Fermi National Laboratory

Final report

Search for di-Higgs resonances decaying to 4 b-jets on CMS at 13 TeV

Summer Internship 2016

Filip Nechanský

Supervisor: Caterina Vernieri

Consultants: Silvio Donato, Jacobo Konigsberg

Fermilab, 2016

Title:

Author: Filip Nechanský

Specialization: Experimental particle physics

Sort of project: Summer Internship project

Supervisor: Caterina Vernieri

Abstract: This measurement uses data collected with the CMS experiment at LHC. It is a search for new particle decaying into pair of Higgs bosons, to probe new physics scenarios. This report summarizes the search for such resonance, where both Higgs boson decay into $b\bar{b}$ pair. Focus is put on my contribution to this search: measurement of the trigger efficiency, optimization of the signal selection criteria.

Key words: New physics, Higgs boson, b-jets, trigger efficiency, diHiggs resonance, CMS

Contents

List of Figures

Introduction

Discovery of the Higgs boson in 2012 marked a huge success for both experiments ATLAS and CMS, but also for Standard model of Physics. After decades of eluding a discovery, this last missing particle was finally found.

However even though standard model is tremendously successful theory, we know it is incomplete. There are many aspects it does not - and cannot - explain. Whether it is e.g. the mass of neutrinos, dark energy or dark matter, there are many experimental observations which point to new physics. This is where Beyond the Standard Model (BSM) theories come in.

Even though there are many different BSM theories, they all predict some new particles - whether they should represent the dark matter, or one of the versions of graviton. It is therefore a task for experimental physicists to push boundaries of the unknown and to uncover new particles - if there are any.

Many of these theories predict a heavy particle decaying into pair of Higgs Bosons. Examples can be e.g. massive KK graviton, which serves as Monte Carlo signal for validation of various methods present in the analysis.

The Higgs boson has many decay channels, where biggest branching ratio (57.7%) is for decay into pair of b quarks. It is therefore logical to use this channel for the search. Hence the channels used in this search is $X \to HH \to b\bar{b}b\bar{b}$ and can be summarized through Feynman diagram in Figure 1.

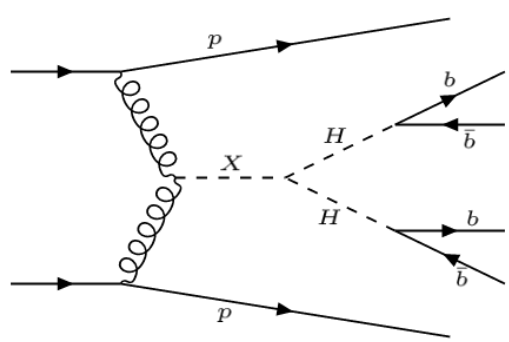


Figure 1: Feynman diagram of particle X production and decay into pair of Higgs bosons, which then decay into $b\bar{b}$ pairs.

This report aims to present one such search done on CMS experiment on the LHC. Specifically it focuses on my contribution to the analysis, though additional parts are also included in order to present more complete picture of the analysis and to demonstrate impact of my studies. First, a short theoretical introduction is given in Chapter 1, explaining basics of Standard Model and hinting at Beyond Standard Model theories.

Chapter 2 shortly depicts the LHC and the CMS detector. Particle identification is also introduced, with focus on identification of b-jets, since they are central part of the analysis. Chapter 3 starts with short introduction of CMS trigger system. This is followed by specification of triggers used in the analysis - two special triggers designed to suppress as much background as possible while maintaining the signal. Rest of the chapter focuses on derivation of trigger efficiency and validation.

The analysis selection is then explained in Chapter 4. This is directly followed by introduction of correction procedures, which aim to improve mass resolution of the Higgs boson and of the resulting X particle, leading to better overall sensitivity to the signal. The last Chapter 5 goes through the rest of the analysis, which was not performed by me, but is important nevertheless.

This search is done for centre-of-mass energies $\sqrt{s} = 13$ TeV and data taken during the year 2016. Similar studies were done for $\sqrt{s} = 8$ TeV[1] and for $\sqrt{s} = 13$ TeV with 2015 dataset[2]. Compared to last year, this search has approximately 10 times bigger statistic and is therefore more sensitive.

Chapter 1

Theoretical introduction

1.1 Standard model

Though not giving a final picture of the universe, the Standard Model of physics is best modern theory describing the fundamental forces and particles in the nature. It describes three of the four fundamental forces - the strong, weak and electromagnetic force. Only the gravitational force is not implemented, though it is assumed to be carried by particle called graviton with spin 2. However this deficiency does not concern us since gravity is by far the weakest of the forces and can be neglected[3].

The standard model represents our best understanding of physics on the lowest scale, but it is known to be incomplete. For example it does not explain the mass of neutrinos or the dark matter. There are many theories trying to explain those problems, as for example Super Symmetry, but none of those have been experimentally confirmed.

1.1.1 Particles of the standard model

The elementary particles are divided between fermions with half-integer spin and bosons with integer spin. The elementary fermions are further split into two groups, quarks and leptons[3]:

Quarks are particles which e.g. build-up the protons and neutrons - fundamental particles of the matter. They are divided in three generations and they have either charge +2/3 (up, strange and top quark) or -1/3 (down, charm and beauty), with the corresponding anti-particle partners. The up and down quarks are the lightest and are stable, unlike the heavier quarks which decay through weak interaction. Quarks are the only elementary particles which interact through all four forces.

Lepton also have three generations, each consisting of charged particle (electron, muon and tauon) and its corresponding neutrino. They interact through weak and electromagnetic interaction. Again, only the lightest electron is stable, while the two heavier lepton decay trough the weak force.

The elementary gauge bosons are representations of the three forces:

Photon carries the electromagnetic force. It has no mass, spin 1 and does not have a charge.

Gluons are responsible for the strong interaction. They carry the colour charge and can therefore self-interact. They are also massless.

Weak bosons are divided between one charged boson W^+ with its corresponding anti-particle W^- and one neutral boson Z. As their name suggest, the carry the weak force. Unlike the photons or gluons they are massive (they are one of the heaviest elementary particles).

Finally, the **Higgs boson** is particle responsible for the mass of elementary particles. Its discovery in 2012 marked a huge success of the standard model and the LHC project.

All the elementary particles are summarized in Figure 1.1.

Aside from the elementary particles, particle physics operates with composite particles. They are exclusively composed of quarks, either of three (with most known examples being proton and neutron)

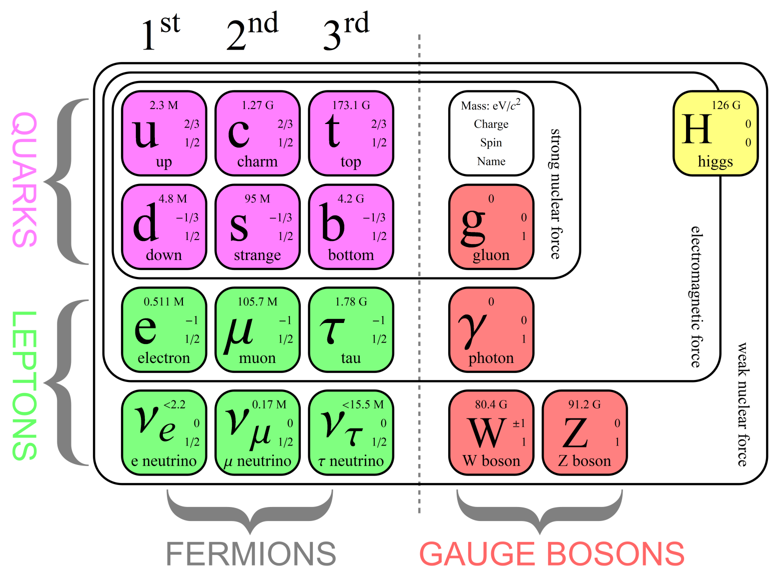


Figure 1.1: Particles of the standard model[4].

or two (e.g. pion). There is strong evidence suggesting composite particle with four or five quarks, but this area is still not explored enough.

Some of the particles more important to this analysis will be established in following sections.

1.1.2 b quark

Bottom (also known as beauty) quark is the second heaviest quark and the fourth heaviest particle of the standard model, with mass around $m_b = 4.66$ GeV and charge -1/3. It is product of all t quark decay and important particle in research of the Higgs boson[5].

The b quark was theorized in 1973 to explain CP violation and was discovered in Fermilab in year 1977 by the E288 experiment. Since quarks cannot exist on their own under standard conditions, they are usually find as part of hadrons. The bottom quark was found as a new meson - the bottomium.

b quark can decay either into c quark or u quark. Since both those decays are propagated through the weak force, the b hadrons have relatively large lifetimes, which is useful feature for their identification.

1.1.3 Higgs Boson

The Higgs Boson was theorized in 1964 as part of mechanism explaining the Electro Weak Symmetry Breaking and origin of mass of most of the elementary particles[1]. Higgs boson eluded discovery for almost five decades, until it was finally found by joint effort of ATLAS[6] and CMS[7] collaborations.

Higgs boson has mass 125.09 $\pm$ 0.21 (stat.) $\pm$ 0.11 (sys.) GeV, no charge or spin and parity +1. Observed decay channels are: Bottom-antibottom pair, W/Z boson pair, tau-antitau pair and photon pair[5].

1.1.4 Quantum chromodynamics

Quantum Chromodynamics is the theory of strong interaction - interaction between quarks and gluons. Similarly to electrodynamics, strong force has a charge. More precisely it has three charges, called

[1]With exception of neutrinos, which may also acquire mass through additional mechanisms.

colours: red and anti-red ($R\bar{R}$), green and anti-green ($G\bar{G}$), and blue and anti-blue ($B\bar{B}$). Colour was first implemented as another degree of freedom, since Δ^{++} would have a symmetrical wave function even though it is a fermion. The number of colours was definitely determined from the ratio of cross-section of $e^+e^- \rightarrow q\bar{q}$ and $e^+e^- \rightarrow \mu^+\mu^-$ processes.

Lagrangian of QCD is based on SU(3) group[2]. Its 9 linearly independent elements can be divided between one singlet $\frac{1}{\sqrt{3}}(R\bar{R} + G\bar{G} + B\bar{B})$ invariant under SU(3) transformation and eight elements affected by such transformation. For this reason QCD has 8 generators - gluons - as carriers of the interactions. Unlike photons, the gluons carry a charge and therefore can emit additional gluons. This small detail has large consequences and it is the main reason why strong interaction is stronger at larger distances[8].

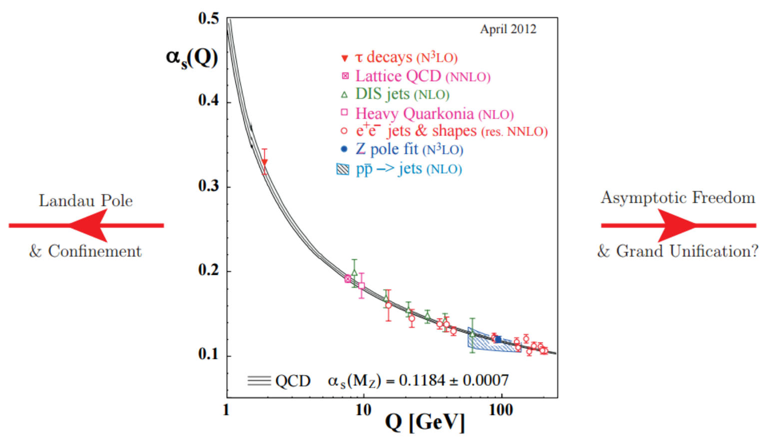


Figure 1.2: Dependence of coupling constant of strong interaction on Q^2[8].

The QCD is also not scale-independent and coupling constant of strong interaction α_s changes as a function of Q^2 similarly to electromagnetic interaction. Nevertheless, here the similarity ends, since the dependence on the momentum transfer is opposite. QCD is strongest at low Q^2 as can be seen in Figure 1.2. The fact that strength of the interaction increases with distance leads to a confinement - quarks are confined in baryons and mesons as colour singlets. For this reason colour cannot be observed directly at lower energies[8].

1.2 Beyond Standard Model

Even though this study is model independent, there are theories which serve as motivation for the analysis, as they predict particle decaying into pair of Higgs bosons. Also they can serve as model signal for an optimization of the procedure. Here is short introduction of two of them:

Warped Extra Dimensions Models (WED)[9] are interesting since they are able to solve the Hiearchy problem[3]. They also predict two particles with diHiggs decay: a spin-two Kaluza-Klein (KK) Graviton or a spin-zero Radion, where invariant mass is expected to be bigger than 500 GeV.

The smaller masses are more interesting for **Next to Minimal SuperSymmetry Standard model (NMSSM)**[10]. Whereas Minimal SuperSymetry (MSSM) presumes one supersymmetric partner to every particle of the standard model, NMSSM adds additional field in order to solve some of the problems of the MSSM. This theory predicts CP-even heavy Higgs Boson decaying into two lighter Higgs boson.

[2]Unitary 3x3 matrices with determinant equal to 1.
[3]Hiearchy problem asks why is there such big difference (10^{32}) between the weak force and gravity.

Chapter 2

Experiment CMS

This chapter describes the experiment CMS, in which this measurement takes place, starting with description of the Large Hadron Collider, which supplies ATLAS with collisions, and ending with a description of individual sub-detectors. All information about the detector and particle identification is taken from [11, 12].

2.1 Large Hadron Collider

Large Hadron Collider (LHC) is a circular proton-proton collider situated at CERN near Geneva, Switzerland. It collides protons with the largest center-of-mass energy in the world, with a maximal planned energy of 14 TeV and current energy of 13 TeV. The high energies enable to study problems on and beyond borders of modern science. For example in the year 2012 both the ATLAS and CMS, the general purpose experiments of the LHC, were able to confirm Higgs boson, a missing particle of the Standard Model .

The tunnel of the LHC is around 27 km long, with additional smaller accelerators providing injecting energy of 900 GeV[1]. There are approximately 2800 bunches at the same time along the ring, with approximately 10^{11} protons present in each bunch. Bunches flow through two separate magnetic channels, which only intersect in four places, where the four experiments are situated. The protons currently collide $40 \cdot 10^6$ times per second. Number of colliding particles and the frequency of their collisions can be summed in luminosity, which is a proportionality factor between cross-section of some process and number of events in which this process occurred[13].

2.2 Detector overview

The **C**ompact **M**uon **S**olenoid (CMS) is a general purpose detector present on the LHC. It is cylindrical detector centred around the beampipe of the accelerator, with diameter of 15 meters, 21.6 meters of length and weight of 14 kilotons. Whole detector is depicted in Figure 2.2.

CMS consists of five main parts: tracker, electromagnetic calorimeter, hadronic calorimeter, magnet and muon spectrometer; which will be briefly summarized in following paragraphs.

The tracker is the closest detector to the beamline, and is mainly responsible for reconstruction of tracks and vertices. It covers central region with $|\eta| < 2.5$ and is made up of two sub-detectors. The first is *Pixel detector*, which is the innermost detector with the biggest granularity in order to be handle large multiplicities present in the area. It has two parts, a barrel with three layer and end-caps, where each cap has two layers. Pixel detector consist of 66 million pixels.

[1]The protons are injected from the CERN accelerator complex, containing for example accelerators Super Proton Synchrotron, Proton Synchrotron or PS Booster

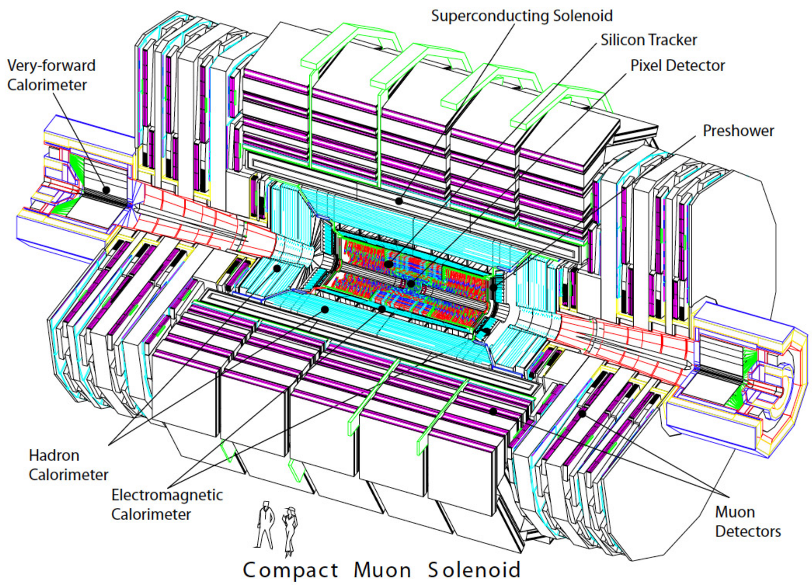


Figure 2.1: Overview of CMS detector and its main components[12].

The outer detector is called Silicon Microstrip detector. Its inner part has four barrel and three end-cap layers, while the outer one has six barrel and nine disc layers. Each layer is either one-sided, providing 2D information about a hit, or double-sided, enabling 3D reconstruction.

The **Electromagnetic calorimeter (ECal)** is responsible for identification of electrons and photons. It uses $PbWO_4$ scintillator as both activator and sampler. This heavy but optically clear material provides an excellent energy resolution and is therefore ideal material for a calorimeter. The ECal consists of a barrel and two end-caps and covers area up to $|\eta| = 3$. In order to distinguish between one high energy photon and two close-by low energy photons, part of the end-cap is preceded by a preshower detector.

Hadrons, as e.g. pions, kaons and protons, are measured by the **Hadronic Calorimeter**. Made up of four parts, most of this detector consists of interchanging layers of heavy material (brass or steel) and scintillating detector. The main part of the detector is responsible for region up to $|\eta| = 3$, while more forward region with $3 < |\eta| < 5$ is covered by the *Hadronic Forward*, a steel/quartz fibre detector. Finally, the *Hadronic Outer* detector helps to contain central showers and is located outside of the magnet described in next part.

Magnetic field is common feature of particle detectors, since it bends charged particle trajectories based on their momentum, and therefore enables to measure their transverse momentum. Hence the **magnet** is essential part of the detector. It is a 13 meters long solenoid, has 6 meters in diameter and magnetic field strength 4 T.

The outermost part of the CMS detector is the **Muon spectrometer**, located within the iron return yoke. Muons are the only particles[2] with high enough penetrability to reach the detector and since they are charged they are bent by the solenoid field.

The muon system has three sub-detectors, one only in the barrel: the *Drift tubes*, which cover area up to $|\eta| < 1.2$; one in both barrel and end-cap: *Cathode strip detector* - with coverage $0.9 < |\eta| < 2.4$; and finally the *Resistive Plate Chambers*, which are located only in the end-cap and reach up to $|\eta| = 1.6$.

[2]With exception of neutrinos, which are practically undetectable by the detector.

2.3 Particle identification and particle flow

Important property of any detector is an ability to distinguish between various particles. Charged particles are in general identified by the inner detector, and if they also deposit energy in the calorimeters, combined information from those detectors enables to estimate particle mass. Muons with $p_T > 3$ GeV are clearly identified by muon spectrometer, while low-energy muons can be be identified by combining information from tracker and calorimeter. The ECal enables to distinguish electrons and photons from other particles, where the latter can be identified by lack of a track in the inner tracker. All those cases are summarized in Figure 2.3.

Objects reconstructed as a combination of information from various sub-detectors are called particle flow objects. This process is done using particle flow algorithm, which is responsible on one side for identification of the particles, as was already mentioned, but also for giving the best estimate of particle properties, again by combing information available from different parts of the detector.

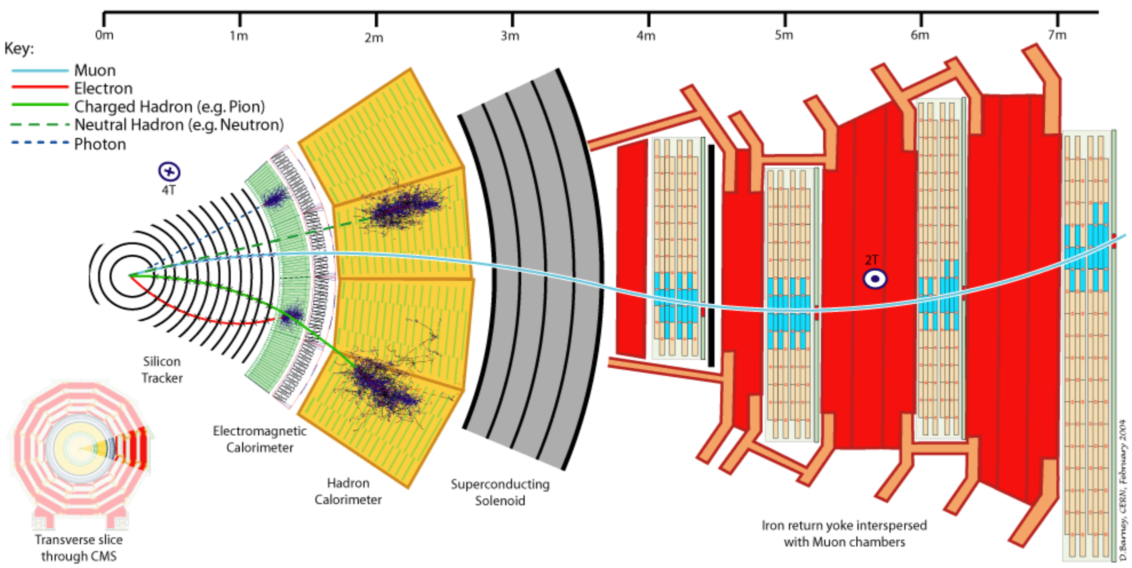


Figure 2.2: Cross-section of a part of the CMS detector. It summarizes transition of various particles trough the detector[14].

2.4 Identification of b-jets

2.4.1 Jets

As was mentioned earlier, b quarks (or coloured particles in general) are subjects to QCD confinement and therefore cannot be observed independently. This leads to fragmentation of the particle, resulting in cone of particles which are in general called jet.

It is nature of jets that they are not clearly defined objects - it is not always obvious which track comes from the fragmentation and which is result of some other interaction. Several algorithms exist designed to reconstruct jets, where the one used on CMS is called anti-k$_T$. It clusters particles beginning with pair which minimizes distance $d_{ij} = min(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2})\frac{\Delta_{ij}^2}{\Delta R^2}$, where $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ and ΔR is parameter of the algorithm in our case set to 0.5.

On CMS, jets can be either reconstructed using calorimeter information, or combined with tracker as Particle flow jets to get more accurate results.

2.4.2 Properties of b-jets

As was mentioned in section 1.1.2, the long lifetime is first property of b-jets which can be used to discriminate them from other jets, since jets are usually produced directly at the primary vertex. Distance of the b hadron decay vertex (secondary vertex) from primary vertex is several millimetres. Due to mass of the b-mesons the invariant mass of the jet will be quite large, as will be its multiplicity.

Another property, which is direct consequence of the previous one, is that most of the tracks will have non-zero impact parameter in regards to the primary vertex. In addition, the impact parameter is signed based on sign of scalar product of the jet direction and vector between primary vertex and point of closest approach of the track. Positive value of impact parameter are then preferred.

2.4.3 Combined Secondary Vertex

The Combined Secondary Vertex (CSV) is a b-tagging tool used on CMS. It uses information about secondary vertex and impact parameter to identify b-jets, though it also enables to discriminate jets without reconstructed secondary vertex by using tracks with large impact parameters and "pseudo" vertices or even when no type of event is present. Some of the variables used by the tool are for example number of tracks at the vertex, type and mass of the vertex, etc. Likelihood ratios are then constructed based on those variables. Information is then combined to develop discriminants, which asign jets value between 0 and 1, where 0 the least b-jet-like and 1 the most.

2.4.4 Combined MultiVariate Algorithm

In order to further improve the signal efficiency, more advanced b-tagging tool is used for offline analysis. It is called Combined Multivariate Algorithm (CMVA) and besides the secondary vertex and track properties exploited by the CSV it also uses information about soft-leptons in the jet. Overall this results in few percent better efficiency and reduction of the multi-jet background. Comparison of both algorithms can be found in Figure 2.4.4.

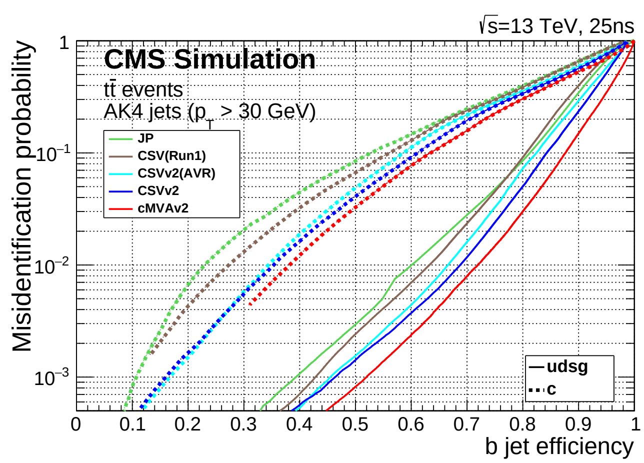


Figure 2.3: Comparison of different b-tagging algorithms used on CMS, from which CSV and CMVA were mentioned in the text and are particularly interesting to us. Displayed is mistag rate for both c jets and $udsg$ jets, both as function of b-tagging efficiency[15].

Both algorithms have defined several working points, which differ by resulting efficiency and probability of mis-identification. In our case medium working points is used, which as average efficiency 72% with $usdg$ mistag rate of about 0.9% and c mistag rate approximately 10%.

Chapter 3

Trigger efficiency

Large frequency of collisions at the LHC directly translates into a large amount of information generated by the CMS. Since it is not possible to extract all the data, only some events can be recorded. It is therefore important to prefer interesting events during this reduction. Hardware or software tool serving this purpose is called a trigger, which decided based on incomplete event information whether the event is interesting or not. Several such triggers are implemented in order to identify various events interesting to the physicist.

3.1 CMS trigger system

On CMS, the trigger system has several levels. First level - Level 1 - reduces frequency from the 40 MHz, which is frequency of the collisions on the LHC, to 50kHz. This stage is purely hardware and works with limited information, for example local energy deposits and track segments. The L1 trigger reaches decision after approximately 3 μs.

The L1 trigger is followed by High Level Trigger, which has access to full event information and can therefore decide better to reject or accept the event. Usually trigger algorithm does not wait for full event reconstruction, but tries to decide as soon as some relevant information is available. This further reduction allows to allocate more time to interesting events. This means that event can be rejected before full track reconstruction.

The HLT is further divided in three subsection. The Level 2 follows directly after L1 and has access to calorimeter and muon system information. At this level jets are partially reconstructed based on calorimeter information. Those are then called CaloJets. Level 2.5 has also access to pixel information and therefore to track and vertex candidates.

Finally, the Level 3 follows after full track reconstruction. At this stage Particle Flow objects are reconstructed. Those are for example jets reconstructed using combined information of tracker and calorimeter, called PF jets[11].

3.2 CSV at the HLT

The CSV algorithm is a multivariate discriminant that has been optimized to use track and primary vertex collections available in offline reconstructed events. The version of this algorithm implemented at Level 3 of the HLT (L3), however, only has access to information available at the trigger as explained below.

Primary Vertex Only information from the pixel detector is used to reconstruct the primary vertex position for the online CSV. Pixel tracks that are compatible with vertices reconstructed through the Fast Primary Vertexing algorithm are used.

Tracks. While full tracker information is available at L3, not all tracks are reconstructed due to limitations of time. The track collection available at L3 consists only of tracks compatible with the "Pixel Primary Vertex" and the 8 jets with highest p_T in the event.

Jets. Only Calojets are available to the HLT CSV (online) algorithm, while PFjets are available to the offline-CSV algorithm.

3.3 Analysis trigger

Triggers used in the analysis are specifically designed to accept as many signal events as possible while reducing the background. This means looking for at least four jets while identifying as many of them as *b*-jets as possible. The former checks on L1 for some general jet signature in the calorimeters, followed by selection on calorimeter and particle flow jets.

b-quark selection is done using online CSV selection at medium working point, which identifies almost 70% of all *b*-jets, while only mis-identifying few percent of non-*b*-jets. This allows to maintain high signal efficiency without need to use tighter selection to keep the event rate low. Both triggers used in the analysis require at least three jets with this *b*-jet selection.

Specifically, following two triggers where used in the analysis:

Quad Jet trigger (QJ):

HLT_BIT_HLT_QuadJet45_TripleBTagCSV_p087_v

- L1 jet activity
- 4 jets $|\eta| < 2.6, p_T > 45$ GeV (Calorimeter and Particle flow level)
- three *b*-tagged jets (CSV medium working point)

Double Jet trigger (DJ):

HLT_BIT_HLT_DoubleJet90_Double30_TripleBTagCSV_p087_v

- L1 jet activity
- 4 jets $|\eta| < 2.6, p_T > 30$ GeV
- 2 jets $|\eta| < 2.6, p_T > 90$ GeV (Calorimeter and Particle flow level)
- three *b*-tagged jets (CSV medium working point)

Main difference is in selection on transverse momentum of both jets. The event are selected such that at least one of both triggers was fired.

3.4 Method to derive trigger efficiency

It is important to derive efficiency of the triggers in order to compare how well is the trigger simulated in the Monte Carlo and account for the different conditions during the data taking by applying Scale factor (SF).

We follow a data driven approach to measure the trigger efficiency. An efficiency of any trigger T can be measured as follows:

$$P(T) = \frac{N(T)}{N_{tot}} \tag{3.1}$$

where $P(T)$ is the probability of the trigger T to pass an event, $N(T)$ is the number of triggered events, and N_{tot} is the number of total events.

We are interested to know the trigger efficiency for the events that pass a particular offline selection S. In that case the trigger efficiency is measured as:

$$P(T|S) = \frac{N(T\&S)}{N(S)} \tag{3.2}$$

where $P(T|S)$ is the probability of the trigger T to pass an event after the selection S, $N(S)$ is the number of events that pass the selection S, and $N(T\&S)$ s the number of events that pass both the selection S and the trigger T.

Then turn-on curve can be interpreted as a set of $P(T|S_i)$, where each S_i corresponds to a bin of the turn-on plot.

Let us assume the trigger is composed by two requirements on some properties of trigger objects, C_1 and C_2. We can write the trigger efficiency as the conditional probability of:

$$\begin{aligned} P(T|S) = & \; P(C_1 \& C_2 \ldots |S) \\ = & \; \frac{N(C_1 \& C_2 \& S)}{N(S)} \\ = & \; \frac{N(C_1 \& C_2 \& S)}{N(C_1 \& S)} \frac{N(C_1 \& S)}{N(S)} \\ = & \; P(C_2|S, C_1) \cdot P(C_1|S). \end{aligned} \tag{3.3}$$

It can be iterated using three or more cuts:

$$\begin{aligned} P(T|S) = & \; P(C_1 \& C_2 \& C_3 \& \ldots |S) \\ = & \; P(C_2 \& C_3 \& \ldots |S, C_1) P(C_1|S) \\ = & \; P(C_3 \& \ldots |S, C_1, C_2) \cdot P(C_2|S, C_1) \cdot P(C_1|S) \\ = & \; \prod_{i=1}^{n} P(C_i|S, C_1, \ldots, C_{i-1}). \end{aligned} \tag{3.4}$$

It means that the efficiency of the trigger $T = C_1 \& \ldots \& C_n$ can be evaluated as the product of the efficiency of single cut given the previous cuts (i.e. $P(C_i|S, C_1, \ldots, C_{i-1})$).

In our case, efficiency is divided into several stages, each studied as a function of some relevant variable , e.g. for Quad Jet trigger:

- L1 as function of sum of p_T of four leading jets ($\sum^4 p_T$)
- Four Calorimeter-jet selection as function of p_T of the fourth jet ($p_{T,4}$)
- Three B-tag jets as function of discriminant of the third jet (CSV_3)
- Four Particle-flow-jets selection as function of p_T of the fourth jet ($p_{T,4}$)

3.5 Results from the data driven estimate

3.5.1 Selection of events to estimate the trigger efficiency

Compared to the previous analysis, where events were chosen by only e.g. single electron trigger, the current one uses much tighter selection, which reflects more the one used in signal extraction. This was done to suppress additional dependencies of the trigger, which lead to significant discrepancies when compared to the signal Monte Carlo. First, a single muon trigger `HLT_BIT_HLT_IsoMu24` with additional cut on quality of the muon is required.

Further selection looks at jet properties. All jets considered are required to have $|\eta| < 2.6$ and should not be pile-up jets. At least four of those jets have to satisfy CMVA>0.185 to pass the preselection. Since we are looking for di-Higgs resonances, a loose Higgs selection is used, where there have to be two pairs of jets compatible with Higgs mass.

3.5.2 QuadJet4 efficiency

The weights `QuadJet45_TripleBTagCSV_p087` can be parametrized as follows:

$$\begin{aligned} w_{\text{Quad}}(p_{T1}, p_{T2}, p_{T3}, p_{T4}, CSV3) = & \; \text{TurnOnL1Pt1PtPt3Pt4}(p_{T1} + p_{T2} + p_{T3} + p_{T4}) \cdot \\ & \text{TurnOnCaloPt4}(p_{T4}) \cdot \text{TurnOnPFPt4}(p_{T4}) \cdot \text{TurnOnCaloCSV3}(CSV3) \end{aligned} \tag{3.5}$$

where the TurnOn for the L1 trigger requirements, Calo-jets and PF-jets selections are shown in Fig.3.1. We allow fit parameters to float to derive their uncertainty. We then will assess a systematic

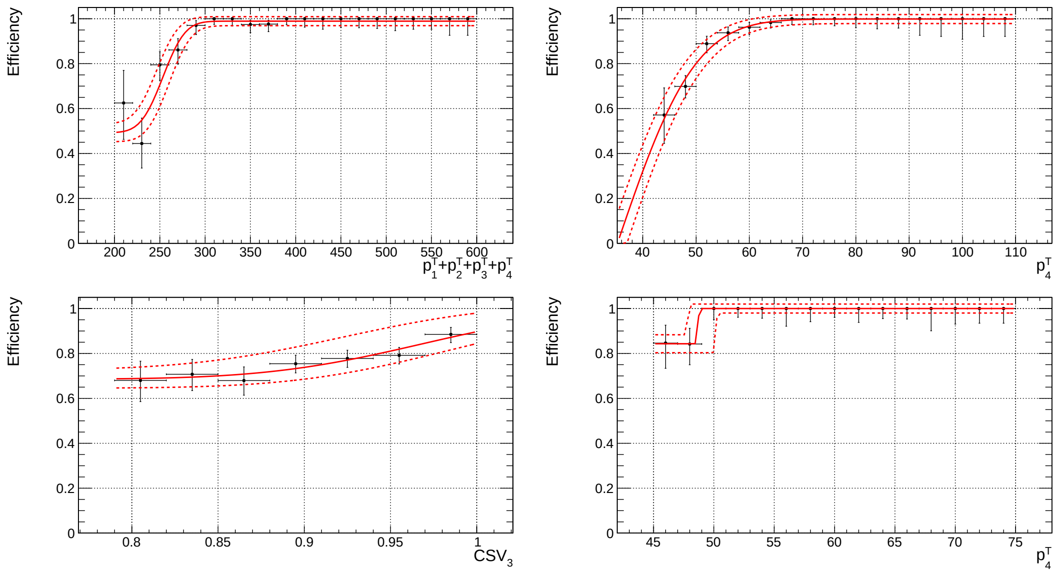


Figure 3.1: TurnOn for the L1 trigger requirements, Calo-jets and PF-jets selections for `QuadJet45_TripleBTagCSV_p087` trigger. Derived using data corresponding to 9.2 fb^{-1}.

uncertainty to this estimate of the trigger efficiency by propagating the fit uncertainty through our event selection in simulated signal events.

The efficiency was validated using the same preselection with extra cut on $p_T > 30$ GeV for the four leading jets. We have compared distributions derived by applying the weights as computed in eq. (3.5) with what we obtain if we apply the trigger bit, reporting a good agreement which validates the method. Results can be found in Figure 3.2.

3.5.3 Double Jet efficiency

The weights for `DoubleJet90_Double30_TripleBTagCSV_p087` can be parametrized as follows:

$$\begin{aligned} w_{\text{Double}}(p_{T1}, p_{T2}, p_{T3}, p_{T4}, CSV3) = \text{TurnOnL1Pt1PtPt3Pt4}(p_{T1} + p_{T2} + p_{T3} + p_{T4})\cdot \\ \cdot\text{TurnOnCaloPt4}(p_{T4}) \cdot \text{TurnOnCaloPt2}(p_{T2}) \cdot \text{TurnOnPFPt4}(p_{T4})\cdot \\ \cdot\text{TurnOnPFPt2}(p_{T2}) \cdot \text{TurnOnCaloCSV3}(CSV3) \end{aligned} \tag{3.6}$$

3.5.4 Combined trigger efficiency

One can describe the combined trigger efficiency as $\epsilon(QJ||DJ) = \epsilon(QJ) + \epsilon(DJ) - \epsilon(QJ\&\&DJ)$. If the correlation between the triggers is negligible, the $\epsilon(QJ\&\&DJ)$ can be computed as $\epsilon(QJ\&\&DJ) = \epsilon(QJ) \cdot \epsilon(DJ)$. We compute explicitly the AND, following the previous approach for the separate trigger paths. First the efficiency for `DoubleJet90_Double30 _TripleBTagCSV_p087` is computed, then followed by the efficiency of `QuadJet45_TripleBTagCSV_p087`. The weights can be then parametrized as follow:

$$w_{\text{AND}}(p_{T1}, p_{T2}, p_{T3}, p_{T4}, CSV3) = w_{\text{Double}}(p_{T1}, p_{T2}, p_{T3}, p_{T4}, CSV3) \cdot w_{\text{Quad}'}(p_{T1}, p_{T2}, p_{T3}, p_{T4}, CSV3) \tag{3.7}$$

where the $w_{\text{Quad}'}$ is computed after w_{Double}.

Finally, the combined efficiency is validated using the formula $\epsilon(QJ||DJ) = \epsilon(QJ) + \epsilon(DJ) - \epsilon(QJ\&\&DJ)$ (see Figure 3.3).

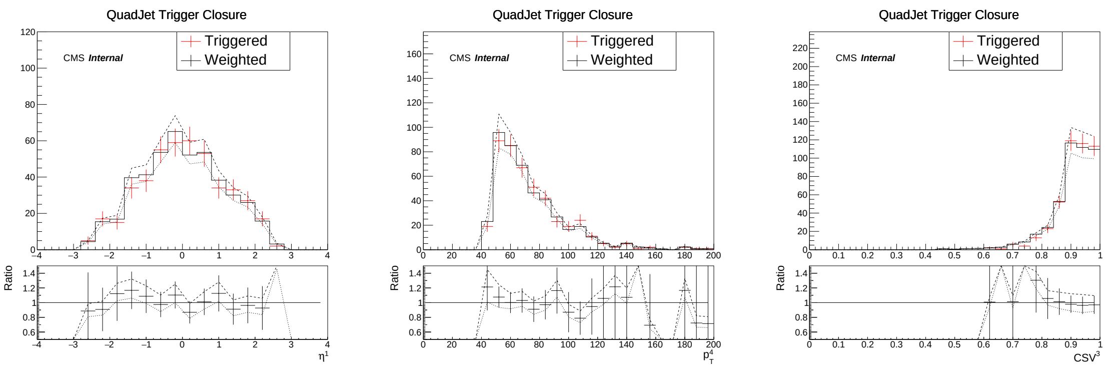


Figure 3.2: η_1, p_{T4} and CSV3 distributions for data events if applying trigger bit (blue) or weights (black) as computed in eq. (3.5) for `QuadJet45_TripleBTagCSV_p087` trigger. Efficiency derived using data (9.23 fb^{-1}).

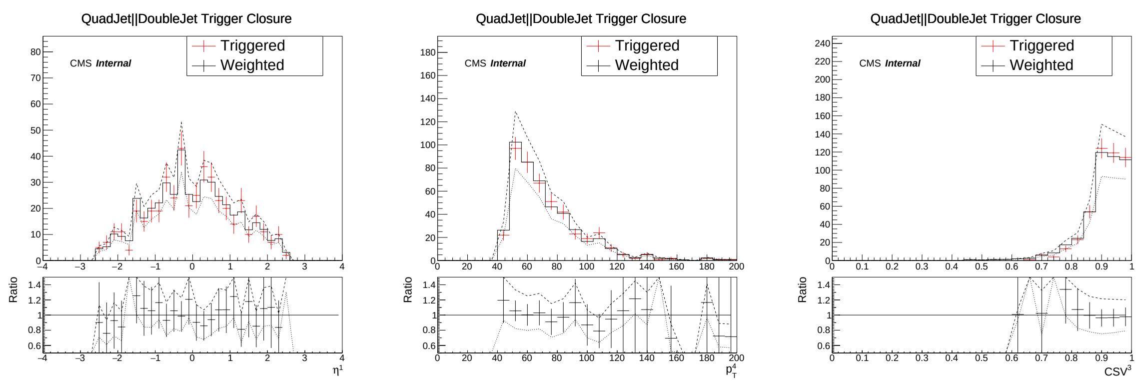


Figure 3.3: η_1, p_{T4} and CSV3 distributions for data events if applying trigger bit (blue) or weights (black) for OR of both triggers. Efficiency derived using data corresponding to 9.2 fb^{-1}.

3.5.5 Trigger efficiency correction factor for signal

We have compared in simulated signal samples the relative trigger efficiency with respect to a pre-selection of 4 b-tagged jets with p $_T$ > 30 GeV derived from simulation and the measured one for the two paths combination (OR). As was mentioned earlier, due to tracking inefficiency, data and MC give different results. The data driven estimate will be used to correct the available simulation. The uncertainty is derived from the $\pm 1\sigma$ fit variations to each turn-on and it will be propagated as systematic uncertainty on the trigger efficiency for each signal mass hypothesis.

3.6 Validation using MC $t\bar{t}$

The tracking in-efficiency experienced during the 2016 data taking (Run B to Run F) is not included in the simulation. Hence we cannot use MC to validate the method against data. To do so, the whole procedure is repeated to measure the efficiency in the same phase space selected in data using a $t\bar{t}$ simulated sample. The selection and the method is same in all regards as for data.

Then the weights computed using the simulated $t\bar{t}$ sample are applied to signal MC and compared to the trigger bit simulation. The closure test as function of the signal mass is reported in Fig. 3.6. The difference between weighted and triggered will be taken as an additional systematic uncertainty.

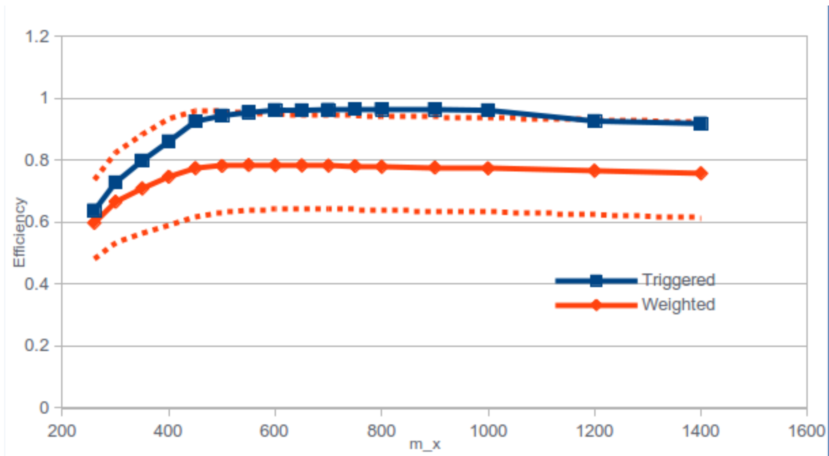


Figure 3.4: Fraction of triggered and weighted events after 4 b-tagged jets preselection for signal sample as function of the signal mass. Efficiency derived using data corresponding to 9.2 fb^{-1} and uncertainty is found to be around 20%.

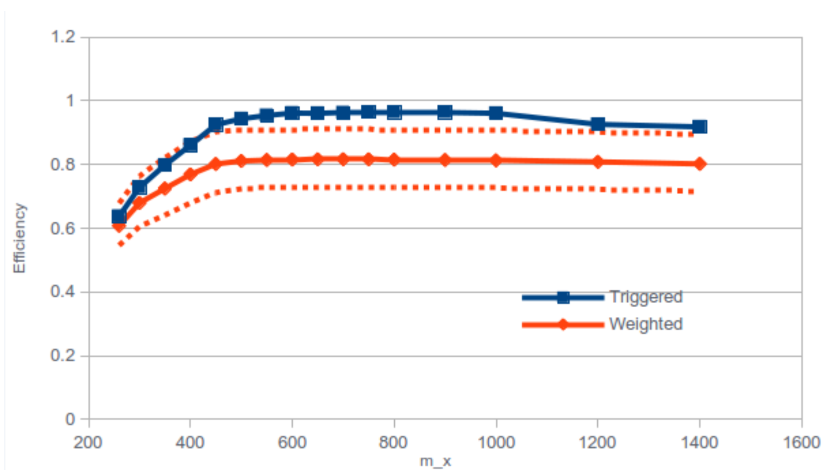


Figure 3.5: Fraction of triggered and weighted events after 4 b-tagged jets preselection for signal sample as function of the signal mass. Efficiency derived using $t\bar{t}$ simulated sample and uncertainty is found to be around 15%.

Chapter 4

Event selection

This search covers a wide range of mass hypotheses for the resonance, m_X, between 260 GeV and 1200 GeV. The kinematics involved in the decay of such a resonance change substantially over this range, and thus two sets of event selection criteria have been designed to efficiently cover the different kinematic regimes. The Low Mass Regime (LMR) event selection criteria are applied on mass hypotheses between 260 GeV and ≈450 GeV, where the decaying Higgs bosons have a lower boost. For mass above 450 GeV, the Medium Mass Regime (MMR) event selection criteria are applied, where the decaying Higgs are sufficiently boosted for their *b*-jets to subtend a small angle between themselves. This makes it relatively easy to determine which *b*-jets originate from the same Higgs. After the selection a signal region is defined, which is region with the biggest signal significance. Only event in this region are then used to look for signal.

4.1 Low Mass Regime

Events are required to contain at least 4 central jets ($|\eta| < 2.5$) with $p_T > 30$ GeV and passing CMVA selection. The distribution of the number of such central jets is displayed in Fig. 4.1 for both simulated signal events ($m_X = 260 - 600$ GeV) and data.

Among all the selected jets we search for HH candidates such that:

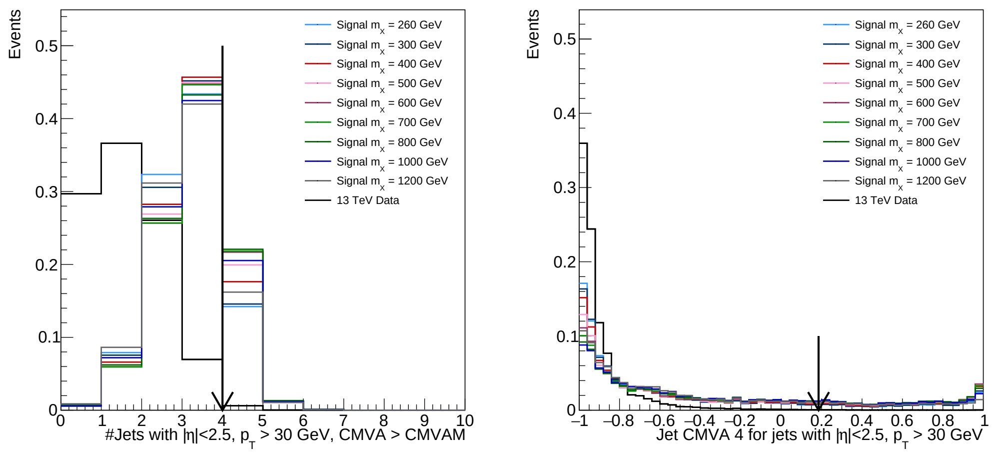
- Starting from the CMVA-leading jet we look for another jet such that m_{H1} lies in the window $m_H(115) \pm 34$ GeV.
- We search for another pair with the same criterion among the jets that are left, to identify H_2.
- In case multiple candidates satisfying the previous criteria are found, the combination which minimizes χ^2[1] is chosen, in order to select the two dijet pairs which are closest to the nominal Higgs mass.

Signal region, in which the search is conducted, is defined as region where $\chi^2 < 1$. For background estimation, additional Sideband region is defined, which contains events with $1 < \chi < 2$ and $(m_{H1} - \bar{m_H})(m_{H2} - \bar{m_H}) < 0$. The distributions of the CMVA for the fourth CMVA-ordered jet are shown in 4.1.

4.2 Medium Mass Regime

The selection criteria for the medium mass regime relies on the boosted topology of the *b*-jet pairs from each Higgs. Events are required to contain at least 4 jets with $p_T > 30$ GeV, $|\eta| < 2.5$ and

[1] $\chi^2 = (m_{H1} - \bar{m_H})^2/\bar{\sigma_H}^2 + (m_{H2} - \bar{m_H})^2/\bar{\sigma_H}^2$ (with $\bar{m_H}, \bar{\sigma_H}$ optimized for biggest signal significance)



(a) (b)

Figure 4.1: (a) Number of jets with $p_T > 30$ GeV, $|\eta| < 2.5$, and CMVAM for signal events and data. (b) CMVA discriminant distributions for the third and forth CMVA-ordered jets.

Table 4.1: Signal Region definition (radius and centre (GeV)) for LMR and MMR

	LMR		MMR	
	center	radius	center	radius
baseline	115	25	120	20
after regression	120	20	125	20

passing CMVA selection. The distribution of the number of such jets in signal and data are displayed in Fig. 4.1. Within the set of such jets in each event, the two di-jet pairs with the minimum χ^2 and such that the jets within a pair are separated by $\Delta R < 1.5$ are selected as the di-Higgs candidates. For $m_X > 450$ GeV, the boost of each Higgs is $\sim 1.8 - 3.5$. Therefore, the opening angle between its decay products, which is roughly $1/\gamma$, is expected to be lower than 1.5. This criterion rejects combinatorial backgrounds from incorrect jet pairing. Signal region and sideband follow the same definition as Low Mass Region, though the mean value and uncertainty may be different.

4.2.1 b-jet momentum correction

The presence of a neutrino in about 35% of b hadron decays makes the b-jet energy resolution worse than light quark and gluon jets. Hence, the searches for the Higgs boson decaying into b quarks have developed a dedicated technique to improve the $b\bar{b}$ pair invariant mass resolution. This is done with a multivariate regression targeting the generator-level p_T of the b-quark.The regression for this analysis was trained using the signal samples. Effect of this correction can be seen in Figure 4.2 for Higgs boson invariant mass, resulting in 10.06% improvement of σ/μ value for $m_X = 300$ GeV signal sample and 11.94% for $m_X = 300$ 800 GeV sample.

After applying b-jet energy corrections the reconstructed Higgs boson invariant mass resolution improves. To take advantage of the improved resolution the Signal Region criteria have been optimized for both LMR and MMR, resulting in a different radius and center. The SR criteria are summarized in Tab. 4.1. In each case the chosen radius and centre have been optimized for the sensitivity.

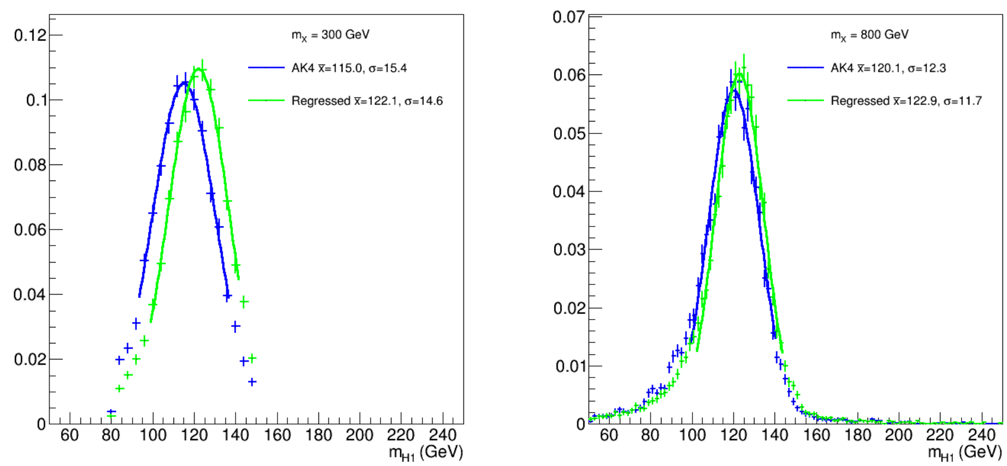


Figure 4.2: Distribution of Higgs mass before and after regression is applied for $m_X = 300$ GeV and $m_X = 800$ signal samples.

4.2.2 Kinematic Constraint on m_H

In this search, the strong kinematic constraints on the invariant mass of the b-jet pairs to the nominal mass of the Higgs boson are exploited to correct the momenta of the reconstructed b-jets. Before reconstructing the 4-momentum vector of the di-Higgs resonance, p_X^μ, the 3-momentum vectors of the jets constituting the Higgs candidates are corrected by this kinematic constraint. The kinematic constraint requires the η, ϕ and p_T resolution for b -jets in order to construct the χ^2 for an event with 4 b-jets that has to be minimized while maintaining $m_H = 125$ GeV.

In Fig. 4.3 the invariant mass distributions for signal events is illustrated for the Low Mass and High Mass regimes before and after the corrections returned by the kinematic constraint are applied. The kinematic constraint is seen to appreciably increase the mean of the signal lineshapes and also improve the mass resolution by factors of 20% to 40%, depending on the mass point under consideration.

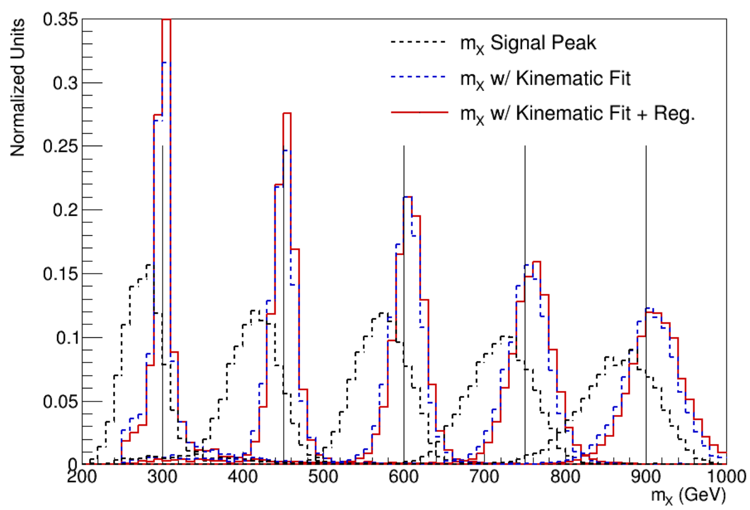


Figure 4.3: Signal line-shapes for the $m_X = 400 - 1200$ GeV mass points (red) overlaid with variations in the line-shape due to the kinematic constraint (black).

Chapter 5

Results

This chapter aims to briefly summarize derivation of expected limits for BSM searches. Even though I have not worked on them, they are important in order to present more complete picture of the analysis

5.1 Expected upper limits

The Asymptotic CL_S method of the Higgs Combination Tool is used to compute the expected upper limits on the signal cross sections at 95% confidence level. These limits for the Low and Medium Mass Regimes, which follow slightly different analysis paths, are shown in Fig. 5.1. Based on the expected sensitivity the transition from LMR to MMR is for mass hypothesis larger than 400 GeV.

We compare with theory prediction for RS kk-Graviton production (both Drell Yan and Gluon Fusion compontents) in the HH final state assuming k=0.1. The results can be found in Fig. 5.1. I

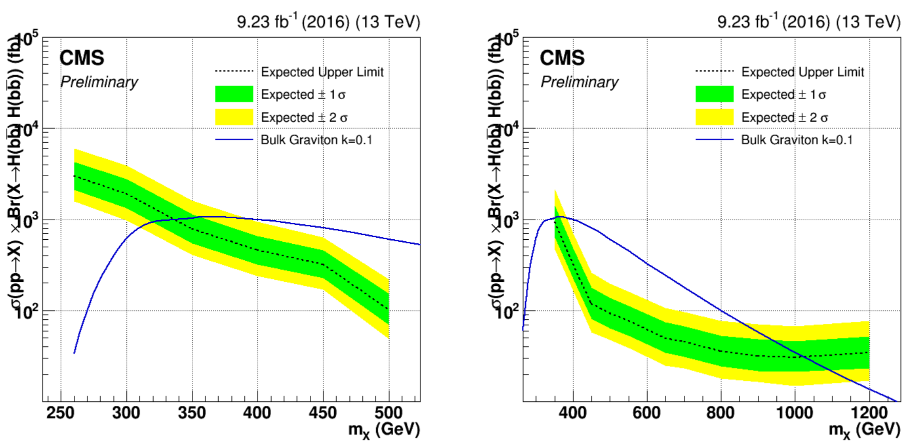


Figure 5.1: The expected upper limit of $\sigma(pp \to X \to HH \to b\bar{b}b\bar{b})$ at 95% confidence level in the Low Mass Regime (left) and Medium mass region (right) using 9.23 fb^{-1} of data. Left is the baseline, Right includes als the b-jet momentum corrections using the regression technique.

The use of the regression technique improves the expected sensitivity by 5-25% as shown in Tab. 5.1.s

Table 5.1: The expected upper limit of $\sigma(pp \to X \to HH \to b\bar{b}b\bar{b})$ at 95% confidence level for LMR and MMR with and without including jet energy correction from regression technique. Values are reported in fb.

	Kin. Fit	Kin. Fit +regr.	Improvement (%)
LMR			
260	3085	2976	3.7
300	1976	1898	4.1
350	808	781	3.5
400	470	455	3.3
MMR			
350	-	925.8	-
400	315.9	321.8	-1.8
450	140.6	118.2	19.0
500	108.9	92.8	17.3
550	89.4	76.7	16.6
650	59.1	49.3	19.9
700	52.2	45.4	15.0
800	40.5	35.6	13.8
900	35.6	31.7	12.3
1000	33.7	30.8	9.4
1200	38.6	34.7	11.2

Summary

This report introduced search for diHiggs resonance, which is a state predicted by some of the Beyond Standard Model theories, as is for example SuperSymmetry. Since Higgs boson has largest branching ratio for decay to pair of b-quarks, final state containing four b-jets is studied. Focus is put on my contribution to the analysis, though all parts of the analysis are mentioned in order to present more complete picture.

Chapter 1 offers summary of the Standard Model followed by theoretical motivation for this search. In Chapter 2, experimental apparatus - the CMS detector - is presented, followed by introduction to particle reconstruction. Focus is put on b-jet identification and summary of two algorithms used for this purpose.

Chapter 3 reviewed triggers used in the analysis. which were specifically designed to select the signal while reducing the background. Efficiency of those triggers then had to be derived in order to be able to correct discrepancies between data and Monte Carlo. Hence, the method used to determine the efficiency is explained, together with its application on data and $t\bar{t}$ sample for a validation. Finally, comparison to signal trigger efficiency is made. This is where I contributed the most, fine-tuning the selection, cross-checking compatibility, expansion of the method for $t\bar{t}$ and producing the results.

Selection used in the analysis was introduced in Chapter 4, complemented by summary of corrections applied. Their purpose is to improve signal efficiency and resolution of the signal. Here I was responsible for optimization of the signal region and study of effects of corrections used on jets. Finally, Chapter 5 goes through expected upper limits. Both last chapters also demonstrate impact of the corrections.

It is already clear, that this analysis is more sensitive to the signal than the previous ones. Future plans include assessment of systematic uncertainties of the trigger efficiency and inclusion of all 2016 datasets. Results of the analysis, once finished, will be published by the CMS experiment.

Bibliography

[1] CMS Collaboration, "Search for resonant pair production of Higgs bosons decaying to two bottom quark–antiquark pairs in proton–proton collisions at 8 TeV," Phys. Lett. B **749** (2015) 560 [arXiv:1503.04114 [hep-ex]]

[2] CMS Collaboration, "Search for resonant pair production of Higgs bosons decaying to two bottom quark-antiquark pairs in proton-proton collisions at 13 TeV," CMS-PAS-HIG-16-002.

[3] David Griffiths, " Introduction to Elementary Particles," Wiley, 2008, 470 ISBN: 978-3-527-40601-2

[4] Group Serra, Standard Model , http://www.physik.uzh.ch/groups/serra/StandardModel.html [Cited on 7. 4. 2016]

[5] Particle Data Group (PDG) http://pdglive.lbl.gov/ [Cited on 14. 9. 2016]

[6] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," Phys. Lett. B **716** (2012) 1 doi:10.1016/j.physletb.2012.08.020 [arXiv:1207.7214 [hep-ex]].

[7] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," Phys. Lett. B **716** (2012) 30 doi:10.1016/j.physletb.2012.08.021 [arXiv:1207.7235 [hep-ex]].

[8] P. Skands, "Introduction to QCD," [arXiv:1207.2389 [hep-ph]].

[9] L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension," Phys. Rev. Lett. **83** (1999) 3370 doi:10.1103/PhysRevLett.83.3370 [hep-ph/9905221].

[10] R. Barbieri, D. Buttazzo, K. Kannike, F. Sala and A. Tesi, "Exploring the Higgs sector of a most natural NMSSM," Phys. Rev. D **87** (2013) no.11, 115018 doi:10.1103/PhysRevD.87.115018 [arXiv:1304.3670 [hep-ph]].

[11] C. Vernieri, "Search for resonances decaying to $H(b\bar{b})$ pairs with the CMS experiment at the LHC," CERN-THESIS-2014-161, CMS-TS-2014-035.

[12] CMS Collaboration, "The CMS experiment at the CERN LHC," JINST **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004

[13] Stephen Myers, “The large hadron collider 2008–2013,” Int. J. Mod. Phys. A, **28** (2013) 1330035 [DOI: 10.1142/S0217751X13300354].

[14] Wyatt A. Behn, European Laboratory for Particle Physics, “The CMS Detector and the Token Bit Manager,” https://www.phys.ksu.edu/reu2014/wabehn/

[15] CMS Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC Run 2,” CMS-PAS-BTV-15-001.