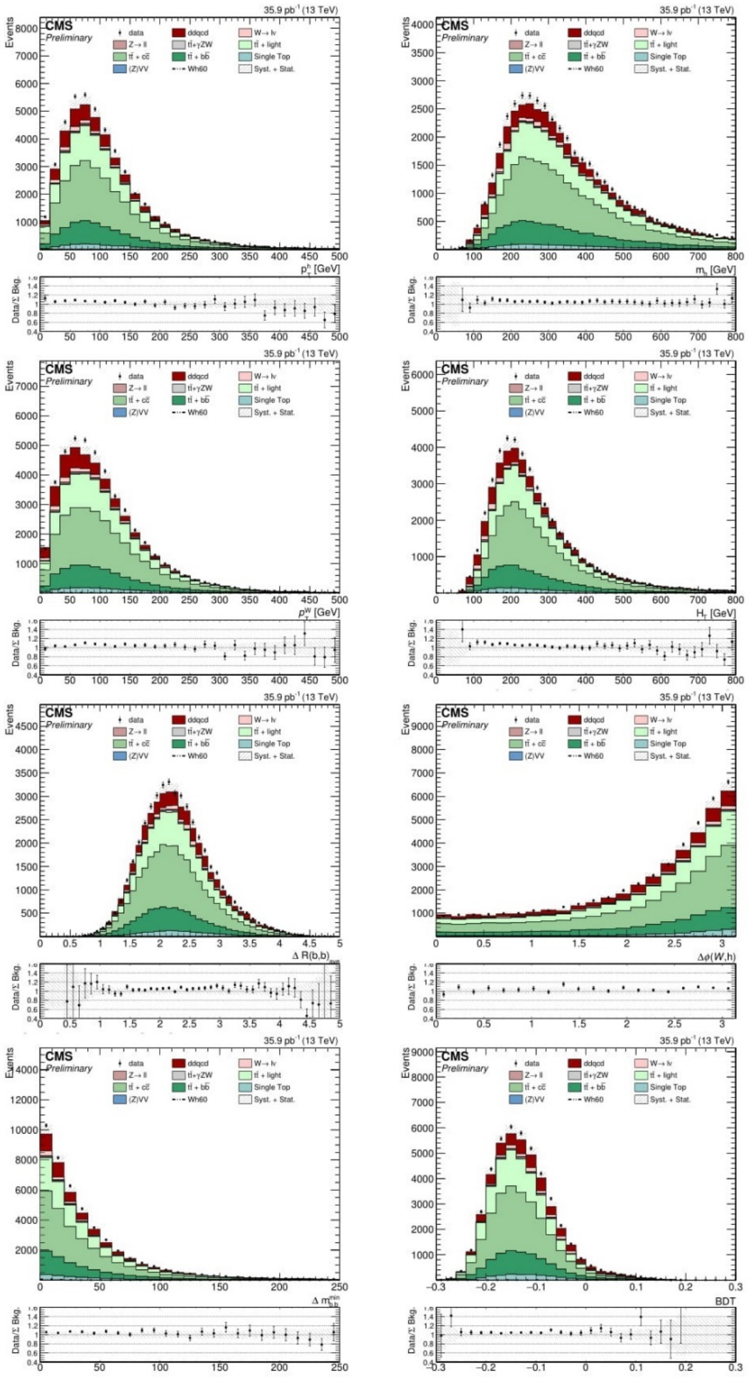


Figure 3-46: Analysis variable distributions in the (4b, 5jets) bin ($t\bar{t}$ + heavy Control Region) for electrons for 4b-jet category.

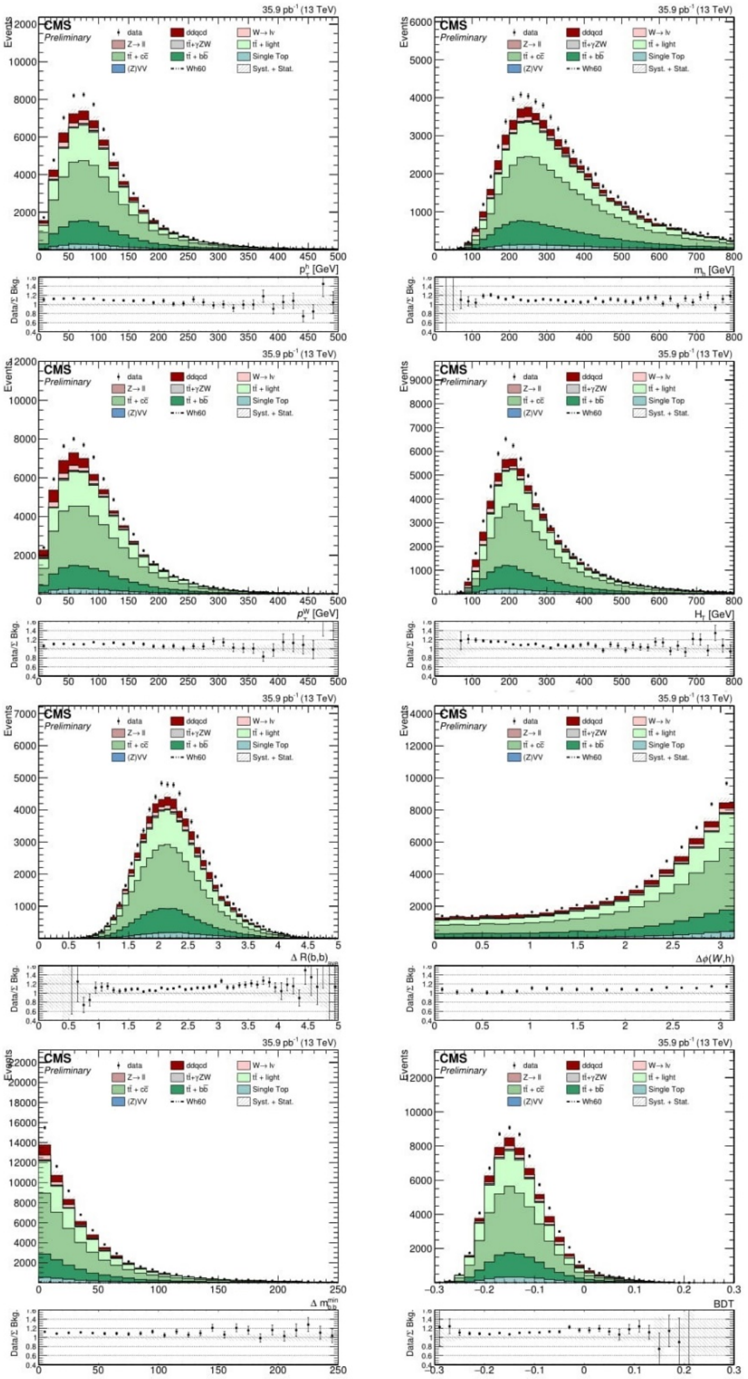


Figure 3-47: Analysis variable distributions in the (4b, 5jets) bin ($t\bar{t} + heavy$ Control Region) for muons for 4b-jet category.

3.7 Systematic Uncertainties

The sources of systematic uncertainties in estimating signal and background yields are discussed in this section. The sources are categorized into instrumental and theoretical uncertainties.

3.7.1 Instrumental Uncertainties

3.7.1.1 Luminosity

The uncertainty of the luminosity measurement is estimated to be 2.5% in 2016.

3.7.1.2 Lepton Efficiency

The E/gamma and Muon groups have recommendations for uncertainties of lepton triggering, identification, and isolation efficiencies. These recommended values have been evaluated by standard tag-and-probe techniques in Z-boson events. The single-electron trigger efficiency is slightly below 100% and thus the systematic uncertainty is estimated to be 1%. The single-muon studies contain the trigger efficiency as a function of the p_T of the Z boson, using their plots a 1% systematic uncertainty for the muon trigger efficiency is assumed. The uncertainties of identification and isolation efficiencies do not exceed 1% for muons and electrons. Therefore, the total uncertainty is 1.6% for muons and 1.5% for electron, which we take as a constant 2% for leptons.

3.7.1.3 Lepton Momentum Scale

The uncertainty of the muon momentum scale is assumed to be 1%. For electrons, the momentum scale uncertainty is estimated to be 0.6% for barrel region, and 1.5% for endcap region. The effect of the lepton p_T uncertainty on E_T^{miss} uncertainty is estimated by changing the nominal energy of the lepton by $\pm 1\sigma$ and then measuring the effect on E_T^{miss}. Table 3-17 shows the result.

Table 3-17: Systematic uncertainty of the electron energy scale. The uncertainty of muons is negligible in all of the processes.

Process	Category	
	3b-tag	4b-tag
$t\bar{t} + light$	0.7%	0.2%
$t\bar{t} + b\bar{b}$	0.3%	0.6%
$W \to lv$	1.3%	5.8%
$pp \to Wh(20)$	0.9%	0.9%
$pp \to Wh(60)$	2.3%	2.7%

3.7.1.4 Jet Energy scale, Resolution and E_T^{miss}

The uncertainties of the jet energy scale, resolution, and unclustered E_T^{miss} scale affect:

- Categorizing of the events as 3b-tag or 4b-tag
- The efficiency of b-tagging identification. B-tagging is applied to jets within the tracker acceptance and $p_T > 20\ GeV$
- The cut on E_T^{miss} and M_T

To estimate these effects, we changed the nominal jet energy scale (or resolution) by $\pm 1\sigma$ according to the prescription of the JetMET group [175, 176]. For the effect of the jet energy scale uncertainty on unclustered E_T^{miss} we used MiniAOD sample following the JetMET group recommendation. The clustered component is added back in order to reassess the E_T^{miss}.

The $\pm 1\sigma$ variations are used to recompute the E_T^{miss} and M_T for each event and then jets are b-tagged if the new corrected $p_T > 20$ GeV and $|\eta| < 2.4$. The result is shown in Table 3-18.

Table 3-18: Systematic uncertainty of the jet energy scale (JES), jet energy resolution(JER) and unclustured E_T^{miss} for the major SM background processes and signal with $m_a = 60\ GeV$.

Process	Category			Category		
	JES	JER	uE_T^{miss}	JES	JER	uE_T^{miss}
$t\bar{t} + light$	0.5%	0.4%	5.0%	1.4%	0.7%	5.6%
$t\bar{t} + c\bar{c}$	1.5%	0.8%	7.0%	0.7%	0.3%	7.2%
$t\bar{t} + b\bar{b}$	1.8%	1.0%	5.6%	2.0%	1.3%	5.4%
$W \to l\upsilon$	4.7%	1.8%	8.0%	3.4%	1.2%	8.2%
$pp \to Wh(60)$	1.6%	0.8%	8.6%	3.6%	2.3%	8.4%

3.7.1.5 Jet b-tagging

Official b-tagging scale factors are applied to jets above 20 GeV in signal and background per recommendations from the BTV[31] group [177]. We read the systematic uncertainties of the b-tag SF from the corresponding CSV file with the official BTagCalibration tool. The impact of this uncertainty is estimated by running the "up" and "down" BDT shape variations of the b-tag scale factors. The systematic uncertainties related to jet b-tagging are thus considered as shape uncertainties.

Soft-b tagging scale factor uncertainties are taken from [178]. The values have been evaluated using $t\bar{t}$ samples with different configurations in the physics modeling and

---

[31] BTV stands for B-Tag & Vertexing. BTV group measures the b-tagging efficiency scale factors for b and light flavor jets.

detector description (MadGraph at LO and Powheg-new tune NLO). The resulting SF(SV) and uncertainty used is 1.05 ± 0.16.

Table 3-19: Systematic uncertainty of the b-tagging scale factors for the major SM backgrounds and signal with $m_a = 60\ GeV$.

Process	b-tagging		c-tagging		mis-tagging	
	3b-tag	4b-tag	3b-tag	4b-tag	3b-tag	4b-tag
$t\bar{t} + light$	2.3%	2.3%	0.1%	0.4%	5.7%	12.3%
$t\bar{t} + b\bar{b}$	2.1%	2.2%	2.7%	3.7%	1.1%	7.0%
$t\bar{t} + c\bar{c}$	2.5%	3.5%	0.7%	1.0%	2.6%	5.5%
$W \to l\upsilon$	0.9%	1.6%	1.6%	4.8%	8.3%	11.6%
$pp \to Wh(60)$	3.0%	4.8%	0.7%	2.2%	0.9%	2.6%

3.7.2 Theoretical uncertainties

3.7.2.1 Limited size of the simulated samples

Although uncertainty due to limited statistics is not a theoretical uncertainty in principle, here it reflects an uncertainty in modelling of processes. In this analysis, limited statistics is mostly present in these samples: signal, dibosons and tribosons background, and also the $t\bar{t} + X$ and single top processes. These uncertainties are shown in Table 3-20 for the samples used in this analysis.[32]

Table 3-20: Systematic uncertainty due to the limited size of the simulated samples for the major SM backgrounds and signal with $m_a = 60\ GeV$.

Process	Category	
	3b-tag	4b-tag
$t\bar{t} + light$	0.3%	1.2%

32- Bin-by-bin statistical uncertainties are also included in the final limit extraction as will be shown later.

$t\bar{t}+c\bar{c}$	0.3%	1.0%
$t\bar{t}+b\bar{b}$	0.9%	2.2%
Single top	0.4%	1.9%
$W \to l\upsilon$	1.9%	8.3%
$Z \to ll$	1.3%	6.2%
$pp \to Wh(60)$	1.4%	3.5%

3.7.2.2 Cross-section uncertainties

We assign a normalization uncertainty to the cross-section of the small electroweak backgrounds of this analysis: dibosons/tribosons (50%), DY+jets (2%), $t\bar{t}+\ light$ (6%) and single top (5%).

3.7.2.3 PDF and α_S uncertainties

All the background and the signal samples in this analysis use NNPDF[33]. The uncertainty on the PDF and α_S[34] are computed according to the PDF4LHC recommendations [179] and are taken as shape uncertainties.

3.7.2.4 Summary of the systematic uncertainties

Figure 3-48 and Figure 3-49 show the individual "up" and "down" shape variations in BDT variable for each systematic uncertainty described above. The plots are shown for the major SM background which is the Top background and for a signal with $m_a = 60\ GeV$).

---

[33] This is a collaboration to extract parton distribution functions.

[34] The strong coupling explained in section 1.1.1.1

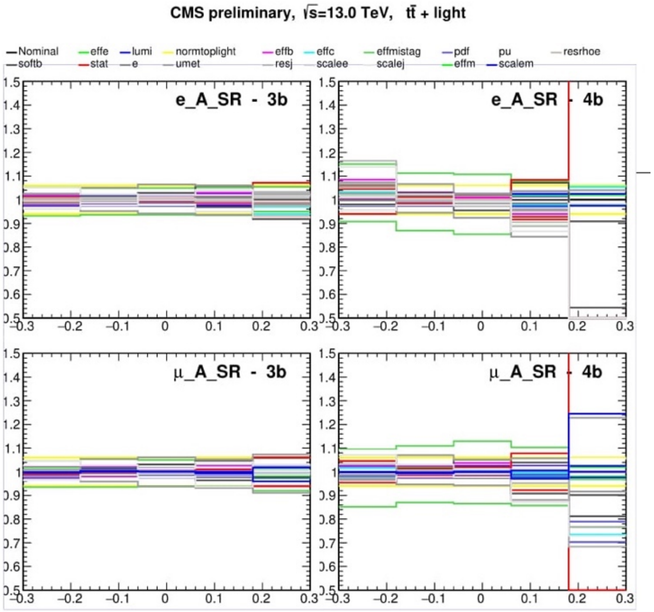


Figure 3-48: Individual “up” and “down” shape variations in BDT variable for each systematic uncertainty for $t\bar{t} + light$ background.

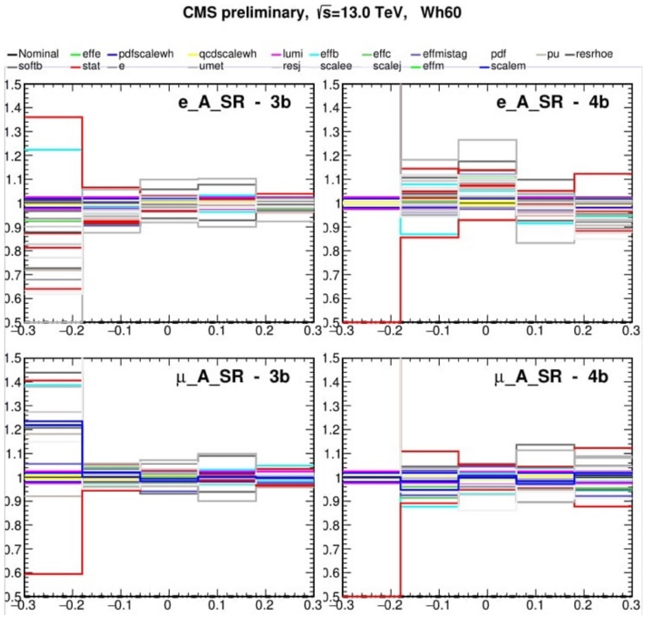


Figure 3-49: Individual “up” and “down” shape variations in BDT variable for each systematic uncertainty for signal with $m_a = 60\ GeV$.

3.8 Statistical analysis

3.8.1 Introduction

The current recommended methods by the LHC Higgs Combination Groups are:

- The hybrid Bayesian-Frequentist limits popularly known as "CLs of LEP or Tevatron type"
- The fully frequentist limits.

In this analysis, we will use the CLs method.

3.8.2 Probability, Inference, and Likelihood

According to the frequentist approach[35], “probability is the long-run relative frequency of an event.” [181]. However, when we have a theory, probability is the chance of an event happening according to the theory. For example, if our theory is that a coin is fair (i.e. Prob(tail) = Prob(head) = 0.5), then we can calculate the chance of having 59 tails and 41 heads. On the other hand, inference is using data to determine an unknown parameter of the theory, μ. To have a more accurate definition, [161] defines inference as “The process of determining an estimated value $\hat{\mu}$ and the corresponding uncertainty $\delta\mu$ of some unknown parameter μ from experimental data”. An example would be finding the chance of Prob(tail) = Prob(head) = 0.5 of a coin by flipping it 1000 times (for example we get

---

[35] 35- We will not talk about the Bayesian approach in this dissertation. However, a good discussion on Bayesian approach can be found in 180. Sivia, D. and J. Skilling, *Data analysis: a Bayesian tutorial*. 2006: OUP Oxford.

540 tails and 460 heads). In inference, the equivalent term for probability is "likelihood". Here are examples of probability and inference problems:

- Probability: Assuming that a coin is fair, what is the probability of getting 540 tails and 460 heads?
- Inference: In an experiment we got 540 tails and 460. What is the likelihood of the coin being fair?

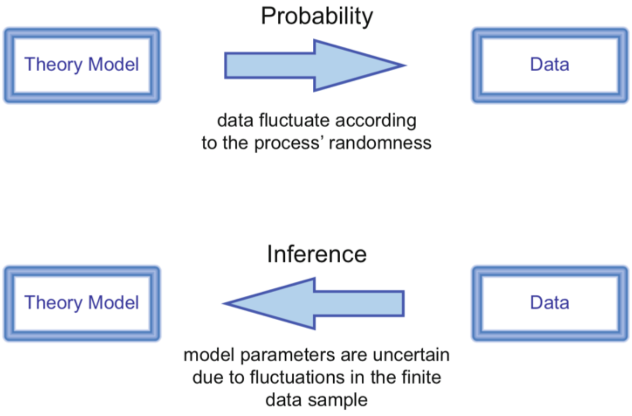


Figure 3-50: Relation between probability and inference. This image is taken from [161]

Our model ($h \rightarrow 2a \rightarrow 4b$ channel in NMSSM) has two parameters of interest (unknown parameters): The mass of the pseudo-scalar Higgs and the branching ratio of the decay. However, in order to eliminate one of these unknown parameters, we have chosen eight masses of $m_a = 12, 15, 20, 25, 30, 40, 50, 60\ GeV$. Therefore, at each m_a, we only have one parameter of interest: the branching ratio[36].

36- Actually, the parameter of interest that we will estimate is the signal strength. However, we have not defined it yet.

The mathematical formula for likelihood is[37]:

$$L(\mu|x) = f_\mu(x) \tag{55}$$

Where f is the PDF of collecting the dataset x if the variable of interest (branching ratio in our case) is μ. [182]

The number of recorded events N is also a variable that depends on μ. For example, the number of events that we expect in the signal region depends on the branching ratio of the $h \rightarrow 2a \rightarrow 4b$ decay. Therefore, we define the extended likelihood function as:

$$L(\mu|x) = P(N(\mu), \mu) \prod_{i=1}^{N(\mu)} f_\mu(x_i) \tag{56}$$

Where $P(N, \mu)$ is the probability distribution of the number of events N (N depends on μ). From our basic knowledge of probability, we know that $P(N, \mu)$ is a Poisson distribution with an average $\upsilon(\mu)$. Therefore:

$$L(\mu|x) = \frac{e^{-\nu(\mu)}\upsilon(\mu)^N}{N!} \prod_{i=1}^{N} f_\mu(x_i) \tag{57}$$

If data contains both signal and background, both signal yield can depend on μ:

$$\nu = s(\mu) + b \tag{58}$$

And:

$$f_\mu(x_i) = \frac{s}{s+b} f_s(\mu|x) + \frac{b}{s+b} f_b(\mu|x) \tag{59}$$

---

[37]- Because in this analysis, we only have one variable of interest μ, we defined the likelihood like this. In general, we can have m variables of interest $\mu_1, \dots, \mu_m$

Where υ, s, and b are the average of the Poisson distribution, number of signal events, and number of background events respectively. Therefore:

$$L(\mu, s, b|x) = \frac{e^{-(s+b)}(s+b)^N}{N!}\prod_{i=1}^{N}\frac{sf_s(\mu|x) + bf_b(\mu|x)}{s+b} = \\ = \frac{e^{-(s+b)}}{N!}\prod_{i=1}^{N} sf_s(\mu|x) + bf_b(\mu|x) \tag{60}$$

"In case of a very large number of measurements N, computing the likelihood function may become unpractical from the numerical point of view, and the implementation could require intensive computing power. Machine precision may also become an issue. For this reason, it is frequently preferred to perform the parameter estimate using a summary of the sample's information obtained by binning the distribution of the random variable and using as information the number of entries in each single bin: $(n_1, \dots, n_N)$, where the number of bins N is typically much smaller than the number of events n." [161] In this analysis, we also use binned BDT distribution with the binning defined in the description of Figure 3-15 and Figure 3-16.

In this analysis, we estimate the background yield in the SR from the amount of the background in the CRs for the $t\bar{t} + X$ and W + Jets backgrounds. However, we have some degree of Data/MC disagreement coming from mostly $t\bar{t} + b\bar{b}$ and W + Jets mismodeling. In order to compensate for the mismodeling we do a simultaneous Maximum Likelihood fit in the SR, $t\bar{t} + HF$ CR, and W + Jets CR. Here, we have three independent sets of data. Therefore:

$$L_{x_s, x_{b1}, x_{b2}}(\theta | x_s, x_{b1}, x_{b2}) = L_{x_s}(\theta | x_s) L_{x_{b1}}(\theta | x_{b1}) L_{x_{b2}}(\theta | x_{b2}) \tag{61}$$

Where x_s, x_{b1}, and x_{b2} are the data in SR, $t\bar{t} + HF$ CR, and W + Jets CR respectively.

3.8.2 Nuisance parameters

[183] defines nuisance parameters as “any parameter which is not of immediate interest, but which must be accounted for in the analysis of those parameters which are of interest”. At CMS, the nuisance parameters come mostly from the theoretical modeling and the response of the experimental detector: the finite resolution of the detector, mis-calibrations, the presence of background, etc. The probability model that describes the response of the CMS detector depends on unknown parameters. These additional unknown parameters are called nuisance parameters. [161]

3.8.3 Maximum Likelihood Method

Estimator is the function that estimates the unknown parameter (the so-called parameters of interest) from the data.[184] The estimator that we will use in this analysis is the maximum likelihood estimator which is the most popular estimate method. [161]

The maximum likelihood method is simply maximizing the likelihood function defined in the previous part. It is also called best fit method because it finds the parameter μ for which the theoretical model best fits the experimental data. [161] In other words, we want the value of μ for which:

$$\frac{\partial L}{\partial \mu} = 0 \tag{62}$$

However, instead of finding the maximum of the likelihood function, it is easier to find the maximum of its logarithm. It is because logarithm turns all multiplications into summation when taking derivative.

$$-\log L(\theta, s, b|x) = s + b + \sum_{i=1}^{N} \log[s f_s(\theta|x) + b f_b(\theta|x)] - \log N! \tag{63}$$

3.8.4 The Wilk's Theorem and the Profile Likelihood

When there is only one parameter of interest μ and some nuisance parameters that we show by $\vec{\theta}$, the Wilk's theorem states that under some condition[38]

$$\chi_r^2(\mu_0) = -2\log \frac{\sup_{\vec{\theta}} \prod_{i=1}^{N} L(\vec{x}_i;\ \mu_0, \vec{\theta})}{\sup_{\mu, \vec{\theta}} \prod_{i=1}^{N} L(\vec{x}_i;\ \mu, \vec{\theta})} \tag{64}$$

behaves as a χ^2 distribution with one degree of freedom when $N \to \infty$[39]. [161]

This is a very useful result as we will see later. We can see eq.64 as the ratio of what we define as the profile likelihood:

$$\boldsymbol{profile\ L} \coloneqq \sup_{\vec{\theta}} \prod_{i=1}^{N} L(\vec{x}_i;\ \mu_0, \vec{\theta}) \tag{65}$$

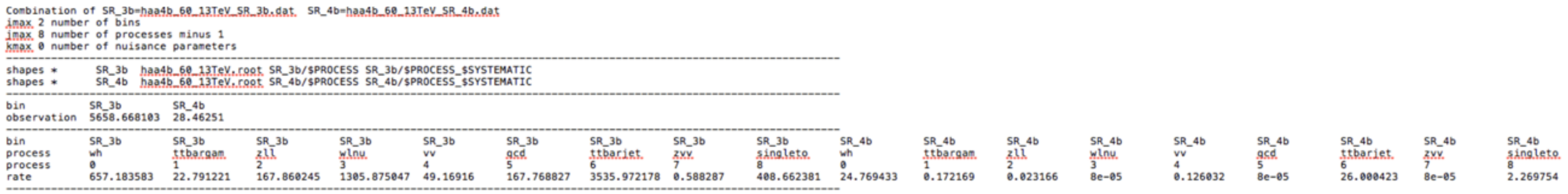
---

[38] Since this dissertation is not theoretical, the author is not trying to be mathematically accurate.

[39] Here μ_0 is the value of the parameter of interest under H_0. In this analysis, μ_0 is the branching ratio of the $h \to 2a \to 4b$ decay under H_0 which is equal to zero.

Since the maximum of the profile likelihood (PL) has exactly the same μ as the maximum of the likelihood itself, it can be used as a computational tool in maximum likelihood calculations.[185, 186]. In general, the profile likelihood can be asymptotically[40] used as an ordinary likelihood. [186] However, the importance of the Profile Likelihood goes far beyond computation as we will see later.

40- Meaning at very large number of events.



```
Combination of SR_3b=haa4b_60_13TeV_SR_3b.dat  SR_4b=haa4b_60_13TeV_SR_4b.dat
imax 2 number of bins
jmax 8 number of processes minus 1
kmax 0 number of nuisance parameters
----------------------------------------------------------------------------------------------------------------------------------
shapes *      SR_3b  haa4b_60_13TeV.root SR_3b/$PROCESS SR_3b/$PROCESS_$SYSTEMATIC
shapes *      SR_4b  haa4b_60_13TeV.root SR_4b/$PROCESS SR_4b/$PROCESS_$SYSTEMATIC
----------------------------------------------------------------------------------------------------------------------------------
bin          SR_3b        SR_4b
observation  5658.668103  28.46251
----------------------------------------------------------------------------------------------------------------------------------
bin          SR_3b        SR_3b        SR_3b        SR_3b        SR_3b        SR_3b        SR_3b        SR_3b        SR_3b        SR_4b        SR_4b        SR_4b        SR_4b        SR_4b        SR_4b        SR_4b        SR_4b        SR_4b
process      wh           ttbargam     zll          wlnu         vv           qcd          ttbarjet     zvv          singleto     wh           ttbargam     zll          wlnu         vv           qcd          ttbarjet     zvv          singleto
process      0            1            2            3            4            5            6            7            8            0            1            2            3            4            5            6            7            8
rate         657.183583   22.791221    167.860245   1305.875047  49.16916     167.768827   3535.972178  0.588287     408.662381   24.769433    0.172169     0.023166     8e-05        0.126032     8e-05        26.000423    8e-05        2.269754
----------------------------------------------------------------------------------------------------------------------------------
```

Figure 3-51: an example of a datacard for the Higgs Combine Tool

3.8.5 Simultaneous maximum likelihood fit of Top and W normalization

We will do a simultaneous fit (eq. 61) on the binned BDT shape of the two control regions ($t\bar{t} + HF$ and W + jets) as well as the signal region using the maximum profile likelihood estimator to derive the normalization of $t\bar{t} + b\bar{b}$ and W + Jets backgrounds using the CMS Higgs Combine tool. The normalizations are scale factors applied to $t\bar{t} + b\bar{b}$ and W + Jets backgrounds to account for Data/MC difference in SR and CRs. These normalizations are tied between the signal and the two control regions when fitting. In order to have a better result, the QCD component in all signal and control regions is replaced by the data-driven estimate (Explained in section 3.6.4

The initial values of the scale factors are determined using the integral of the histograms for the corresponding $t\bar{t} + b\bar{b}$ and W + Jets processes and the scale factor parameters are therefore treated as multiplicative factors of the nominal scale factor.

The CMS Higgs Combine tool is a command line interface to the RooFit/RooStats used widely inside CMS. [65] RooStats is the statistical framework for combination of the results in Higgs boson searches and measurements that the CMS and ATLAS collaborations agreed to use. [186-189]. In this framework, the distributions of signal and background events, parameters of interest, nuisance parameters, and observed data are encoded in a ROOT file called the workspace.

The input to the Higgs Combine Tool is a data card which defined the details of the experiment. [65] Figure 3-51 shows an example data card and what parameters are in it. The step to convert the datacard to a binary workspace can be performed by the text2workspace.py file.

The fitted values of $t\bar{t}+b\bar{b}$ and W + Jets normalizations are shown in Table 3-21. The extracted scale factors are compatible with similar observations in the VH(bb) analyses like [190]

Table 3-21: data/MC normalizations for the $t\bar{t}+b\bar{b}$ and W + Jets backgrounds for the SR+CRs simultaneous fit. The values refer to the simultaneous control regions plus signal region including channel masking in the SRs.

Background	B only fit	B+S fit
$t\bar{t}+b\bar{b}(e)$	1.53 ± 0.27	1.52 ± 0.28
$t\bar{t}+b\bar{b}(\mu)$	1.47 ± 0.26	1.47 ± 0.27
$W+3b(e)$	1.09 ± 0.12	1.09 ± 0.13
$W+3b(\mu)$	1.31 ± 0.09	1.31 ± 0.10
$W+4b(e)$	1.16 ± 0.32	1.15 ± 0.33
$W+4b(\mu)$	1.24 ± 0.19	1.23 ± 0.19

3.8.6 Post-fit results and fit diagnostics

Figure 3-52 to Figure 3-59 show the pre- and post-fit BDT distributions in $W/t\bar{t}+LF$ and $t\bar{t}+HF$ control regions. Uncertainties have been calculated considering the full list of systematic uncertainties described in section 3.7 by re-sampling of the fit covariance matrix, thereby accounting for the full correlation between the parameters of the fit.

The plots also show the ratio of data event yield to MC event yield with statistical and systematic uncertainties both considered in the error band. As expected, including the control regions in the fit helped constraining the systematical uncertainties significantly as can be seen in the post-fit distributions and the data-MC ratio.

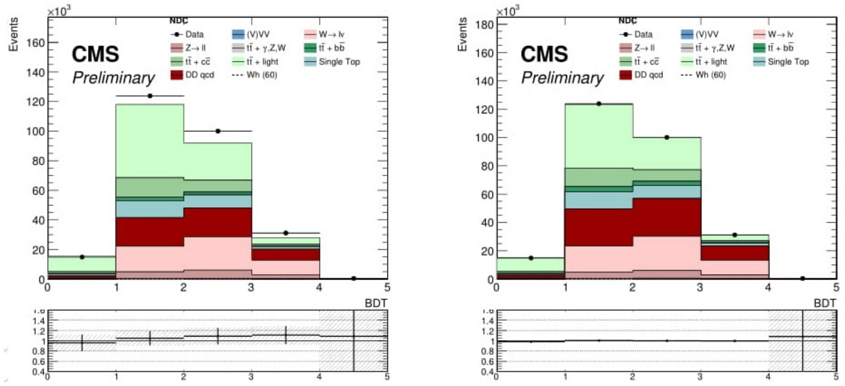


Figure 3-52: Pre-fit (left) and post-fit (right) BDT distributions in the $W/t\bar{t} + light$ CR - (2b, 3jets) , for electrons

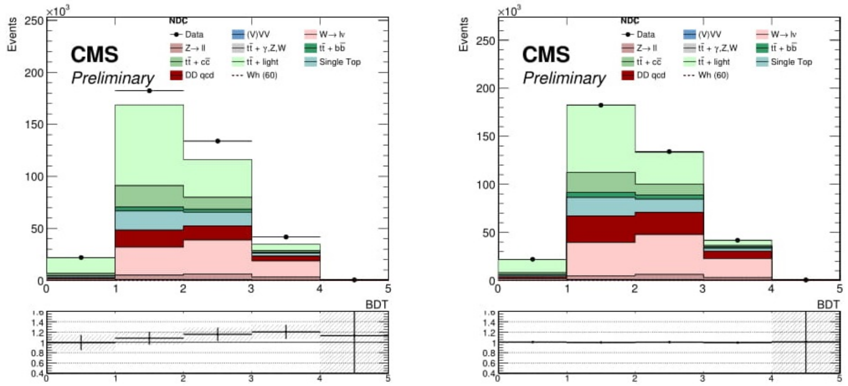


Figure 3-53: Pre-fit (left) and post-fit (right) BDT distributions in the $W/t\bar{t} + light$ CR - (2b, 3jets) , for muons

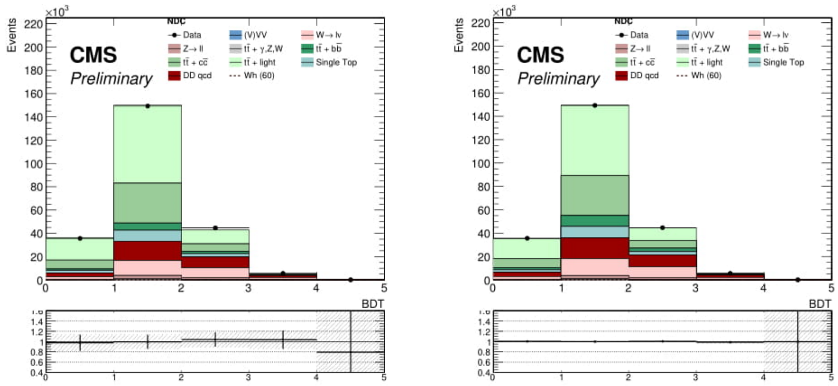


Figure 3-54: Pre-fit (left) and post-fit (right) BDT distributions in the $W/t\bar{t} + light$ CR - (2b, 4jets) , for electrons

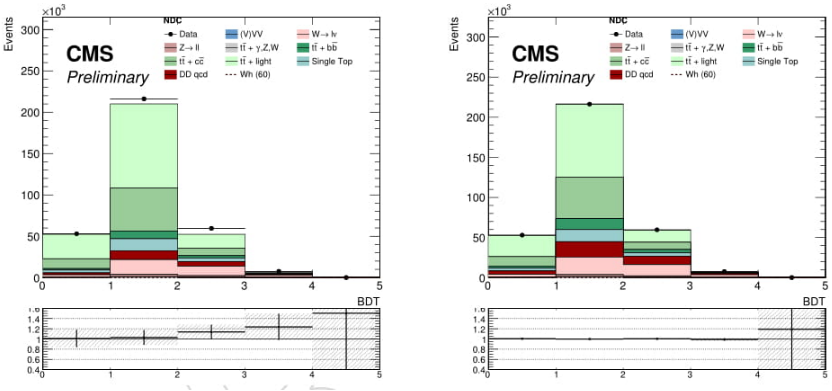


Figure 3-55: Pre-fit (left) and post-fit (right) BDT distributions in the $W/t\bar{t} + light$ CR - (2b, 4jets) , for muons

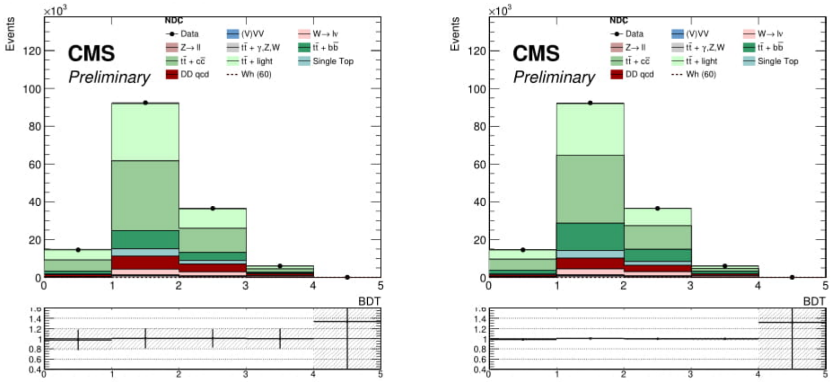


Figure 3-56: Pre-fit (left) and post-fit (right) BDT distributions in the $t\bar{t} + HF$ CR - (3b, 5jets), for electrons

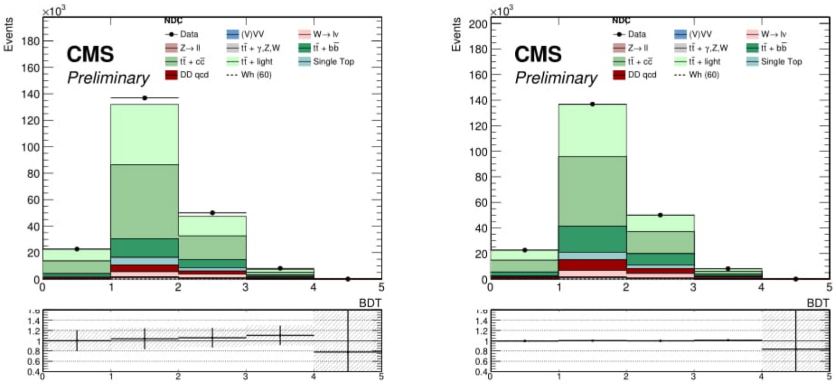


Figure 3-57: Pre-fit (left) and post-fit (right) BDT distributions in the $t\bar{t} + HF$ CR - (3b, 5jets), for muons

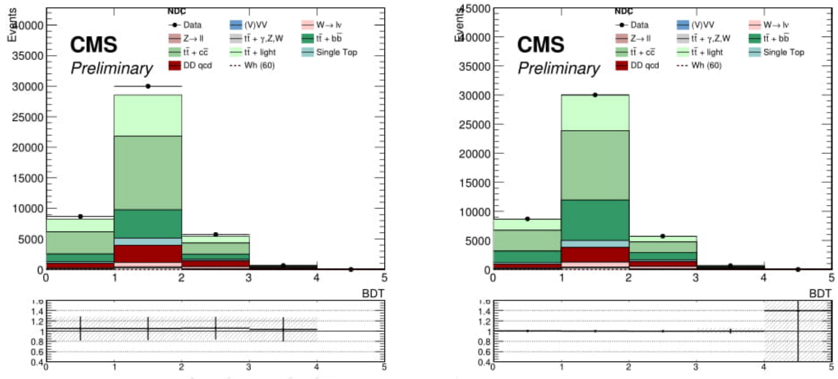


Figure 3-58: Pre-fit (left) and post-fit (right) BDT distributions in the $t\bar{t} + HF$ CR - (4b, 5jets), for electrons

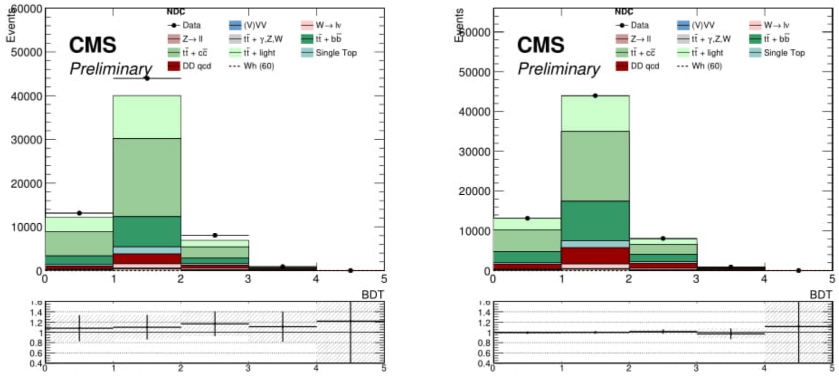


Figure 3-59: Pre-fit (left) and post-fit (right) BDT distributions in the $t\bar{t} + HF$ CR - (4b, 5jets), for muons

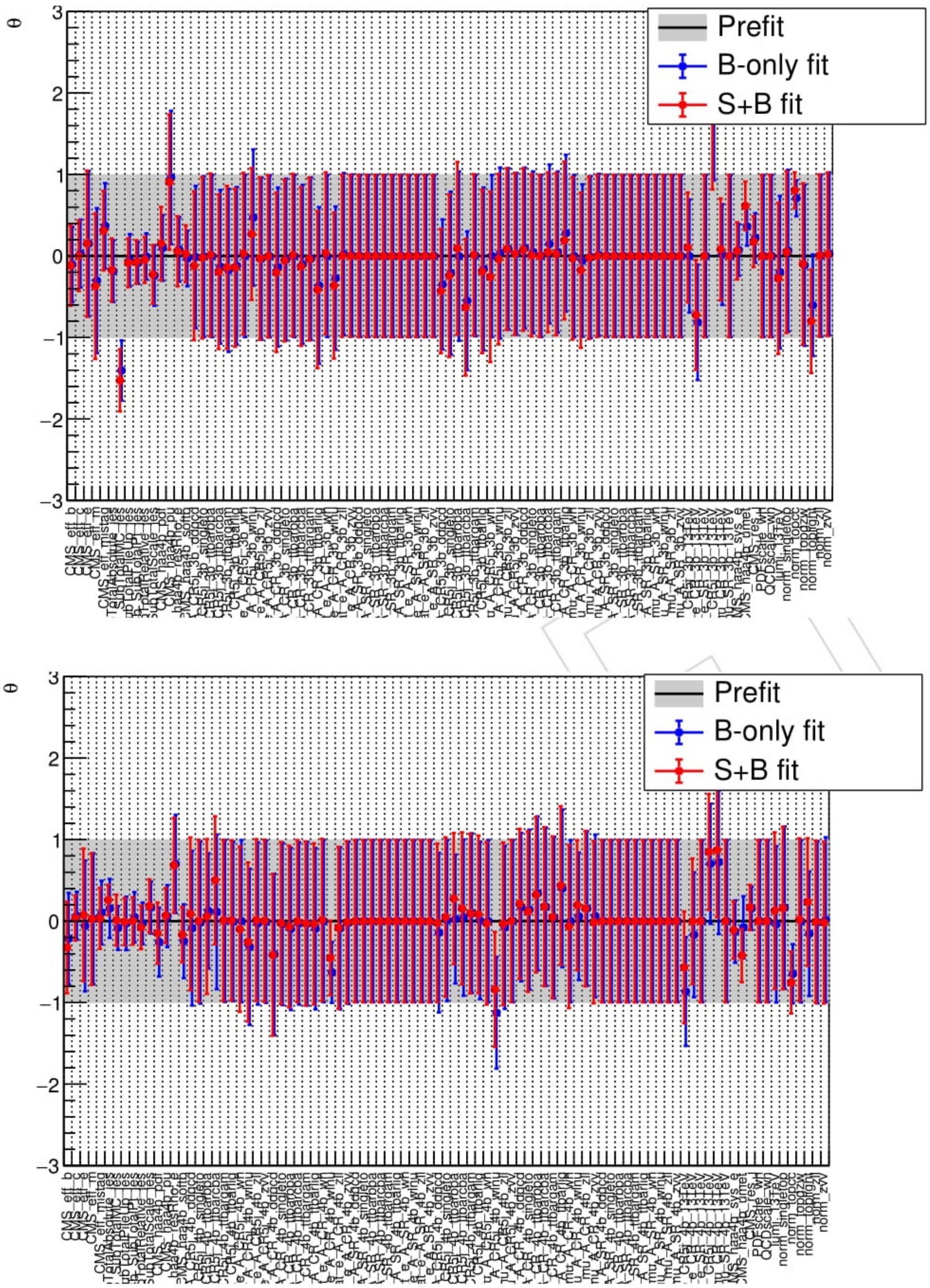


Figure 3-60: Pre-fit and post-fit values of the nuisance parameters. For each of those parameters, the plot shows the change in value and the post-fit uncertainty, both normalized to their input values. Upper and lower plots correspond to the 3b-tag and 4b-tag categories respectively.

3.9 Signal Extraction

3.9.1 Introduction

The goal of this analysis is to evaluate the $h \rightarrow 2a \rightarrow 4b$ decay that is predicted in some BSM theories using the data taken at CMS in 2016. This is done by looking at the BDT distribution of 2016 data to see if it is more consistent with the background BDT distribution or the background + signal BDT distribution. Hence, we will start by identifying and stating the null (H_0) and alternative (H_{alt}) hypotheses. The null hypothesis is the one that we will try to disprove. If we succeed, then the alternative hypothesis is automatically proved. However, if we fail to disprove H_0, it means that the data is more consistent with the absence of signal than its presence. In this case, the level of the absence of signal is characterized by an interval called an exclusion limit.[184] In our analysis, the exclusion limit is an upper limit on the of the $h \rightarrow 2a \rightarrow 4b$ decay at 95% confidence level as we will explain.

In this analysis, we define the null hypothesis as the hypothesis that $h \rightarrow 2a \rightarrow 4b$ channel does not happen (background only hypothesis) and the alternative hypothesis as that it does happen (background+signal hypothesis). Two things about the definition of the H_0 and H_{alt} hypotheses:

- It is also possible to choose H_0: background+signal and H_{alt}: background only. i.e. trying to disprove the $h \rightarrow 2a \rightarrow 4b$ channel. However, the abovementioned choice makes more logical sense.

- The disproval of the H_0 can happen in two ways: either if we have too much excess of data or too much lack of data compared to the background estimation. However, the presence of $h \to 2a \to 4b$ channel shows itself only as the excess of data in this analysis. Therefore, the alternative hypothesis is one-sided. We will add this one-sidedness to our mathematics later in this section.

3.9.2 The test statistic

In order to quantitatively evaluate the hypotheses, we reduce the data to one single value which is the numerical summary of that data sample. This single numerical value is called the test statistic and is usually shown by t. And its probability distribution can be derived under H_0 and H_{alt} from the probability distribution of the observable quantities in the experiment. [161, 191] We want the test statistic to be discriminating between a dataset with and without signal. Therefore, its probability distribution in general is different for the signal and background datasets (Figure 3-61). [161]

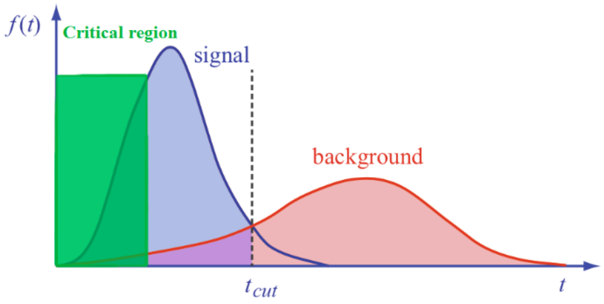


Figure 3-61: The test statistic variable in general has a different probability distribution for signal and background. This image is taken from [161]

After choosing the test statistic, we have to define a critical region (C_α) in its probability distribution (Figure 3-61). If the value of the test statistic for the 2016 CMS data falls in the critical region, we claim that we have disproved H_0 and discovered H_{alt}! We want this critical region

- To be as far as possible from the H_0 (background only hypothesis) probability distribution
- The probability that the test statistic value for a background only dataset falls in the critical region be less than α. i.e. $\Pr(t \in C_\alpha | H_0) < \alpha$. We define the confidence level as $(1 - \alpha)$ and it is expressed in percent. [192, 193]

3.9.3 Type-I and Type-II Errors

By looking at Figure 3-61 two important questions come to mind: What if the data is actually only background, however its test statistic value falls in the critical region by chance? Or what if the data contains signal, however the test statistic value falls out of the critical region by chance? These errors are called type-I and type-II errors:

- Type I error: Rejecting H_0 when it is actually true. The probability of Type-I error is α
- Type-II error: Erroneously failing to reject H_0 when it is false. We show the probability of Type-II error by β. [184]

Defining the critical region is kind of arbitrary. We can narrow C_α to decrease the probability of type-I error, however we will increase the probability of type-II error. Also, if we widen C_α to decrease the probability of type-II error, we will automatically increase the probability of type-I error[41]. Therefore, we have to trade off type-I error against type-II

41- We can even plot the type-I vs type-2 error probability in a ROC curve like Figure 3-13.

error. This trade off depends on every analysis. Depending on whether we prefer to avoid type-I error more or type-II error, we can have two different kinds of analysis:

- Best discovery potential: If our motivation for the analysis is discovery, avoiding type-I error would be more important than avoiding type-II error. This is because in high energy physics, we want to be very confident before claiming a discovery. Therefore, at CERN the threshold of 5.0 standard errors (it will be explained later) is used which corresponds to an alpha-level of approximately 0.000000573. [194][42]
- Best limit setting: However, if our goal is not discovery, probably it is because we already know that the signal is weak if not absent. Therefore, our main goal is to see any sign of the signal. In other words, we want to maximize prob(accept H_{alt} | H_{alt} = TRUE). We call this probability the statistical power:

$$\begin{aligned} POWER = Prob(accept\ H_{alt}|H_{alt} = TRUE) = \\ = 1 - Prob(reject\ H_{alt}|H_{alt} = TRUE) = 1 - \beta \end{aligned} \tag{66}$$

  This analysis is also a best limit analysis. This is because we already know that none of the analyses in Table 3-1 have seen the $h \to 2a \to 4b$ signal. Therefore, we will use $1 - \beta$ (statistical power) to enhance our chance of seeing the signal as we will see later in this part. [184, 196]

---

[42] 42- "It should be emphasized that in an actual scientific context, rejecting the background-only hypothesis in a statistical sense is only part of discovering a new phenomenon. One's degree of belief that a new process is present will depend in general on other factors as well, such as the plausibility of the new signal hypothesis and the degree to which it can describe the data. Here, however, we only consider the task of determining the p-value of the background-only hypothesis; if it is found below a specified threshold, we regard this as discovery" 195. Cowan, G., et al., *Asymptotic formulae for likelihood-based tests of new physics*. 2011. **71**(2): p. 1554.

In this part, we explained the hypothesis testing in a nut shell. The first step is to choose a test statistic. The most common test statistic in experimental high energy physics are the Neyman-Pearson (NP) and the Profile Likelihood (PL) [196]. In this analysis, we will use the Profile Likelihood test statistic.

3.9.4 Signal strength

In a counting experiment like the present one, if n is the number of observed events

$$n = \mu . s + b \tag{67}$$

Where b is the expected background. The signal strength (μ) is defined as:

$$\mu = \frac{\sigma_{obs}}{\sigma_{SM}} \tag{68}$$

3.9.5 The Neyman–Pearson Lemma and the Profile Likelihood test statistic

The Neyman-Pearson lemma states that the optimal test statistic is given by the ratio of the likelihood functions $L(\vec{x}|H_1)$ and $L(\vec{x}|H_0)$. Optimal here means that for a fixed α (confidence level) the selection that corresponds to the largest $1-\beta$ (statistical power) is given by: [161, 197]

$$\lambda(\vec{x}) = \frac{L(\vec{x}|H_1)}{L(\vec{x}|H_0)} \tag{69}$$

However, as we explained in section 3., we can use the profile likelihood as a regular likelihood at asymptotic regions (high number of events). Profile likelihood has the advantage that we know its distribution asymptotically. Therefore, we will use the Profile log-Likelihood ratio test statistic defined as:

$$q_\mu = -2\ln\frac{L\left(\mu,\hat{\hat{\theta}}_\mu\right)}{L(\hat{\mu},\hat{\theta}_\mu)} \tag{70}$$

Where μ is the parameter of interest and θ represent the nuisance parameters. One hat stands for the Maximum Likelihood Estimator (MLE) while the double hat means constrained MLE, i.e. the MLE of θ, fixing μ. In this case, we say that θ is profiled. [196] If the parameter of interest is the signal strength (defined in eq.68), we can adjust the PL test statistic to account only for excess of data:

$$q_0 = \begin{cases} -2\ln\frac{L\left(0,\hat{\hat{\theta}}\right)}{L(\hat{\mu},\hat{\theta})} & if\ \hat{\mu} \geq 0 \\ 0 & if\ \hat{\mu} < 0 \end{cases} \tag{71}$$

3.9.6 The p-value, the significance level, and the CLs technique

Claiming a discovery of a new signal after collecting data requires showing that the sample is sufficiently inconsistent with the background-only hypothesis. The test statistic t can be used as the measure of inconsistency between the data and H_0. The p-value (standing for probability value) is the probability that the test statistic t assumes a value equal or greater than the observed one from only pure background fluctuations.[161] Therefore, the smaller the p-value, the more improbable it is that the outcome comes from pure fluctuation and the more likely it comes from actual signal events. [184]

Rather than quoting the p-value, it is often easier to report the equivalent number of standard deviations excluding the central area in a normal distribution and leaving the tail that has the same p-value (Figure 3-62). In other words, it is preferred to report the '$Z\sigma$' significance corresponding to a given p-value. [161] The significance level can be calculated using this transformation:

$$p = \int_{Z}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 - \phi(Z) = \phi(-Z) \tag{72}$$

Conventionally, 3σ is referred to as evidence and 5σ as discovery. [161]

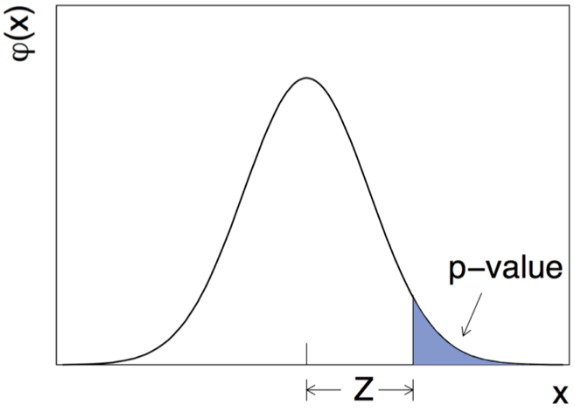


Figure 3-62: The relationship between the p-value and significance level(Z). This image is taken from [196]

As explained in section 3.9.3 , this analysis is a best limit analysis with weak signal if it exists. In other words, the statistical power is small in this analysis and the experiment does not have sensitivity to reject the H_{alt} with high power because it almost rejects H_0 as well, as seen in Figure 3. [196] The CLs technique suggests using a modified p-value that is enhanced by statistical power. Therefore, this modified p-value can be written as [198]:

$$p'_{s+b} = \frac{p_{s+b}}{1-p_b} \tag{73}$$

Where p_{s+b} and p_b are the p-values for the signal+background and background-only hypotheses. In other words, we are normalizing the p-value by the acceptance probability of H_b. [196] “In most of the cases, the probabilities p_{s+b} and p_b are not trivial to obtain analytically and are determined numerically using pseudo-experiments generated by Monte Carlo.” [161]

As [196] explains: “The CLs method lacks a frequentist coverage. However, it lacks it in places where the experiment is insensitive to the expected signal! And this is not necessarily a disadvantage from the physicists’ point of view! Here is what happens: One uses the Neyman-Pearson likelihood ratio as a test statistic. When the expected signal is very low, the two PDFs are almost overlapping (Figure 3-63). The background might fluctuate down resulting in a very small p_{s+b}. As a result, we are tempted to exclude the signal hypothesis. However, it is not the signal hypothesis that is excluded, but the signal+background hypothesis. It is the small expected signal s << s + b that is leading to a false exclusion. To protect against such an inference, one uses the modified p-value as a criterion for taking a decision of rejecting the signal hypothesis.”

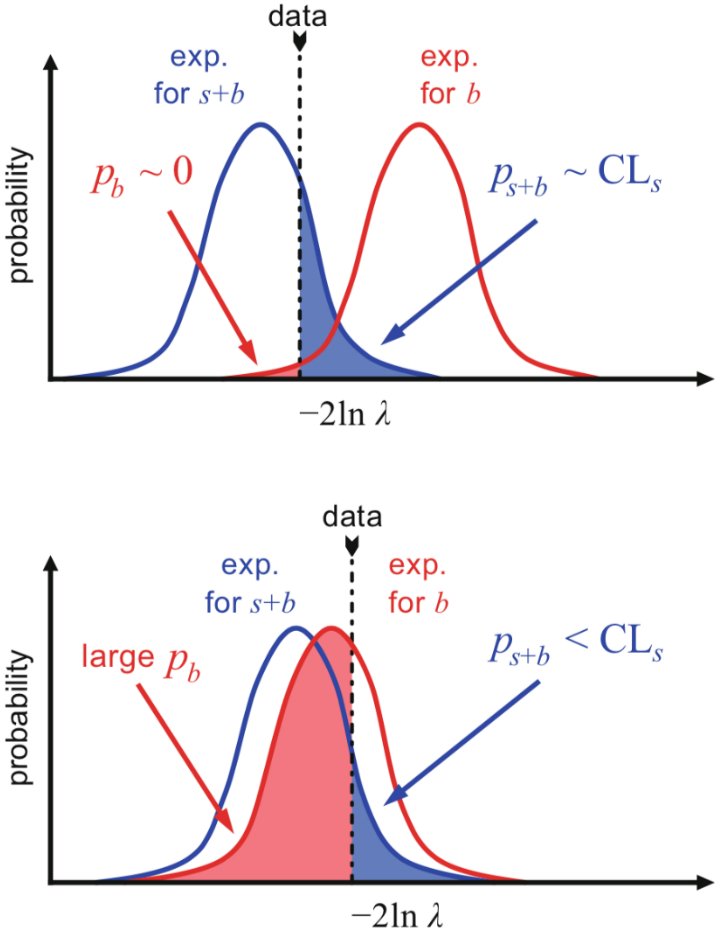


Figure 3-63: This plot shows the reasoning of the CLs method. Top plot shows two well separated distributions of the test statistic and the bottom plot shows two distributions that are largely overlapping where the experiment has poor sensitivity to the signal. This image is taken from [161]

3.9.7 Upper limits on the signal strength

We can assess the sensitivity of an experiment using a special dataset called an Asimov dataset [131], which is a dataset that all maximum likelihood estimators of all parameters are at their true values. Practically, it can be constructed from the total number of predicted signal and background events. In other words, the number of entries in the i^{th} bin is

$$p'_{s+b} = A_i(\mu, \theta) = \mu s_i(\theta) + b_i(\theta) \tag{74}$$

where A is the Asimov data histogram, and $s_i(\theta)$ and $b_i(\theta)$ are the predictions of the signal and background events respectively. And nuisance parameters θ show the systematic uncertainties. The Asimov dataset is built assuming no statistical uncertainties on the signal or background.

If CLs p-value $< \alpha$, then the signal is excluded at $100 \times (1 - \alpha)\%$ CLs confidence level. We use 95% CL that is standard at CERN. Therefore, in order to have the exclusion limit for each m_a we have to find the signal strength for which CLs p-value = 0.05.

As the Asimov dataset does not consider statistical variations, it is interesting to see how the exclusion limits change as a result of statistical variations in the data. In order to asses this, we recompute the expected limit for ± 1 and ± 2 standard deviations of the signal strength. The resulting expected limits then are added to the expected limit plot as ± 1 and ± 2 error bands. While we can use the covariance matrix to calculate the standard deviation on the signal strength, a computationally easier procedure is the approximation of $\sigma^2 = \frac{\mu^2}{q_\mu}$ where μ is the signal strength, and q_μ (eq.71) is computed using the Asimov dataset [131]. We used the Asimov dataset to refit for the upper limits. In this fit, we used the data-driven QCD background estimate described in subsection 3.6.4. Also, we used the normalization values for the Top and W backgrounds that were extracted in section 3.8.5.

The expected upper limit on the signal strength for the 3b-tag and 4b-tag categories are shown in Figure 3-64 and Figure 3-65. The results that are presented using the standard model cross section $\sigma(pp \to Wh) = 1.37\, pb^{-1}$ with $BR(h \to aa) \times BR(a \to bb)^2 = 1$. Expected upper limit means that instead of data, sum of MC background datasets is used to derive the upper limit. In other words, expected upper limit is the upper limit assuming no signal and only SM background. This expected upper limit is compared with observed upper limit derived using actual data taken at CMS. However, as we are still blind, we cannot look at the data yet. Therefore, only the expected upper limit is shown here.

After receiving permission to look at the data and adding the observed upper limit, the areas above the curve in Figure 3-64 and Figure 3-65 will be excluded at 95% confidence level. The y-axis is the signal strength defined as $\frac{\sigma(Wh\to 2a\to 4b)}{\sigma(Wh)}$. It is impossible that $\sigma(Wh \to 2a \to 4b) < \sigma(Wh)$ as $\sigma(Wh)$ is more general. However, keep in mind that we have assumed that $BR(h \to aa) \times BR(a \to bb)^2 = 1$ to derive the upper limit as we do not know the true values of the branching ratios. Therefore, our goal should be to exclude areas much lower than $\mu = \frac{\sigma_{95\%}}{\sigma_{th}} = 1$.

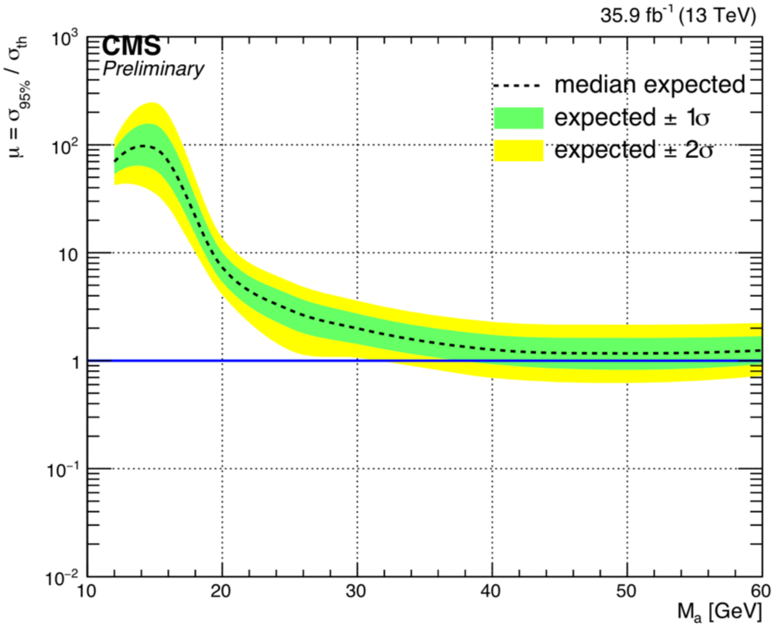


Figure 3-64: Expected limit on the signal strength when only SRs included in the fit. The solid blue line indicates the SM cross-section of $\sigma(pp \rightarrow Wh)$ with $BR(h \rightarrow aa) \times BR(a \rightarrow bb)^2 = 1$.

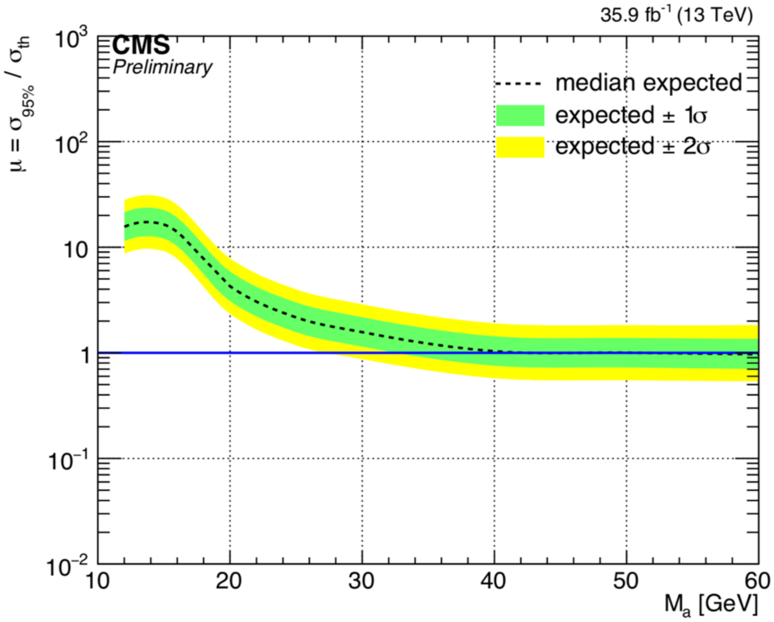


Figure 3-65: Expected limit on the signal strength with a simultaneous fit in the control and signal regions. The signal region data have been masked in the likelihood (blinded mode). The solid blue line indicates the SM cross-section $\sigma(pp \rightarrow Wh)$ with $BR(h \rightarrow aa) \times BR(a \rightarrow bb)^2 = 1$.

Figure 3-66 to Figure 3-69 show the data-MC comparison for analysis variables in the signal regions of the 3b-tag and 4b-tag categories after the fit. We can see a pretty good Data/MC agreement here. Also, Figure 3-70 and Figure 3-71 show the pre-fit and post-fit BDT shapes. The background components in these plots are prior to fitting to data. Also, as we are still blind, we are not allowed to look at the last two BDT bins which are the most sensitive ones. Therefore, these two bins are shaded.

The event yields for all the SM backgrounds and expected signal in 2016 data are shown in Table 3-22 and Table 3-23 for the 3b-tag and 4b-tag categories. The data yield in the tables match with the total background because we are working in a blind condition and data used here is the sum of MC background.

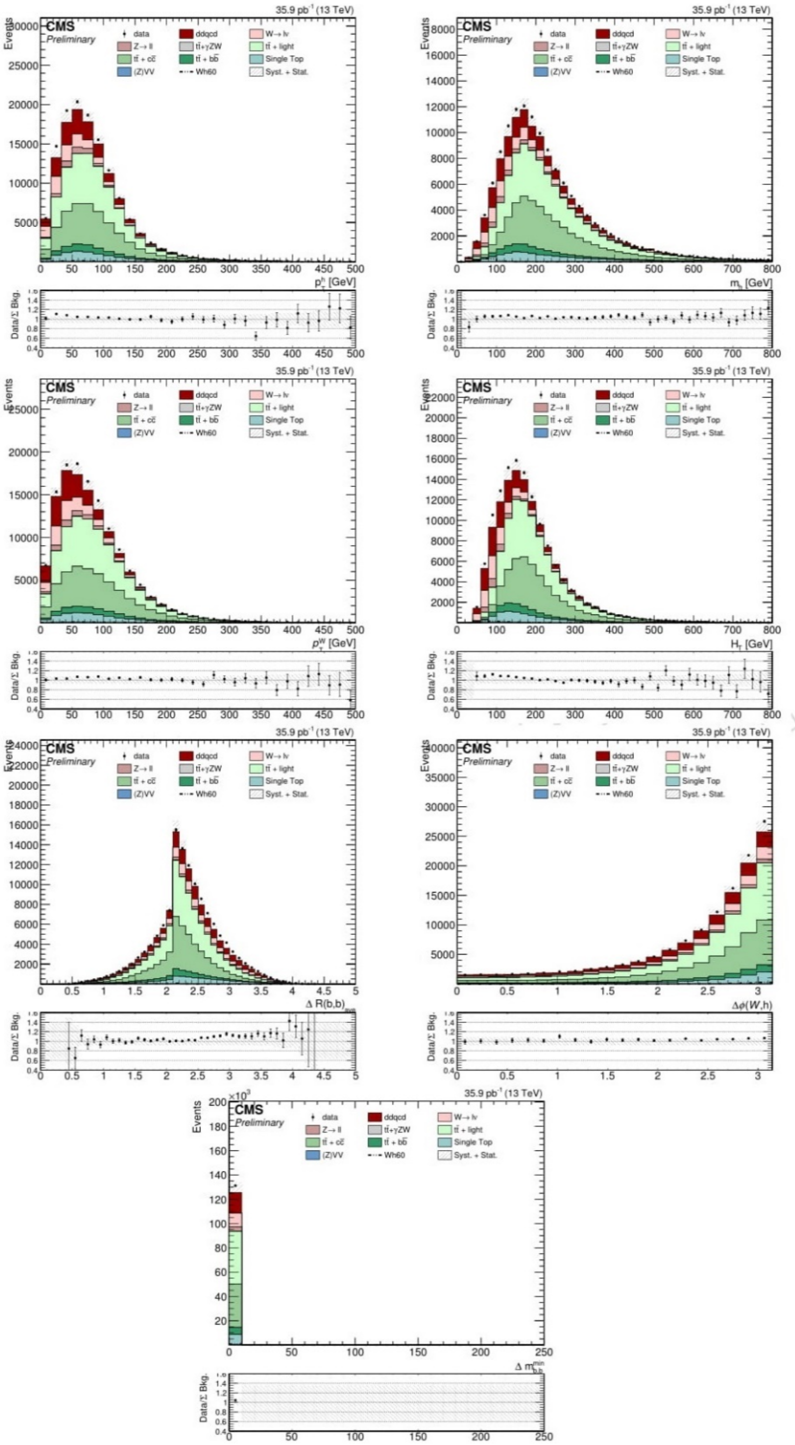


Figure 3-66: Analysis variable distributions in the Signal Regions - (3b, 3jets) bin, in the electron channel. Note that Δm_{bb}^{min} is not defined in the 3b-tag category.

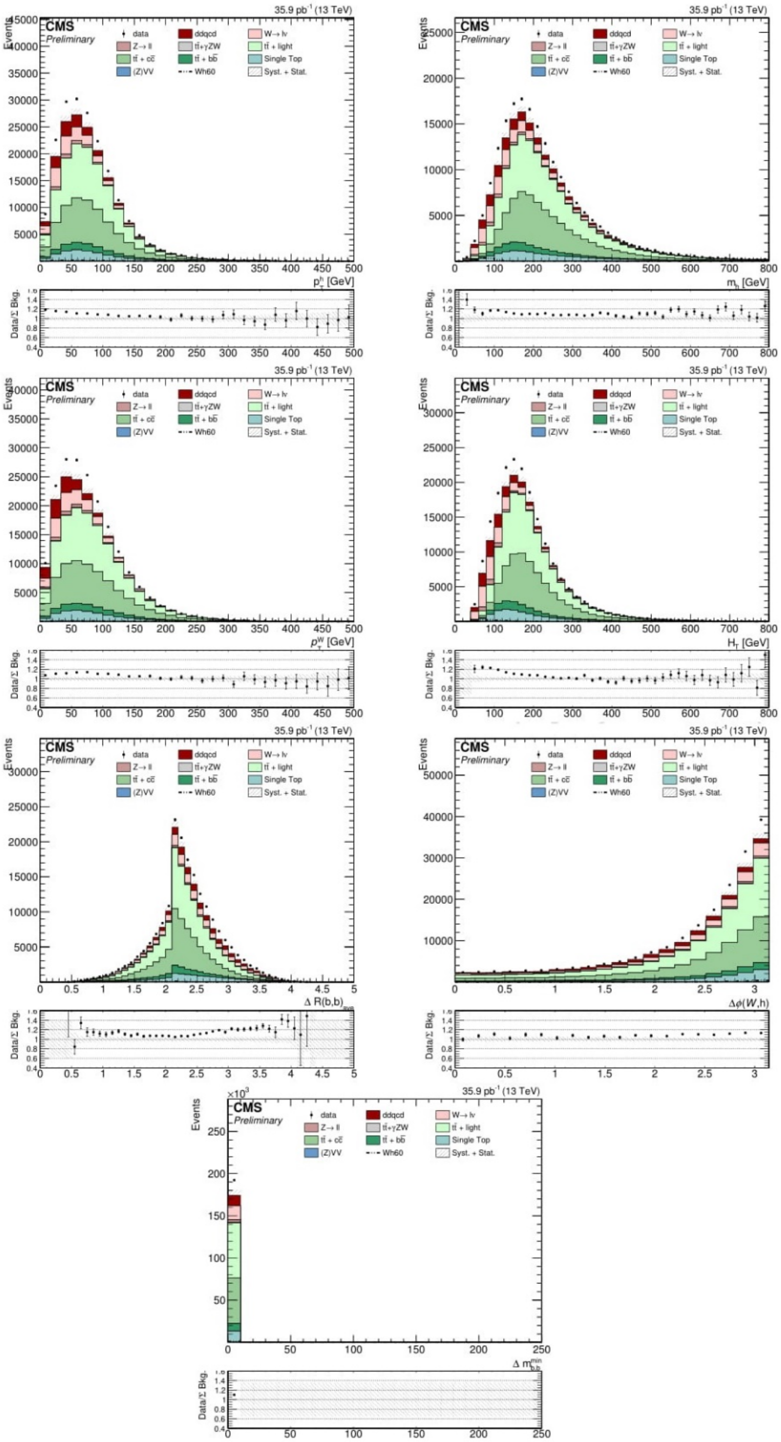


Figure 3-67: Analysis variable distributions in the Signal Regions - (3b, 3jets) bin, in the muon channel. Note that Δm_{bb}^{min} is not defined in the 3b-tag category.

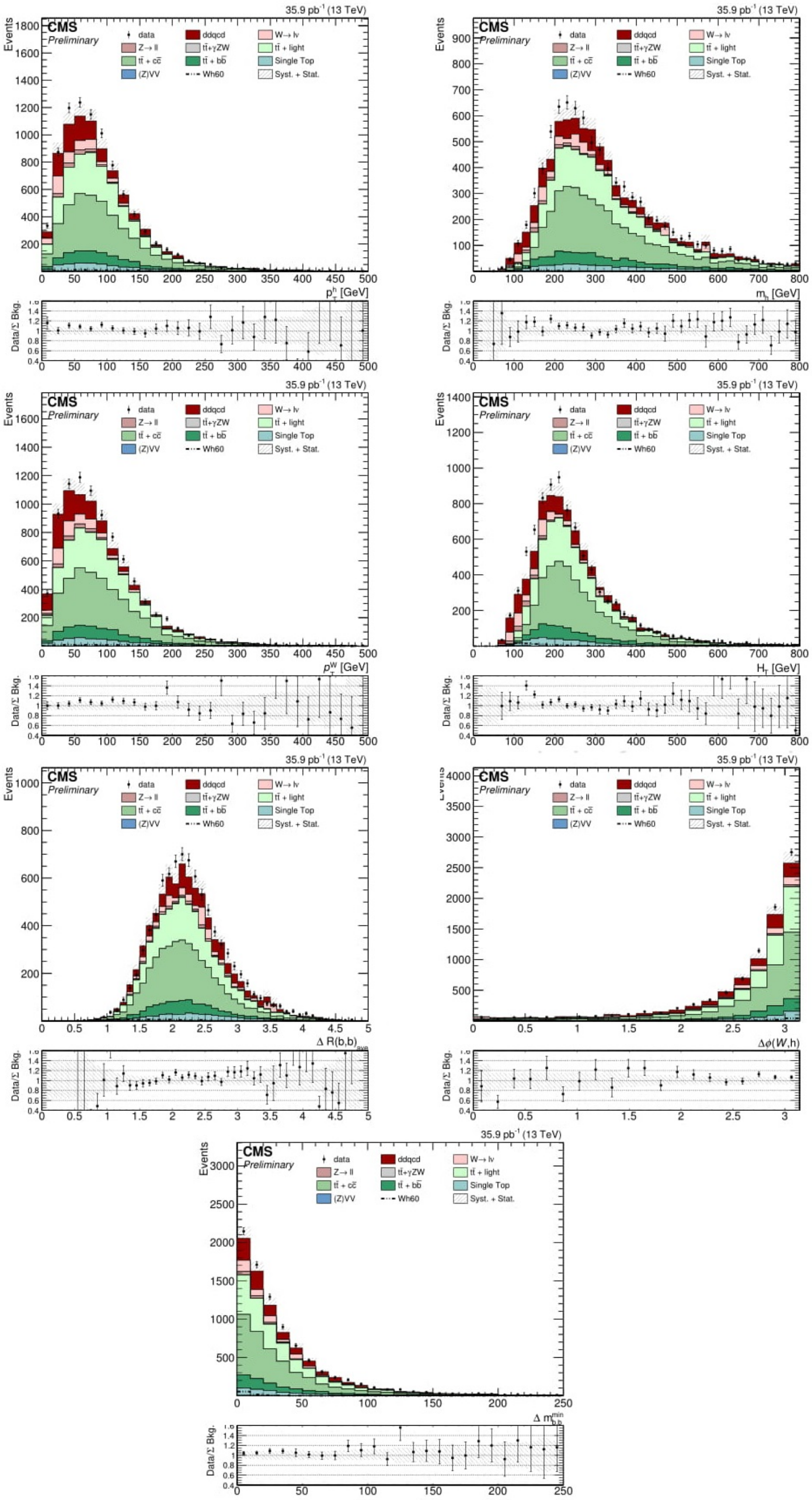


Figure 3-68: Analysis variable distributions in the Signal Regions - (4b, 4jets) bin, in the electron channel.

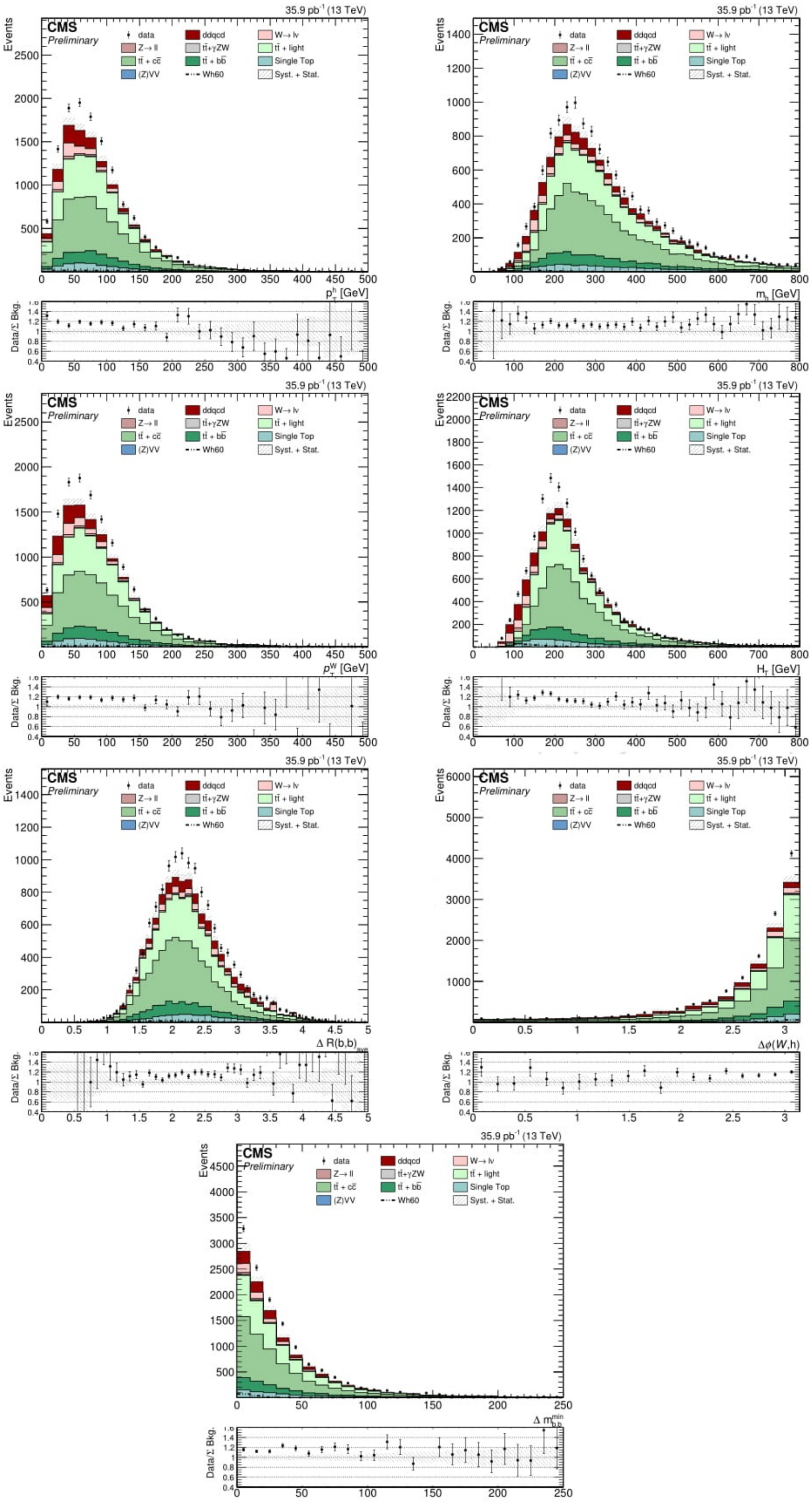


Figure 3-69: Analysis variable distributions in the Signal Regions - (4b, 4jets) bin, in the muon channel.

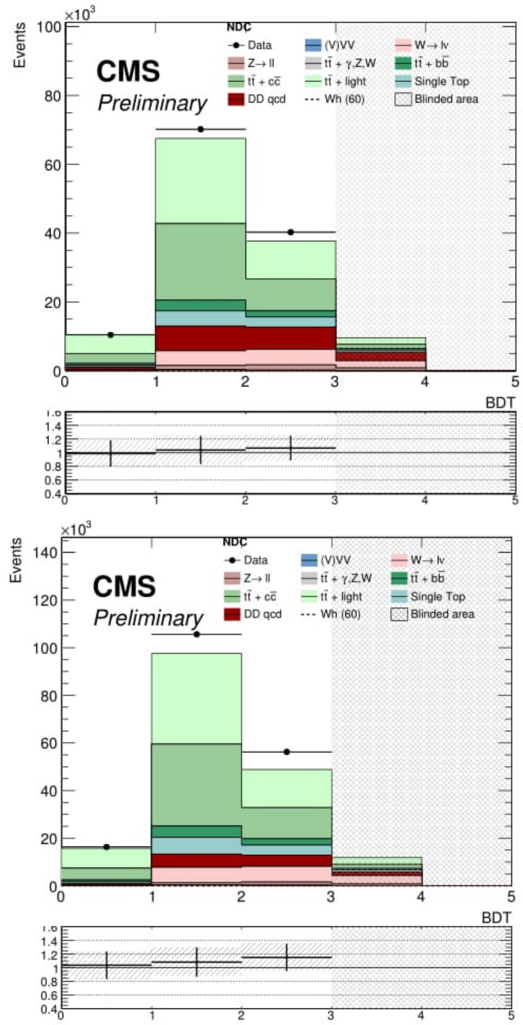


Figure 3-70: Pre-fit (left) and post-fit (right) BDT shapes in the electron (upper) and muon (lower) channels for the 3b-tag category, used as input in the final upper limit extraction. The last two bins in BDT are the most sensitive regions and are blinded here as we still do not have permission to look at data. The X-axis labels count the number of bins in BDT.

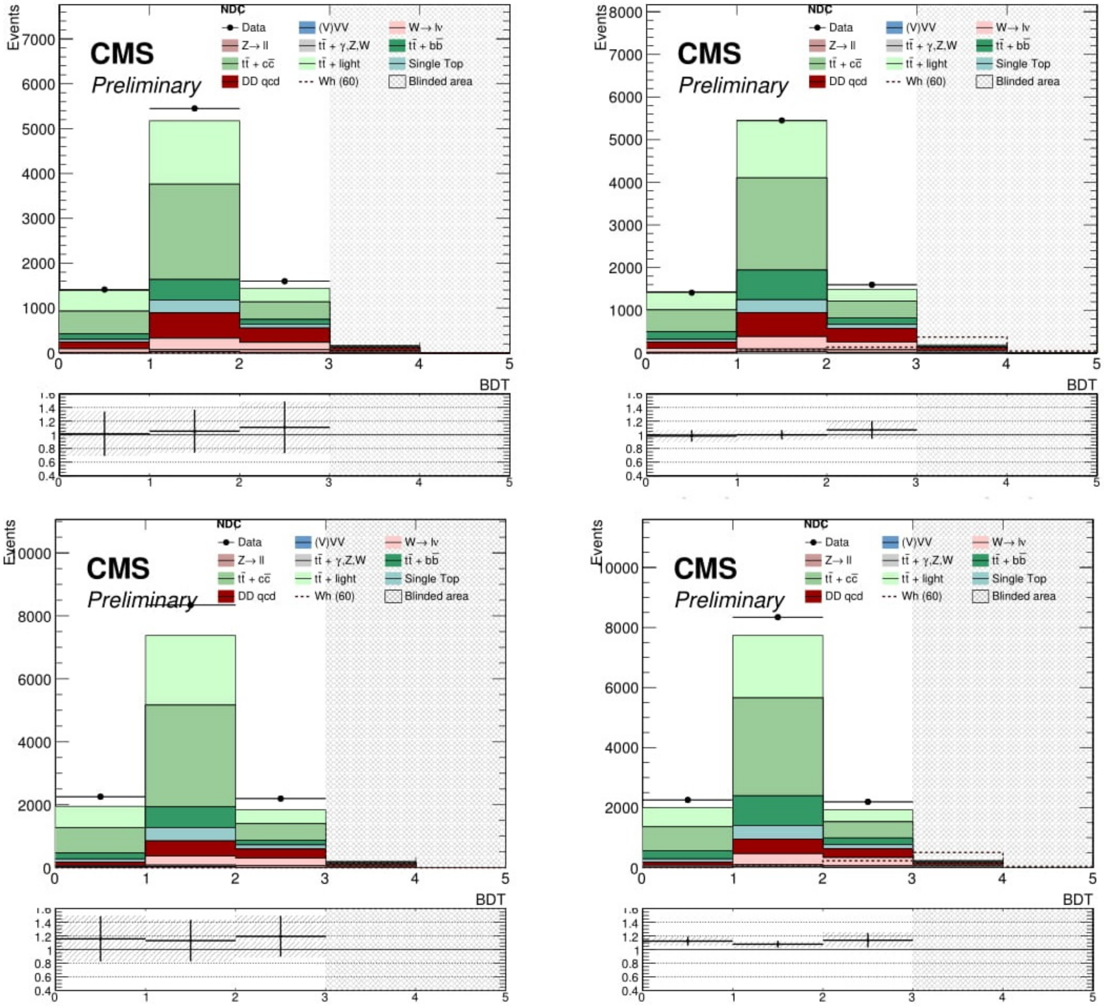


Figure 3-71: Pre-fit (left) and post-fit (right) BDT shapes in the electron (upper) and muon (lower) channels for the 4b-tag event categories, used as input in the final upper limit extraction. The last two bins in BDT are the most sensitive regions and are blinded here as we still do not have permission to look at data. The X-axis labels count the number of bins in BDT.

Table 3-22: Event yields for MC background processes and data, for the 3b-tag category. Uncertainties for backgrounds as well as for signal include the statistical and systematic components. The data yield match with the total background because we are working in a blind condition and the data have been masked in the signal region.

channel	Inclusive - 3b	muons SR - 3b	electrons SR - 3b
(V)VV	$\mathbf{593.4 \pm 14.2^{+55.3}_{-68.3}}$	$344.8 \pm 10.9^{+35.6}_{-59.3}$	$248.6 \pm 9.1^{+42.3}_{-33.9}$
Single top	$\mathbf{(213.7 \pm 0.6^{+14.8}_{-15.5}) \times 10^2}$	$(129.8 \pm 0.4^{+12.0}_{-13.2}) \times 10^2$	$(838.2 \pm 3.6^{+86.0}_{-82.5}) \times 10^1$
$t\bar{t} + b\bar{b}$	$\mathbf{(147.5 \pm 0.8^{+7.8}_{-8.0}) \times 10^2}$	$(884.8 \pm 6.0^{+62.6}_{-64.9}) \times 10^1$	$(590.4 \pm 4.9^{+46.7}_{-46.3}) \times 10^1$
$t\bar{t} + c\bar{c}$	$\mathbf{(906.2 \pm 1.9^{+56.3}_{-52.4}) \times 10^2}$	$(546.0 \pm 1.5^{+46.1}_{-43.2}) \times 10^2$	$(360.1 \pm 1.2^{+32.3}_{-29.7}) \times 10^2$
$t\bar{t} + light$	$\mathbf{(109.4 \pm 0.2^{+6.8}_{-7.0}) \times 10^3}$	$(658.1 \pm 1.6^{+56.4}_{-58.4}) \times 10^2$	$(435.8 \pm 1.3^{+37.2}_{-39.2}) \times 10^2$
$t\bar{t} + \gamma$ZW	$\mathbf{800.6 \pm 21.5^{+62.9}_{-64.7}}$	$477.0 \pm 16.6^{+34.3}_{-57.6}$	$323.5 \pm 13.6^{+52.8}_{-29.5}$
$Z \to \ell\ell$	$\mathbf{(707.2 \pm 6.1^{+99.1}_{-90.3}) \times 10^1}$	$(331.0 \pm 4.2^{+69.2}_{-55.0}) \times 10^1$	$(376.2 \pm 4.4^{+70.8}_{-71.5}) \times 10^1$
$W \to \ell\nu$	$\mathbf{(304.9 \pm 3.7^{+35.0}_{-34.0}) \times 10^2}$	$(183.4 \pm 2.7^{+29.2}_{-28.7}) \times 10^2$	$(121.5 \pm 2.5^{+19.2}_{-18.2}) \times 10^2$
dd QCD	$\mathbf{(289.7 \pm 2.5 \pm 103.8) \times 10^2}$	$(121.2 \pm 1.2 \pm 60.6) \times 10^2$	$(168.5 \pm 2.2 \pm 84.2) \times 10^2$
Total	$\mathbf{(304.0 \pm 0.5^{+10.4}_{-10.4}) \times 10^3}$	$\mathbf{(176.8 \pm 0.4^{+6.1}_{-6.1}) \times 10^3}$	$\mathbf{(127.2 \pm 0.4^{+8.4}_{-8.4}) \times 10^3}$
Data	304049	**176839**	**127210**
Wh20	$\mathbf{(125.3 \pm 1.3^{+11.3}_{-10.5}) \times 10^1}$	$766.3 \pm 10.3^{+91.4}_{-90.6}$	$487.1 \pm 8.2^{+66.8}_{-52.9}$
Wh60	$\mathbf{(180.1 \pm 1.5^{+15.2}_{-14.9}) \times 10^1}$	$(107.0 \pm 1.2^{+12.9}_{-11.5}) \times 10^1$	$731.4 \pm 9.7^{+79.6}_{-94.2}$

Table 3-23: Event yields for MC background processes and data, for the 4b-tag category. Uncertainties for backgrounds as well as for signal include the statistical and systematic components. The data yield match with the total background because we are working in a blind condition and the data have been masked in the signal region.

channel	Inclusive	muons SR - 4b	electrons SR - 4b
(V)VV	$\mathbf{31.3 \pm 3.3^{+7.7}_{-7.3}}$	$16.4 \pm 2.2^{+6.2}_{-3.3}$	$14.9 \pm 2.4^{+4.6}_{-6.5}$
Single top	$\mathbf{(110.8 \pm 1.3^{+11.1}_{-11.0}) \times 10^1}$	$662.0 \pm 10.3^{+96.6}_{-84.1}$	$446.4 \pm 8.6^{+54.4}_{-71.0}$
$t\bar{t} + b\bar{b}$	$\mathbf{(171.3 \pm 2.6^{+11.9}_{-13.0}) \times 10^1}$	$(102.3 \pm 2.0^{+10.2}_{-11.1}) \times 10^1$	$689.5 \pm 16.9^{+62.1}_{-67.9}$
$t\bar{t} + c\bar{c}$	$\mathbf{(767.5 \pm 5.6^{+60.9}_{-68.5}) \times 10^1}$	$(461.3 \pm 4.3^{+49.1}_{-57.9}) \times 10^1$	$(306.2 \pm 3.5^{+36.1}_{-36.5}) \times 10^1$
$t\bar{t} + light$	$\mathbf{(562.5 \pm 4.7^{+59.7}_{-64.0}) \times 10^1}$	$(340.1 \pm 3.7^{+50.3}_{-53.1}) \times 10^1$	$(222.4 \pm 3.0^{+32.2}_{-35.6}) \times 10^1$
$t\bar{t} + \gamma$ZW	$\mathbf{55.5 \pm 5.1^{+10.7}_{-9.0}}$	$33.2 \pm 3.9^{+6.1}_{-5.9}$	$22.3 \pm 3.3^{+8.8}_{-6.8}$
$Z \to \ell\ell$	$\mathbf{316.3 \pm 12.9^{+56.4}_{-55.5}}$	$153.0 \pm 8.9^{+43.1}_{-37.6}$	$163.4 \pm 9.3^{+36.4}_{-40.8}$
$W \to \ell\nu$	$\mathbf{(125.5 \pm 7.3^{+19.3}_{-19.0}) \times 10^1}$	$683.4 \pm 41.1^{+168.6}_{-115.3}$	$571.1 \pm 60.1^{+93.8}_{-151.6}$
dd QCD	$\mathbf{(202.3 \pm 6.4 \pm 71.7) \times 10^1}$	$951.4 \pm 37.2 \pm 475.7$	$(107.2 \pm 5.2 \pm 53.6) \times 10^1$
Total	$\mathbf{(198.0 \pm 1.3^{+7.2}_{-7.2}) \times 10^2}$	$\mathbf{(115.4 \pm 0.8^{+4.8}_{-4.8}) \times 10^2}$	$\mathbf{(826.6 \pm 9.4^{+53.6}_{-53.6}) \times 10^1}$
Data	19802	**11537**	**8265**
Wh20	$\mathbf{94.9 \pm 3.7^{+12.9}_{-10.5}}$	$60.1 \pm 2.9^{+8.7}_{-9.5}$	$34.8 \pm 2.2^{+9.5}_{-4.3}$
Wh60	$\mathbf{224.8 \pm 5.4^{+24.1}_{-24.4}}$	$135.2 \pm 4.2^{+18.3}_{-21.4}$	$89.5 \pm 3.4^{+15.6}_{-11.8}$

The impact of a nuisance parameter (NP) θ on the parameter of interest (POI) μ is defined as the shift $\Delta\mu$ that is induced as θ is fixed and brought to its $+1\sigma$ or -1σ post-fit values while all other parameters profiled as normal. This is effectively a measure of the correlation between the NP and the POI and is useful to determine which NPs have the largest effect on the uncertainty of the POI. For the sake of comparison, we first show the impacts and pulls on the signal strength for the case where only the signal regions are present in the likelihood fit. This is shown in Figure 3-72 for the limit extraction with a mass signal point $m_a = 60\ GeV$ and the 3b-tag and 4b-tag categories combined. The corresponding impacts and pulls for the simultaneous fit with control regions included are shown in Figure 3-73. We observe that in case of the simultaneous fit including the control regions, nuisances related to the JES and uMET are overconstrainted. A decorrelation of the JES uncertainties into six subcomponents depending on the jet p_T and eta has been implemented as per the JetMET group prescription [199]. The nuisance related to the unclustered MET shows the highest impact and is also highly overconstrained. The reason for this has been shown to come from the overestimation of this uncertainty while using the old-style recipe for smearing the unclustered part in E_T^{miss} using a 10% resolution for the objects. Nuisances related to the systematic uncertainty on the fake-lepton background estimation are also slightly overconstrainted due to the conservative 50% uncertainty assigned on the fake-lepton predicted yields.

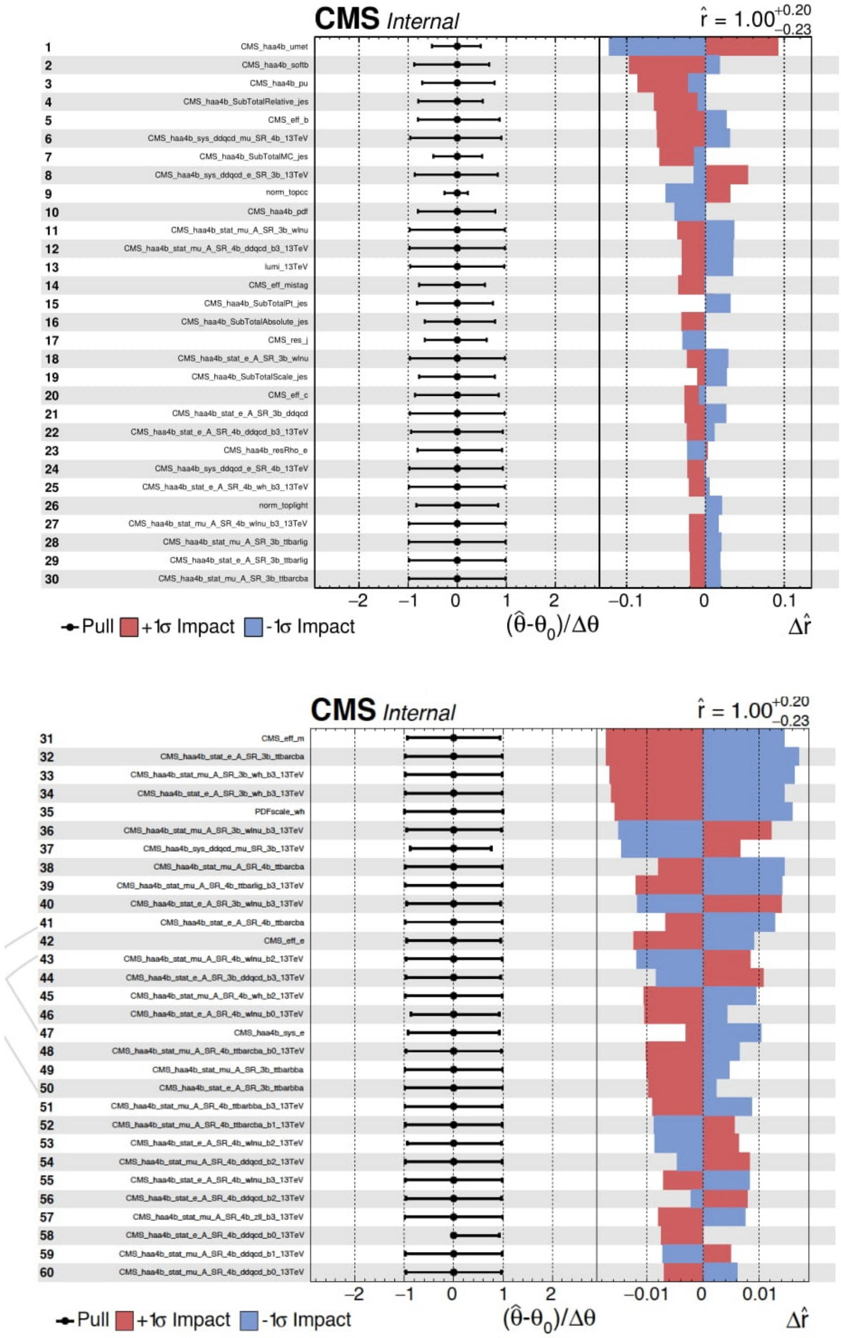


Figure 3-72: Impacts and pulls on signal strength for the systematic sources ranked in terms of importance. The Asimov S+B fit in the Signal Regions only with $m_a = 60\ GeV$ signal hypothesis is used.

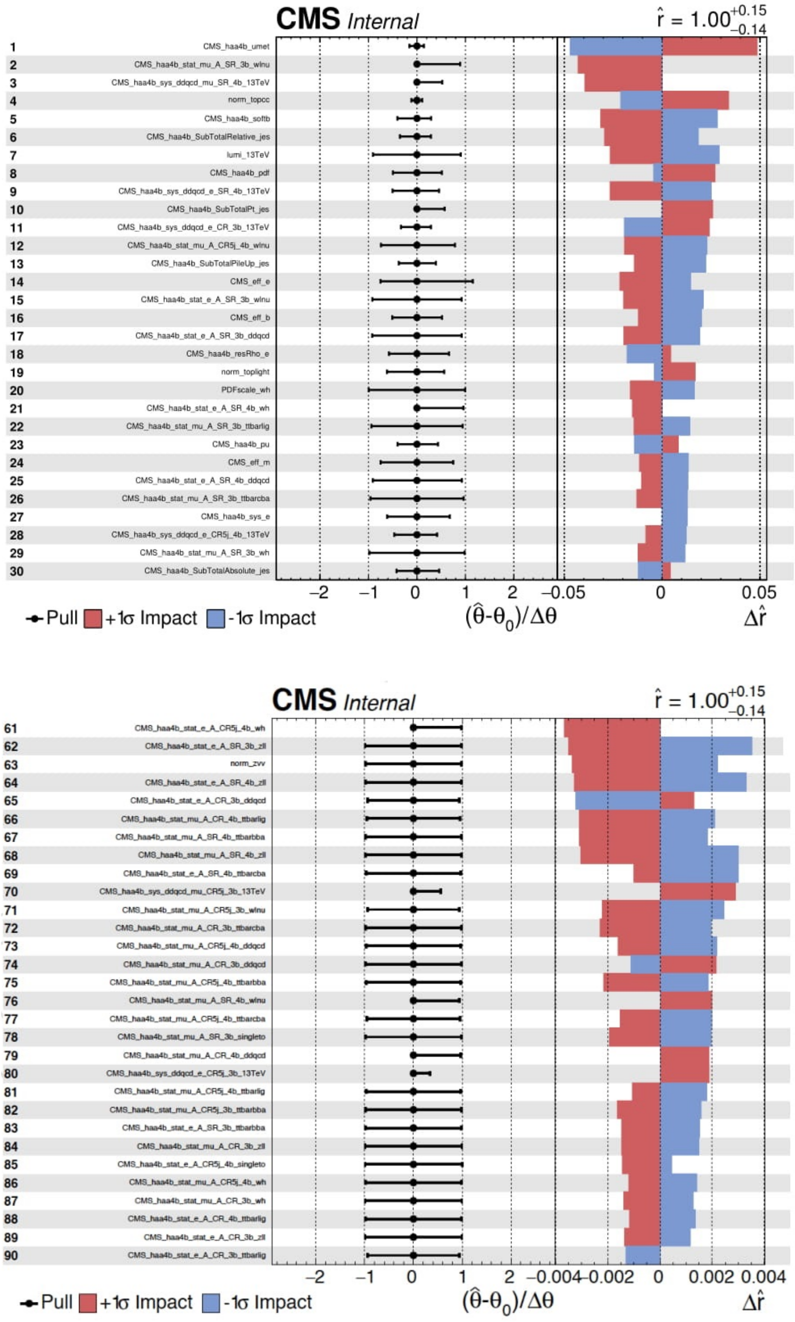


Figure 3-73: Impacts and pulls on signal strength for the systematic sources ranked in terms of importance. The Asimov S+B simultaneous fit in the Control Regions with $m_a = 60$ GeV signal hypothesis is used.

3.10 Summary

In this dissertation, I explained the analysis documented in AN2018_154 using 35.9 fb^{-1} of 13 TeV center of mass pp collision data recorded in 2016 by the Compact Muon Solenoid experiment at the Large Hadron Collider. The goal of this analysis is to put an upper limit on the signal strength of $h \rightarrow 2a \rightarrow 4b$ decay predicted by a few well-motivated beyond-standard-model theories. Here, a is a lighter pseudoscalar Higgs.

In order to isolate the signal, first a few pre-selection requirements were applied to suppress SM background events (section 3.4.2.). Then we trained a BDT multivariable discriminator to be able to efficiently distinguish between the remaining background and signal events. And finally, this binned BDT distribution was used by the Higgs Combine tool to put an upper limit on the signal strength of the $h \rightarrow 2a \rightarrow 4b$ decay.

The main backgrounds in this analysis are $t\bar{t} + X$ and W + jets. We defined and used control regions to estimate their shape and yield in the signal region. The shapes of the $t\bar{t} + X$ and W + jets backgrounds were taken from MC while their normalizations were taken from data using a simultaneous fit in the signal and control regions. Another important background is QCD. In this analysis, we suffered from poor QCD MC statistics. Therefore, QCD background estimation as well as QCD shape and yield in the control regions were derived from the so called ABCD method.

Unfortunately, at the time of writing this dissertation, we are still not allowed to look at the data (we are blind). Therefore, the upper limit plot in this dissertation lacks the observed curve. Interested reader can find the updated plot with observed curve in AN2018_154.

Bibliography

1. Barlow, D.J.T.E.D., *The Handy Physics Answer Book*. 2011. **76**(9): p. 61.
2. Van Melsen, A.G., *From atomos to atom: The history of the concept atom*. 2004: Courier Corporation.
3. Einstein, A.J.A.d.p., *On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat*. 1905. **17**: p. 549-560.
4. Thomson, J.J.J.T.L., Edinburgh,, D.P. Magazine, and J.o. Science, *XL. Cathode rays*. 1897. **44**(269): p. 293-316.
5. Rutherford, E.J.T.L., Edinburgh,, D.P. Magazine, and J.o. Science, *LXXIX. The scattering of α and β particles by matter and the structure of the atom*. 1911. **21**(125): p. 669-688.
6. Chadwick, J.J.P.R.S.L.A., *The existence of a neutron*. 1932. **136**(830): p. 692-708.
7. Dissertori, G., I.G. Knowles, and M. Schmelling, *Quantum chromodynamics: high energy experiments and theory*. Vol. 115. 2003: Oxford University Press.
8. Woit, P., Woit, and Bartolini, *Quantum Theory, Groups and Representations*. 2017: Springer.
9. Martin, B.R. and G. Shaw, *Particle physics*. 2017: John Wiley & Sons.
10. Starkman, G. *The Standard Model of particle physics: The absolutely amazing theory of almost everything*. 2018; Available from: http://theconversation.com/the-standard-model-of-particle-physics-the-absolutely-amazing-theory-of-almost-everything-94700.
11. Burgess, C. and G. Moore, *The standard model: A primer*. 2006: Cambridge University Press.
12. Yukawa, H.J.P.o.t.P.-M.S.o.J.r.S., *On the interaction of elementary particles. I*. 1935. **17**: p. 48-57.
13. Wikipedia_Contributors, "*Strong interaction*, in *Wikipedia, The Free Encyclopedia*. 2018.
14. Aaij, R., et al., *Observation of J/ψ p Resonances Consistent with Pentaquark States in Λ b 0→ J/ψ K− p Decays*. 2015. **115**(7): p. 072001.
15. De Angelis, A. and M.J.M. Pimenta, *Introduction to particle and astroparticle physics: multimessenger astronomy and its particle physics foundations*. 2018: Springer.
16. Narison, S., *QCD as a theory of hadrons: from partons to confinement*. Vol. 17. 2004: Cambridge University Press.
17. Boelaert, N., *Dijet Angular Distributions in Proton-Proton Collisions: At√ s= 7 TeV and√ s= 14 TeV*. 2011: Springer Science & Business Media.
18. Paschos, E.A., *Electroweak theory*. 2007: Cambridge University Press.
19. Boyarkin, O.M., *Introduction to physics of elementary particles*. 2007: Nova Publishers.

20. Weber, H., *Search for Supersymmetry in Hadronic Final States: Evolution Studies of the CMS Electromagnetic Calorimeter*. Springer Theses. 2015: Springer. 213.
21. Wikipedia_contributors, *Electroweak interaction*, in *Wikipedia, The Free Encyclopedia*. 2018.
22. Thomson, M., *Modern particle physics*. 2013: Cambridge University Press.
23. Greiner, W., B. Müller, and D.A. Bromley, *Gauge theory of weak interactions*. Vol. 5. 1996: Springer.
24. Baggott, J., *Higgs: the invention and discovery of the'God Particle'*. 2012: Oxford University Press.
25. Mandl, F. and G. Shaw, *Quantum field theory*. 2010: John Wiley & Sons.
26. Bettini, A., *Introduction to elementary particle physics*. 2008.
27. Wikipedia_contributors, *Parton (particle physics)*, in *Wikipedia, The Free Encyclopedia*. 2018.
28. Kovchegov, Y.V. and E. Levin, *Quantum chromodynamics at high energy*. 2012: Cambridge University Press.
29. Höche, S., *Introduction to parton-shower event generators*, in *Journeys Through the Precision Frontier: Amplitudes for Colliders: TASI 2014 Proceedings of the 2014 Theoretical Advanced Study Institute in Elementary Particle Physics*. 2016, World Scientific. p. 235-295.
30. Sjöstrand, T., S. Mrenna, and P.J.J.o.H.E.P. Skands, *PYTHIA 6.4 physics and manual*. 2006. **2006**(05): p. 026.
31. Placakyte, R.J.a.p.a., *Parton Distribution Functions*. 2011.
32. Ellis, R.K., W.J. Stirling, and B.R. Webber, *QCD and collider physics*. 2003: Cambridge university press.
33. The_CMS_collaboration. *Summaries of CMS cross section measurements*. 2018; Available from: https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined.
34. Martin, S.P., *A supersymmetry primer*, in *Perspectives on supersymmetry*. 1998, World Scientific. p. 1-98.
35. WK, C.H., *Lepton And Photon Interactions At High Energies: Lepton-photon 2003-Proceedings Of The Xxi International Symposium*. 2004: World Scientific.
36. Wikipedia_contributors, *Physics beyond the Standard Model*, in *Wikipedia, The Free Encyclopedia*. 2018.
37. Mazure, A. and V. Le Brun, *Matter, Dark Matter, and Anti-matter: In Search of the Hidden Universe*. 2011: Springer Science & Business Media.
38. Dimopoulos, S., S. Raby, and F.J.P.R.D. Wilczek, *Supersymmetry and the scale of unification*. 1981. **24**(6): p. 1681.
39. Pralavorio, P.J.a.p.a., *Particle Physics and Cosmology*. 2013.
40. Wikipedia_contributors, *Naturalness (physics)*. 2018.
41. Vempati, S.K.J.a.p.a., *Introduction to MSSM*. 2012.
42. Zeune, L., *Constraining Supersymmetric Models: Using Higgs Physics, Precision Observables and Direct Searches*. 2015: Springer.

43. Miller, D. and R.J.a.p.h.-p. Nevzorov, *The Peccei-Quinn axion in the next-to-minimal supersymmetric standard model*. 2003.
44. Ricci-Tam, F., *Search for New Light Higgs Bosons in Boosted Tau Final States with the CMS Experiment*. 2016, UC, Davis.
45. Wikipedia_contributors, *Next-to-Minimal Supersymmetric Standard Model*, in *Wikipedia, The Free Encyclopedia*. 2018.
46. Maniatis, M.J.I.J.o.M.P.A., *The next-to-minimal supersymmetric extension of the standard model reviewed*. 2010. **25**(18n19): p. 3505-3602.
47. Nakamura, K., P.D.G.J.J.o.P.G. Nuclear, and P. Physics, *Review of particle physics*. 2010. **37**(7A): p. 075021.
48. Weinberg, S.J.P.R.L., *A new light boson?* 1978. **40**(4): p. 223.
49. Wilczek, F.J.P.R.L., *Problem of Strong p and t Invariance in the Presence of Instantons*. 1978. **40**(5): p. 279.
50. Mou, Q., H. Wu, and S.J.a.p.a. Zheng, *Light Neutralino Dark Matter in the NMSSM*. 2017.
51. Demidov, S., D. Gorbunov, and D.J.J.o.H.E.P. Kirpichnikov, *Split NMSSM with electroweak baryogenesis*. 2016. **2016**(11): p. 148.
52. Wikipedia_contributors, *Baryogenesis*, in *Wikipedia, The Free Encyclopedia*. 2018.
53. Chao, A. and W. Chou, *Reviews of accelerator science and technology*. Vol. 1. 2008: World Scientific.
54. Evans, L.R., *The Large Hadron Collider: a marvel of technology*. 2009: EPFL Press.
55. Airapetian, A., et al., *ATLAS detector and physics performance: Technical Design Report, 1*. 1999: ATLAS-TDR-014.
56. Allday, J., *Quarks, leptons and the big bang*. 2016: CRC Press.
57. Collaboration, C.J.C.l., *Technical proposal*. 1994. **94**: p. 38.
58. George, K.J.C.J.P., *The LHCb Experiment*. 2003. **54**(LHCb-2003-130): p. A437-A440.
59. Avati, V., et al., *The TOTEM experiment*. 2004. **34**(1): p. s255-s268.
60. Adriani, O., et al., *The LHCf detector at the CERN large hadron collider*. 2008. **3**(08): p. S08006.
61. in *BBC News*. 2016.
62. Evans, L. and P.J.J.o.i. Bryant, *LHC machine*. 2008. **3**(08): p. S08001.
63. Myers, S.J.I.J.o.M.P.A., *The large hadron collider 2008–2013*. 2013. **28**(25): p. 1330035.
64. Rabbertz, K.J.S.T.M.P., *Jet Physics at the LHC: The Strong Force beyond the TeV Scale*. 2017. **268**: p. pp. 1-214.
65. ; Available from: https://indico.cern.ch/event/518474/contributions/1198681/attachments/1267581/1877122/Calorimetry-lecture-to-Southampton-students-4May2016-Cockerill-compressed.pdf.

66. Rossi, L., *The High Luminosity LHC Project*, in *THE FUTURE OF OUR PHYSICS INCLUDING NEW FRONTIERS: Proceedings of the International School of Subnuclear Physics*. 2017, World Scientific. p. 371-391.
67. Sirunyan, A.M., et al., *Particle-flow reconstruction and global event description with the CMS detector*. 2017. **12**(10): p. 3.
68. CMS collaboration %J CERN, C.-L.-.-L.-P.-. *Technical proposal for the phase-II upgrade of the CMS detector*. 2015.
69. Anelli, G., et al., *The totem experiment at the cern large hadron collider*. 2008. **3**(08): p. S08007.
70. Bayatian, G., et al., *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. 2006, CMS-TDR-008-1.
71. Breskin, A. and R. Voss, *The CERN large hadron collider: accelerator and experiments*. 2009: Cern Geneva.
72. Chatrchyan, S. and E.J.J.o.i.-B. de Wolf, , currens, *The CMS experiment at the CERN LHC*. 2008. **3**: p. S08004, 1-S08004, 361.
73. McLean, C.A., *Search for High-Mass Top Quark Pair Resonances with the CMS Experiment*. 2018, University of California, Davis.
74. CMS collaboration %J Technical Design Report CMS. CERN, G., *CMS technical design report for the pixel detector upgrade*. 2012.
75. Instrumentation, C.c.J.J.o., *Description and performance of track and primary-vertex reconstruction with the CMS tracker*. 2014. **9**(10): p. P10009.
76. Heller, R.E., *A search for supersymmetry with the CMS detector in the single-lepton final state using the sum of masses of large-radius jets*. 2018, UC, Santa Barbara.
77. Annenkov, A., et al., *Lead tungstate scintillation material*. 2002. **490**(1-2): p. 30-50.
78. Green, D., *The physics of particle detectors*. Vol. 12. 2000: Cambridge University Press.
79. Graham, D.J. and C. Seez, *Simulation of Longitudinal Light Collection Uniformity in PbWO4 crystals*. 1996.
80. Lee, S., et al., *On the limits of the hadronic energy resolution of calorimeters*. 2018. **882**: p. 148-157.
81. S08004, C.c.J., *The cms experiment at the cern lhc. JINST 3*. 2008. **10**: p. 1748-0221.
82. TC collaboration %J CERN, G., *The CMS hadron calorimeter project: Technical Design Report. Technical Design Report CMS*. 1997.
83. Cadamuro, L.J.J.o.I., *The CMS Level-1 trigger system for LHC Run II*. 2017. **12**(03): p. C03021.
84. Speer, T., et al., *Track reconstruction in the CMS tracker*. 2006. **559**(1): p. 143-147.
85. Cucciarelli, S., et al., *Track reconstruction, primary vertex finding and seed generation with the Pixel Detector*. 2006, CERN-CMS-NOTE-2006-026.
86. Curtin, D., et al., *Exotic decays of the 125 GeV Higgs boson*. 2014. **90**(7): p. 075004.

87. Chatrchyan, S., et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*. 2012. **716**(1): p. 30-61.
88. Aad, G., et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*. 2012. **716**(1): p. 1-29.
89. Collaboration, A. and T. collaboration. *Updated coupling measurements of the Higgs boson with the ATLAS detector using up to 25 fb− 1 of proton-proton collision data*. 2014. ATLAS-CONF-2014-009.
90. collaboration, C., *Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV*. 2013, CMS-PAS-HIG-13-005.
91. B, A.c.J.P.L., *Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC*. 2013. **726**(1): p. 88-119.
92. Allbrooke, B., et al., *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√ s= 7 and 8 TeV*. 2016. **45**.
93. Beluffi, C., et al., *Search for a standard model-like Higgs boson in the mu+ mu- and e+ e-decay channels at the LHC*. 2015: p. 184.
94. collaboration, C., *Search for the exotic decay of the Higgs boson to two light pseudoscalar bosons with two taus and two muons in the final state at√ s= 8 TeV*. 2015, CMS-PAS-HIG-15-011.
95. Chang, S., et al., *Nonstandard Higgs boson decays*. 2008. **58**: p. 75-98.
96. *Higgs Exotic Decay*. 2018 2018-09-12]; Available from: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWGExoticDecay.
97. Craig, N., J. Galloway, and S.J.a.p.a. Thomas, *Searching for signs of the second Higgs doublet*. 2013.
98. Grinstein, B. and P.J.J.o.H.E.P. Uttayarat, *Carving out parameter space in type-II two Higgs doublets model*. 2013. **2013**(6): p. 94.
99. Chen, C.-Y., S. Dawson, and M.J.P.R.D. Sher, *Heavy Higgs searches and constraints on two Higgs doublet models*. 2013. **88**(1): p. 015018.
100. Belanger, G., et al., *Status of invisible Higgs decays*. 2013. **723**(4-5): p. 340-347.
101. Coleppa, B., F. Kling, and S.J.J.o.H.E.P. Su, *Constraining type II 2HDM in light of LHC Higgs searches*. 2014. **2014**(1): p. 161.
102. Baglio, J., et al., *Benchmarks for Higgs boson pair production and heavy Higgs boson searches in the two-Higgs-doublet model of type II*. 2014. **90**(1): p. 015008.
103. Wang, L. and X.-F.J.J.o.H.E.P. Han, *Status of the aligned two-Higgs-doublet model confronted with the Higgs data*. 2014. **2014**(4): p. 128.
104. Celis, A., V. Ilisie, and A.J.J.o.H.E.P. Pich, *LHC constraints on two-Higgs doublet models*. 2013. **2013**(7): p. 53.
105. Christensen, N., et al., *Low-mass Higgs bosons in the NMSSM and their LHC implications*. 2013. **2013**(8): p. 19.
106. King, S.F., et al., *Natural NMSSM Higgs bosons*. 2013. **870**(2): p. 323-352.
107. Cerdeno, D.G., P. Ghosh, and C.B.J.J.o.H.E.P. Park, *Probing the two light Higgs scenario in the NMSSM with a low-mass pseudoscalar*. 2013. **2013**(6): p. 31.

108. Cao, J., et al., *A light Higgs scalar in the NMSSM confronted with the latest LHC Higgs data*. 2013. **2013**(11): p. 18.
109. Ahriche, A., A. Arhrib, and S.J.J.o.H.E.P. Nasri, *Higgs phenomenology in the two-singlet model*. 2014. **2014**(2): p. 42.
110. Chalons, G. and F.J.P.R.D. Domingo, *Analysis of the Higgs potentials for two doublets and a singlet*. 2012. **86**(11): p. 115024.
111. Karapostoli, G. *Searches for non-Standard Model decays of the Higgs to two light bosons in ATLAS and CMS*. in *The International Conference on Supersymmetry and Unification of Fundamental Interactions* 2018. Barcelona, Spain.
112. Bernon, J., B. Dumont, and S.J.P.R.D. Kraml, *Status of Higgs couplings after run 1 of the LHC*. 2014. **90**(7): p. 071301.
113. Belanger, G., et al., *Global fit to Higgs signal strengths and couplings and implications for extended Higgs sectors*. 2013. **88**(7): p. 075008.
114. Khachatryan, V., et al., *Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs*. 2014. **736**: p. 64-85.
115. Heng, Z., L. Shang, and P.J.J.o.H.E.P. Wan, *Pair production of a 125 GeV Higgs boson in MSSM and NMSSM at the ILC*. 2013. **2013**(10): p. 47.
116. Sirunyan, A.M., et al., *Search for a massive resonance decaying to a pair of Higgs bosons in the four b quark final state in proton–proton collisions at s= 13TeV*. 2018. **781**: p. 244-269.
117. Khachatryan, V., et al., *Search for heavy resonances decaying to two Higgs bosons in final states containing four b quarks*. 2016. **76**(7): p. 371.
118. Aaboud, M., et al., *Search for the Higgs boson produced in association with a vector boson and decaying into two spin-zero particles in the H⟶ aa⟶ 4 b channel in pp collisions at $$\sqrt {s}= 13$$ TeV with the ATLAS detector*. 2018. **2018**(10): p. 31.
119. Aaboud, M., et al., *Search for pair production of Higgs bosons in the b b¯ b b¯ final state using proton-proton collisions at s= 13 TeV with the ATLAS detector*. 2016. **94**(5): p. 052002.
120. Aad, G., et al., *Search for Higgs boson pair production in the $$ b\bar {b} b\bar {b} $$ b b¯ b b¯ final state from pp collisions at $$\sqrt {s}= 8$$ s= 8 TeVwith the ATLAS detector*. 2015. **75**(9): p. 412.
121. Aad, G., et al., *Searches for Higgs boson pair production in the h h⟶ b b τ τ, γ γ W W*, γ γ b b, b b b b channels with the ATLAS detector*. 2015. **92**(9): p. 092004.
122. Aaboud, M., et al., *Search for the Higgs boson produced in association with a W boson and decaying to four b-quarks via two spin-zero particles in pp collisions at 13 TeV with the ATLAS detector*. 2016. **76**(11): p. 605.
123. Chatrchyan, S., et al., *Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks*. 2014. **89**(1): p. 012003.
124. Sirunyan, A.M., et al., *Observation of Higgs boson decay to bottom quarks*. 2018. **121**(12): p. 121801.
125. Aaboud, M., et al., *Observation of H⟶ bb¯ decays and VH production with the ATLAS detector*. 2018. **786**: p. 59-86.

126. Campbell, J.M., R.K. Ellis, and C.J.J.o.H.E.P. Williams, *Vector boson pair production at the LHC*. 2011. **2011**(7): p. 18.
127. Alwall, J., et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*. 2014. **2014**(7): p. 79.
128. Sjöstrand, T., S. Mrenna, and P.J.C.P.C. Skands, *A brief introduction to PYTHIA 8.1*. 2008. **178**(11): p. 852-867.
129. Nason, P., B.J.A.R.o.N. Webber, and P. Science, *Next-to-leading-order event generators*. 2012. **62**: p. 187-213.
130. Alioli, S., et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*. 2010. **2010**(6): p. 43.
131. Agostinelli, S., et al., *GEANT4—a simulation toolkit*. 2003. **506**(3): p. 250-303.
132. Alwall, J., et al., *MadGraph 5: going beyond*. 2011. **2011**(6): p. 128.
133. BenjF. 2018; Available from: http://feynrules.irmp.ucl.ac.be/attachment/wiki/NMSSM/p1_ufo.tgz.
134. Dittmaier, S., et al., *Handbook of LHC Higgs cross sections: 1. Inclusive observables*. 2011.
135. Dittmaier, S., et al., *Handbook of LHC Higgs cross sections: 2. Differential distributions*. 2012.
136. Heinemeyer, S., et al., *Handbook of LHC Higgs cross sections: 3. Higgs properties*. 2013.
137. Fartoukh, S.J.P.R.S.T.-A. and Beams, *Pile up management at the high-luminosity LHC and introduction to the crab-kissing concept*. 2014. **17**(11): p. 111001.
138. Gardi, E., N. Glover, and A. Robson, *Lhc Phenomenology*. 2015: Springer.
139. Adam, W., et al., *Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC*. 2005. **31**(9): p. N9.
140. Twiki_page. *MET Filter Recommendations for Run II*. 2018 [cited 2018; Available from: https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFiltersRun2.
141. Twiki_page. *Beam Halo in CMS*. 2018 [cited 2018; Available from: https://twiki.cern.ch/twiki/bin/view/Sandbox/MarcoCardaciSandbox.
142. Nghiem, P., et al., *Beam halo definitions and its consequences*. 2012.
143. Orfanelli, S., et al., *A novel Beam Halo Monitor for the CMS experiment at the LHC*. 2015. **10**(11): p. P11011.
144. arXiv:.02702, C.c.J.a.p., *Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at sqrt (s)= 8 TeV*. 2015.
145. Iiyama, Y., *Search for Supersymmetry in pp Collisions at√ s= 8 TeV with a Photon, Lepton, and Missing Transverse Energy*. 2017: Springer.
146. Sani, M. *New Cut-Based New Cut-Based Electron Id Electron Id*.
147. Twiki_page. *Cut Based Electron ID for Run 2*. 2018 [cited 2018; Available from: https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2.
148. Micheli, F. *Cut Based Electron ID for Run 2*

2018 [cited 2018 8/26/2018]; Available from: https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2.

149. Djouadi, A. *The Higgs mechanism and the origin of mass*. in *AIP Conference Proceedings 8*. 2012. AIP.

150. Cacciari, M., G.P. Salam, and G.J.J.o.H.E.P. Soyez, *The anti-kt jet clustering algorithm*. 2008. **2008**(04): p. 063.

151. Cacciari, M., G.P. Salam, and G.J.T.E.P.J.C. Soyez, *FastJet user manual*. 2012. **72**(3): p. 1896.

152. Hinzmann., A. *Jet Identification*

2018 [cited 2018 8/26/2018]; Available from: https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID.

153. Kravchenko, I. *EgammaPFBasedIsolationRun2*. Twiki 2018 09/05/2016 [cited 2018 8/26/2018]; Available from: https://twiki.cern.ch/twiki/bin/view/CMS/EgammaPFBasedIsolationRun2.

154. collaboration, C. and C.C.J.C.P.A.S. CMS-PAS-PFT-10-002, *Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7 TeV*. 2010. **23**.

155. Bachtis, M., *Hadronic Tau Identification and Trigger*, in *Heavy Neutral Particle Decays to Tau Pairs*. 2014, Springer. p. 77-89.

156. Tanabashi, M., et al., *APS: Review of Particle Physics*. 2018. **98**: p. 030001.

157. collaboration, C., *Identification of b quark jets at the CMS Experiment in the LHC Run 2*. 2016, CMS-PAS-BTV-15-001.

158. Instrumentation, C.c.J.J.o., *Identification of b-quark jets with the CMS experiment*. 2013. **8**(04): p. P04013.

159. CristinaAnaMantillaSuarez. *CMS Physics Object School: b-tagging Exercise*. 2018 1/09/2018 [cited 2018 8/27/2018]; Available from: https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideCMSDataAnalysisSchoolLPC2018BTaggingExercise.

160. Botta, C., et al., *Search for direct production of top squark pairs in the fully-hadronic final state with data collected in pp collisions at ps = 13 TeV in 2016*. CMS Note 2016/437, 2016.

161. Lista, L., *Statistical methods for data analysis in particle physics*. Vol. 941. 2017: Springer.

162. Autermann, C., *Boosted Decision Trees*. 2007.

163. Hoecker, A., et al., *TMVA users guide—toolkit for multivariate data analysis, PoSACAT 040*. 2015.

164. Hedrich, N.S., *Boosted regression trees in the Higgs→ ττ channel*. 2013.

165. Karamaoun, A., *BDT-Based Single Tau Selection Algorithm for the Level-2 ATLAS Hadronic Tau Trigger*. 2013: University of Alberta (Canada).

166. Morningstar, A., *A BDT optimization study and assessment of deep learning in selecting VBF events in the $ H\to ZZ^{*}\to4l $ channel*. 2015.

167. Bradski, G. and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*. 2008: " O'Reilly Media, Inc.".

168. Mangano, M.L.J.T.E.P.J.C., *Standard Model backgrounds to supersymmetry searches*. 2009. **59**(2): p. 373.
169. Giammanco, A.J.R.i.P., *Single top quark production at the LHC*. 2016. **1**: p. 1-12.
170. Kenyon, I.R.J.R.o.P.i.P., *The drell-yan process*. 1982. **45**(11): p. 1261.
171. Aaltonen, T., et al., *Search for standard model Higgs boson production in association with a W boson using a matrix element technique at CDF in p p¯ collisions at s= 1.96 TeV*. 2012. **85**(7): p. 072001.
172. Gomez-Ambrosio, R.J.a.p.a., *VBS studies in CMS*. 2018.
173. Regina Demina, Y.-T.D., Mario Galanti, Aran Garcia-Bellido, Otto Hindrichs, and Mauro Verzetti, *Measurement of differential t ¯t production cross sections in lepton + jets final states at 13 TeV* AN2016_020, 2016.
174. Aldaya, T.A., C. Diez Pardos, A. Grohsjean, J. Hauk, J. Keaveney, M. Savitskyi, *Measurement of the differential cross section for top quark pair production in the dilepton final state at* $\sqrt{s}$ *= 13 TeV* AN2017_022, 2017.
175. Twiki_page, *Recommended Jet Energy Corrections and Uncertainties For Data and MC*. https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC, 2018.
176. Twiki_page, *Jet Energy Resolution*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution, 2018.
177. *https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80XReReco*. Available from: https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80XReReco.
178. UCSB, U., CERN,Rice_University, *Search for direct production of top squark pairs in the fully-hadronic final state with data collected in pp collisions at 2016*. CMS Note 2016/437, 2016.
179. Butterworth, J., et al., *PDF4LHC recommendations for LHC run II*. 2016. **43**(2): p. 023001.
180. Sivia, D. and J. Skilling, *Data analysis: a Bayesian tutorial*. 2006: OUP Oxford.
181. Hogg, R.V. and E.A. Tanis, *Probability and statistical inference*. 2009: Pearson Educational International.
182. Wikipedia_contributors, *Likelihood function*. https://en.wikipedia.org/w/index.php?title=Likelihood_function&oldid=868528972, 2018.
183. Tabak, J., *Probability and statistics: The science of uncertainty*. 2014: Infobase Publishing.
184. Clarke, R.N., *A Search for Lepton-Flavor-Violating Decays of the 125 GeV Higgs Boson with Hadronically Decaying Tau Leptons in the 20.3 fb–1, ps= 8 TeV Dataset Collected in 2012 by the ATLAS Detector at the Large Hadron Collider*. 2016: University of California, Berkeley.
185. Ibragimov, I.d.A. and R.Z. Has' Minskii, *Statistical estimation: asymptotic theory*. Vol. 16. 2013: Springer Science & Business Media.
186. Bolthausen, E., E. Perkins, and v.d.V. Aad, *Lectures on Probability Theory and Statistics: Ecole D'Ete de Probabilites de Saint-Flour XXIX-1999*. 2002: Springer Science & Business Media.
187. Verkerke, W. and D.J.D.v. Kirkby, *RooFit users manual v2. 91*. 2008: p. 2.91-33.

188. ATLAS, C., *Collaborations, LHC Higgs Combination Group, Procedure for the LHC higgs boson search combination in Summer 2011*. 2011, CMS-NOTE-2011-005.
189. Schott, G.J.a.p.a., *RooStats for searches*. 2012.
190. Sirunyan, A.M., et al., *Evidence for the Higgs boson decay to a bottom quark–antiquark pair*. 2018. **780**: p. 501-532.
191. Wikipedia_contributors, *Test statistic*. https://en.wikipedia.org/w/index.php?title=Test_statistic&oldid=863096900, 2018.
192. Harvill, J.L., *Essentials of Statistical Inference*. 2008, Taylor & Francis.
193. Mann, P.S., *Introductory statistics*. 2007: John Wiley & Sons.
194. Kottemann, J.E., *Illuminating statistical analysis using scenarios and simulations*. 2017: John Wiley & Sons.
195. Cowan, G., et al., *Asymptotic formulae for likelihood-based tests of new physics*. 2011. **71**(2): p. 1554.
196. Gross, E.J.C.Y.R.S.P., *Practical Statistics for High Energy Physics*. 2017. **4**: p. 165.
197. contributors, W., *Neyman–Pearson lemma*. https://en.wikipedia.org/w/index.php?title=Neyman%E2%80%93Pearson_lemma&oldid=864995010, 2018.
198. Junk, T.J.N.I., S. Methods in Physics Research Section A: Accelerators, Detectors, and A. Equipment, *Confidence level computation for combining searches with small statistics*. 1999. **434**(2-3): p. 435-443.
199. Twiki_page, *Jet energy scale uncertainty sources*. 2018.