Search for excited quark states decaying to qW/qZ

David Leppla-Weber

Abstract

A search for an excited quark state, called q*, is presented using data recorded by the CMS experiment during the years 2016, 2017 and 2018 with a centre-of-mass energy of $\sqrt{s} = 13$ TeV and a total integrated luminosity of $137.19\,\mathrm{fb}^{-1}$. By analysing its decay channels to q + W and q + Z that further decay to $q + q\bar{q}$, resulting in two jets in the final state, the q* can be excluded up to a mass of 6.1 (qW) TeV resp. 5.5 TeV (qZ) with a confidence level of 95 %. This limit is about 1 TeV higher than the limits found by a previous research of data collected by CMS in 2016 [6], excluding the q* particle up to a mass of 5.0 TeV resp. 4.7 TeV. Also a comparison of the new DeepAK8 [11] and the older N-subjettiness [15] tagger is conducted, showing that the newer DeepAK8 tagger, based on a deep neural network, is currently approximately at the same level as the N-subjettiness tagger, but has the potential to further improve in performance, between others because of an improved training that was just published.

Zusammenfassung

In dieser Arbeit wird eine Suche nach angeregten Quarkzuständen, genannt q*, durchgeführt. Dafür werden Daten mit einer gesamten integrierten Luminosität von $137.19\,\mathrm{fb}^{-1}$ analysiert, welche über die Jahre 2016, 2017 und 2018 bei einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV vom CMS Experiment aufgenommen wurden. Indem der Zerfall des q* Teilchens zu q + W und q + Z untersucht wird, kann dieses mit einem Konfidenzniveau von 95 % bis zu einer Masse von 6.1 TeV (qW) bzw. 5.5 TeV (qZ) ausgeschlossen werden. Dieses Limit liegt etwa 1 TeV höher, als das von vorhergegangener Forschung [6] gesetzte von 5.0 TeV bzw. 4.7 TeV. Dabei wird der neue DeepAK8 Tagger [11], welcher auf einem neuronalen Netzwerk basiert, mit dem älteren N-Subjetiness Tagger [15] verglichen. Das Endergebnis der beiden Tagger unterscheidet sich kaum, jedoch gibt es beim DeepAK8 Tagger noch potential zur Verbesserung, unter anderem durch ein verbessertes Traininig, welches vor kurzem veröffentlicht wurde.

Contents

1 Introduction

The Standard Model is a very successful theory in describing most of the effects on a particle level. But it still has a lot of shortcomings that show that it isn't yet a full 'theory of everything'. To solve these shortcomings, lots of theories beyond the standard model exist that try to explain some of them.

One category of such theories is based on a composite quark model. Quarks are currently considered elementary particles by the Standard Model. The composite quark models on the other hand predict that quarks consist of particles unknown to us so far or can bind to other particles using unknown forces. This could explain the symmetries between particles and reduce the number of constants needed to explain the properties of the known particles. One common prediction of those theories are excited quark states. Those are quark states of higher energy that can decay to an unexcited quark under the emission of a boson. This thesis will search for their decay to a quark and a W/Z boson. The W/Z boson then decays in the hadronic channel, to two more quarks. The endstate of this decay has only quarks, making Quantum Chromodynamics effects the main background.

In a previous research [6], a lower limit for the mass of an excited quark has already been set using data from the 2016 run of the Large Hadron Collider with an integrated luminosity of $35.92\,\text{fb}^{-1}$. Since then, a lot more data has been collected, totalling to $137.19\,\text{fb}^{-1}$ of data usable for research. This thesis uses this new data as well as a new technique to identify decays of highly boosted particles based on a deep neural network. By using more data and new tagging techniques, it aims to either confirm the existence of the q* particle or improve the previously set lower limit of 5 TeV respectively 4.7 TeV for the decay to qW respectively qZ on its mass to even higher values. It will also directly compare the performance of this new tagging technique to an older tagger based on jet substructure studies used in the previous research.

In chapter 2, a theoretical background will be presented explaining in short the Standard Model, its shortcomings and the theory of excited quarks. Then, in chapter 3, the Large Hadron Collider and the Compact Muon Solenoid, the detector that collected the data for this analysis, will be described. After that, in chapters 4-7, the main analysis part follows, describing how the data was used to extract limits on the mass of the excited quark particle. At the very end, in chapter 8, the results are presented and compared to previous research.

2 Theoretical motivation

This chapter presents a short summary of the theoretical background relevant to this thesis. It first gives an introduction to the standard model itself and some of the issues it raises. It then goes on to explain the background processes of quantum chromodynamics and the theory of q*, which will be the main topic of this thesis.

2.1 Standard model

The Standard Model of physics proved to be very successful in describing three of the four fundamental interactions currently known: the electromagnetic, weak and strong interaction. The fourth, gravity, could not yet be successfully included in this theory.

The Standard Model divides all particles into spin-$\frac{n}{2}$ fermions and spin-n bosons, where n could be any integer but so far is only known to be one for fermions and either one (gauge bosons) or zero (scalar bosons) for bosons. Fermions are further classified into quarks and leptons. Quarks and leptons can also be categorized into three generations, each of which contains two particles, also called flavours. For leptons, the three generations each consist of a lepton and its corresponding neutrino, namely first the electron, then the muon and third, the tau. The three quark generations consist of first, the up and down, second, the charm and strange, and third, the top and bottom quark. So overall, their exists a total of six quark and six lepton flavours. A full list of particles known to the standard model can be found in fig. 1. Furthermore, all fermions have an associated anti particle with reversed charge.

The matter around us, is built from so called hadrons, that are bound states of quarks, for example protons and neutrons. Long lived hadrons consist of up and down quarks, as the heavier ones over time decay to those.

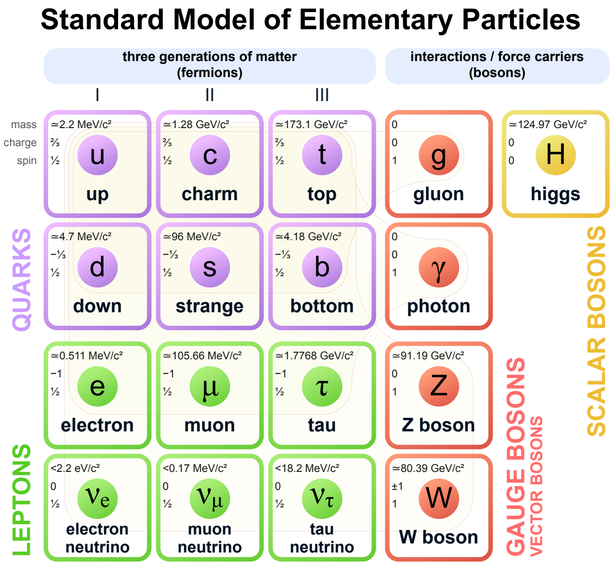

Figure 1: Elementary particles of the Standard Model and their mass charge and spin.

The gauge bosons, namely the photon, $W^{\pm}$ bosons, Z^0 boson, and gluon, are mediators of the different forces of the standard model.

The photon is responsible for the electromagnetic force and therefore interacts with all electrically charged particles. It itself carries no electromagnetic charge and has no mass. Possible interactions are either scattering or absorption. Photons of different energies can also be described as electromagnetic waves of different wavelengths.

The $W^\pm$ and Z^0 bosons mediate the weak force. All quarks and leptons carry a flavour, which is a conserved value. Only the weak interaction breaks this conservation, a quark or lepton can therefore, by interacting with a $W^\pm$ boson, change its flavour. The probabilities of this happening are determined by the Cabibbo-Kobayashi-Maskawa matrix:

$$V_{CKM} = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.224 & 0.974 & 0.042 \\ 0.008 & 0.041 & 0.999 \end{pmatrix} \tag{1}$$

The probability of a quark changing its flavour from i to j is given by the square of the absolute value of the matrix element V_{ij}. It is easy to see, that the change of flavour in the same generation is way more likely than any other flavour change.

Due to their high masses of 80.39 GeV resp. 91.19 GeV, the $W^\pm$ and Z^0 bosons themselves decay very quickly. Either in the leptonic or hadronic decay channel. In the leptonic channel, the $W^\pm$ decays to a lepton and the corresponding anti-lepton neutrino, in the hadronic channel it decays to a quark and an anti-quark of a different flavour. Due to the Z^0 boson having no charge, it always decays to a fermion and its anti-particle, in the leptonic channel this might be for example a electron - positron pair, in the hadronic channel an up and anti-up quark pair. This thesis examines the hadronic decay channel, where both vector bosons essentially decay to to quarks.

The quantum chromodynamics (QCD) describes the strong interaction of particles. It applies to all particles carrying colour (e.g. quarks). The force is mediated by gluons. These bosons carry colour as well, although they don't carry only one colour but rather a combination of a colour and an anticolour, and can therefore interact with themselves and exist in eight different variants. As a result of this, processes, where a gluon decays into two gluons are possible. Furthermore the strength of the strong force, binding to colour carrying particles, increases with their distance making it at a certain point more energetically efficient to form a new quark - antiquark pair than separating the two particles even further. This effect is known as colour confinement. Due to this effect, colour carrying particles can't be observed directly, but rather form so called jets that cause hadronic showers in the detector. Those jets are cone like structures made of hadrons and other particles. The effect is called Hadronisation.

2.1.1 Shortcomings of the Standard Model

While being very successful in describing the effects observed in particle colliders or the particles reaching earth from cosmological sources, the Standard Model still has several shortcomings.

- **Gravity**: as already noted, the standard model doesn't include gravity as a force.
- **Dark Matter**: observations of the rotational velocity of galaxies can't be explained by the known matter. Dark matter currently is our best theory to explain those.
- **Matter-antimatter asymmetry**: The amount of matter vastly outweights the amount of antimatter in the observable universe. This can't be explained by the standard model, which predicts a similar amount of matter and antimatter.
- **Symmetries between particles**: Why do exactly three generations of fermions exist? Why is the charge of a quark exactly one third of the charge of a lepton? How are the masses of the particles related? Those and more questions cannot be answered by the standard model.
- **Hierarchy problem**: The weak force is approximately 10^{24} times stronger than gravity and so far, there's no satisfactory explanation as to why that is.

2.2 Excited quark states

One category of theories that try to explain the symmetries between particles of the standard model are the composite quark models. Those state, that quarks consist of some particles unknown to us so far. This could explain the symmetries between the different fermions. A common prediction of those models are excited quark states (q*, q**, q***...). Similar to atoms, that can be excited by the absorption of a photon and can then decay again under emission of a photon with an energy corresponding to the excited state, those excited quark states could decay under the emission of any boson. Quarks are smaller than 10^{-18} m. This corresponds to an energy scale of approximately 1 TeV. Therefore the excited quark states are expected to be in that region. That will cause the emitted boson to be highly boosted.

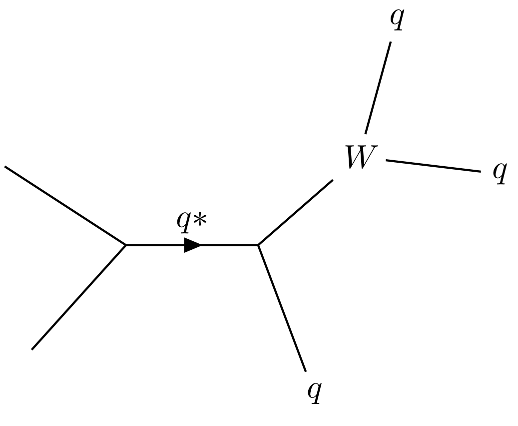

Figure 2: Feynman diagram showing a possible decay of a q* particle to a W boson and a quark with the W boson also decaying to two quarks.

This thesis will search data collected by the CMS in the years 2016, 2017 and 2018 for the single excited quark state q* which can decay to a quark and any boson. An example of a q* decaying to a quark and a W boson can be seen in fig. 2. As explained in sec. 2.1, the vector boson can then decay either in the hadronic or leptonic decay channel. This research investigates only the hadronic channel with two quarks in the endstate. Because the boson is highly boosted, those will be very close together and therefore appear to the detector as only one jet. This means that the decay of a q* particle will have two jets in the endstate (assuming the W/Z boson decays to two quarks) and will therefore be hard to distinguish from the QCD background described in sec. 2.2.1.

The choice of only examining the decay of the q* particle to the vector bosons is motivated by the branching ratios calculated for the decay [2]:

Table 1: Branching ratios of the decaying q* particle.

decay mode	br. ratio [%]	decay mode	br. ratio [%]
$U^* \to ug$	83.4	$D^* \to dg$	83.4
$U^* \to dW$	10.9	$D^* \to uW$	10.9
$U^* \to u\gamma$	2.2	$D^* \to d\gamma$	0.5
$U^* \to uZ$	3.5	$D^* \to dZ$	5.1

The decay to the vector bosons have the second highest branching ratio. The decay to a gluon and a quark is the dominant decay, but virtually impossible to distinguish from the QCD background described in the next section. This makes the decay to the vector bosons the obvious choice.

To reconstruct the mass of the q* particle from an event successfully recognized to be the decay of such a particle, the dijet invariant mass has to be calculated. This can be achieved by adding their four momenta, vectors consisting of the energy and momentum of a particle, together. From the four momentum it's easy to derive the mass by solving $E = \sqrt{p^2 + m^2}$ for m.

This theory has already been investigated in [6] analysing data recorded by CMS in 2016, excluding the q* particle up to a mass of 5 TeV resp. 4.7 TeV for the decay to qW resp. qZ analysing the hadronic decay of the vector boson. This thesis aims to either exclude the particle to higher masses or find a resonance showing its existence using the higher center of mass energy of the LHC as well as more data that is available now.

2.2.1 Quantum Chromodynamic background

In this thesis, a decay with two jets in the endstate will be analysed. Therefore it will be hard to distinguish the signal processes from QCD effects. Those can also produce two jets in the endstate, as can be seen in fig. 3. They are also happening very often in a proton proton collision, as it is happening in the Large Hadron Collider. This is caused by the structure of the proton.

It not only consists of three quarks, called valence quarks, but also of a lot of quark-antiquark pairs connected by gluons, called the sea quarks, that exist due to the self interaction of the gluons binding the three valence quarks. Therefore the QCD multijet backgroubd is the dominant background of the signal described in sec. 2.2.

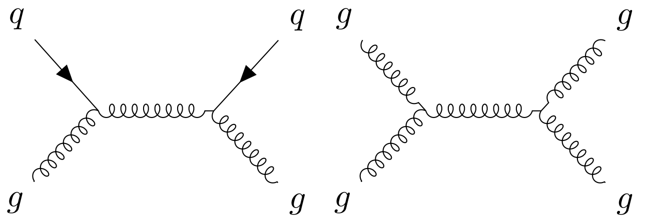

Figure 3: Two examples of QCD processes resulting in two jets.

3 Experimental Setup

Following on, the experimental setup used to gather the data analysed in this thesis will be described.

3.1 Large Hadron Collider

The Large Hadron Collider is the world's largest and most powerful particle accelerator [**website**]. It has a perimeter of 27 km and can accelerate two beams of protons to an energy of 6.5 TeV resulting in a collision with a centre of mass energy of 13 TeV. It is home to several experiments, the biggest of those are ATLAS and the Compact Muon Solenoid (CMS). Both are general-purpose detectors to investigate the particles that form during particle collisions.

Particle colliders are characterized by their luminosity L. It is a quantity to be able to calculate the number of events per second generated in a collision by $N_{event} = L\sigma_{event}$ with σ_{event} being the cross section of the event. The luminosity of the LHC for a Gaussian beam distribution can be described as follows:

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \tag{2}$$

Where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ_r the relativistic gamma factor, ϵ_n the normalised transverse beam emittance, β^* the beta function at the collision point and F the geometric luminosity reduction factor due to the crossing angle at the interaction point:

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{-1/2} \tag{3}$$

At the maximum luminosity of $10^{34}\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, $N_b = 1.15 \cdot 10^{11}$, $n_b = 2808$, $f_{rev} = 11.2\,\mathrm{kHz}$, $\beta^* = 0.55\,\mathrm{m}$, $\epsilon_n = 3.75\,\mu\mathrm{m}$ and $F = 0.85$.

To quantify the amount of data collected by one of the experiments at LHC, the integrated luminosity is introduced as $L_{int} = \int L dt$.

explain pdf -> not all 13 TeV available for collision

3.2 Compact Muon Solenoid

The data used in this thesis was recorded by the Compact Muon Solenoid (CMS). It is one of the four main experiments at the Large Hadron Collider. It can detect all elementary particles of the standard model except neutrinos. For that, it has an onion like setup, as can be seen in fig. 4. The particles produced in a collision first go through a tracking system. They then pass an electromegnetic as well as a hadronic calorimeter. This part is surrounded by a superconducting solenoid that generates a magenetic field of 3.8 T. Outside of the solenoid are big muon chambers. In 2016 the CMS captured data of an integrated luminosity of 37.80 fb^{-1}. In 2017 it collected 44.98 fb^{-1} and in 2018 63.67 fb^{-1}. Of that data, in 2016 35.92 fb^{-1}, in 2017 41.53 fb^{-1} and in 2018 59.74 fb^{-1} were usable for research. Therefore the combined integrated luminosity of data usable for research is 137.19 fb^{-1}.

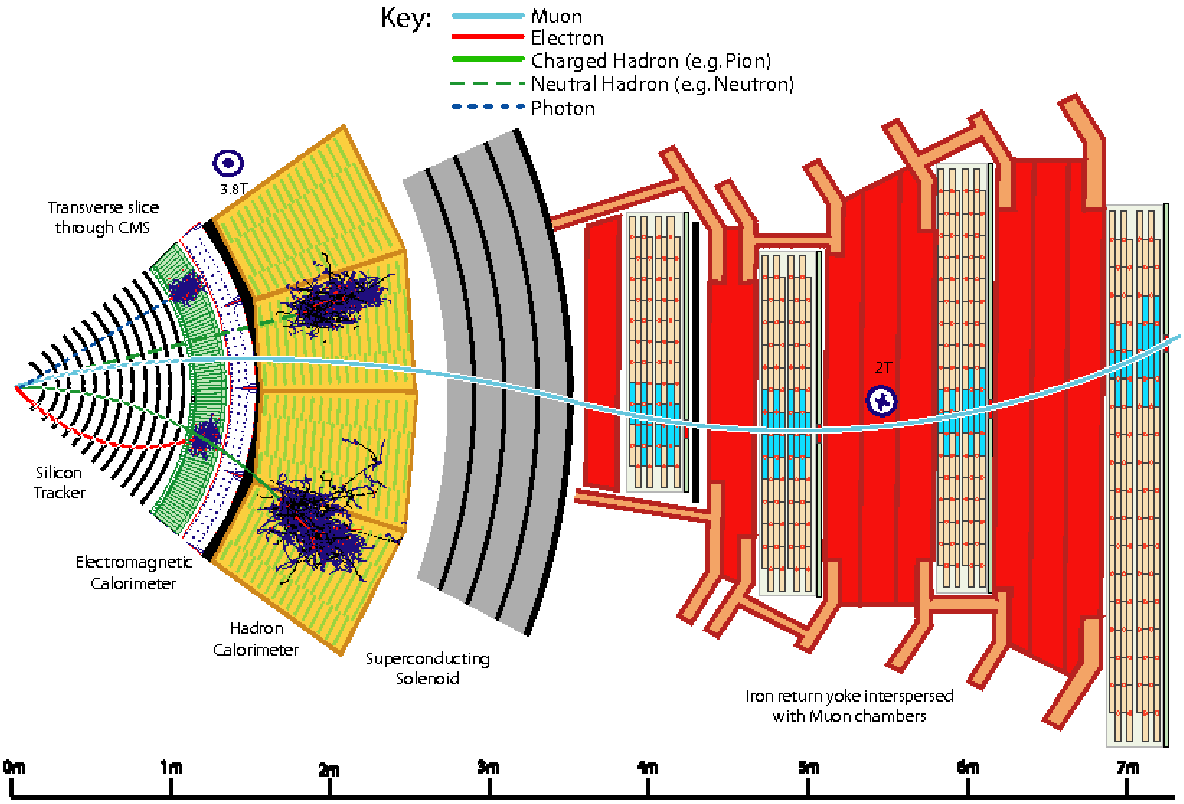

Figure 4: The setup of the Compact Muon Solenoid showing its onion like structure, the different detector parts and where different particles are detected [13].

3.2.1 Coordinate conventions

Per convention, the z axis points along the beam axis in the direction of the magnetic fields of the solenoid, the y axis upwards and the x axis horizontal towards the LHC centre. The azimuthal angle ϕ, which describes the angle in the x - y plane, the polar angle θ, which describes the angle in the y - z plane and the pseudorapidity η, which is defined as $\eta = -ln\left(tan\frac{\theta}{2}\right)$ are also introduced. The coordinates are visualised in fig. 5. Furthermore, to describe a particle's momentum, often the transverse momentum, p_t is used. It is the component of the momentum transversal to the beam axis. Before the collision, the transverse momentum obviously has to be zero, therefore, due to conservation of energy, the sum of all transverse momenta after the collision has to be zero, too. If this is not the case for the detected events, it implies particles that weren't detected such as neutrinos.

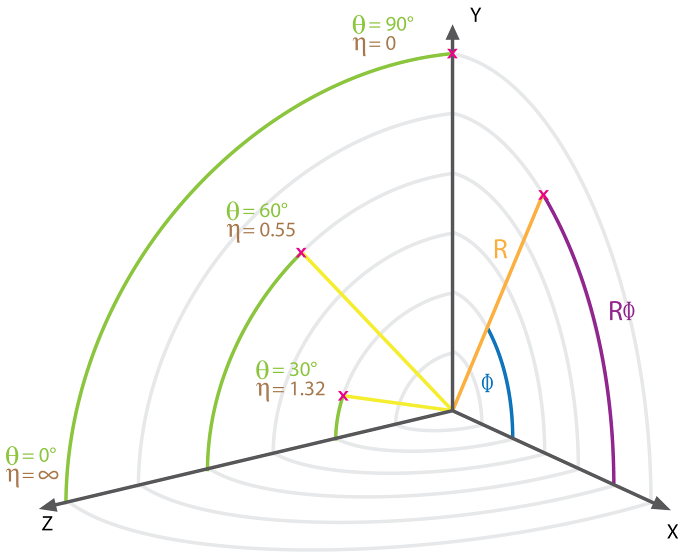

Figure 5: Coordinate conventions of the CMS illustrating the use of η and ϕ. The Z axis is in beam direction. Taken from https://inspirehep.net/record/1236817/plots

3.2.2 The tracking system

The tracking system is built of two parts, closest to the collision is a pixel detector and around that silicon strip sensors. They are used to reconstruct the tracks of charged particles, measuring their charge sign, direction and momentum. They are as close to the collision as possible to be able to identify secondary vertices.

3.2.3 The electromagnetic calorimeter

The electromagnetic calorimeter measures the energy of photons and electrons. It is made of tungstate crystal and photodetectors. When passed by particles, the crystal produces light in proportion to the particle's energy. This light is measured by the photodetectors that convert this scintillation light to an electrical signal. To measure a particles energy, it has to leave its whole energy in the ECAL, which is true for photons and electrons, but not for other particles such as hadrons and muons. Those have are of higher energy and therefore only leave some energy in the ECAL but are not stopped by it.

3.2.4 The hadronic calorimeter

The hadronic calorimeter (HCAL) is used to detect high energy hadronic particles. It surrounds the ECAL and is made of alternating layers of active and absorber material. While the absorber material with its high density causes the hadrons to shower, the active material then detects those showers and measures their energy, similar to how the ECAL works.

3.2.5 The solenoid

The solenoid, giving the detector its name, is one of the most important features. It creates a magnetic field of 3.8 T and therefore makes it possible to measure momentum of charged particles by bending their tracks.

3.2.6 The muon system

Outside of the solenoid there is only the muon system. It consists of three types of gas detectors, the drift tubes, cathode strip chambers and resistive plate chambers. It covers a total of $0 < |\eta| < 2.4$. The muons are the only detected particles, that can pass all the other systems without a significant energy loss.

3.2.7 The Trigger system

The CMS features a two level trigger system. It is necessary because the detector is unable to process all the events due to limited bandwidth. The Level 1 trigger reduces the event rate from 40 MHz to 100 kHz, the software based High Level trigger is then able to further reduce the rate to 1 kHz. The Level 1 trigger uses the data from the electromagnetic and hadronic calorimeters as well as the muon chambers to decide whether to keep an event. The High Level trigger uses a streamlined version of the CMS offline reconstruction software for its decision making.

3.2.8 The Particle Flow algorithm

The particle flow algorithm is used to identify and reconstruct all the particles arising from the proton - proton collision by using all the information available from the different sub-detectors of the CMS. It does so by extrapolating the tracks through the different calorimeters and associating clusters they cross with them. The set of the track and its clusters is then no more used for the detection of other particles. This is first done for muons and then for charged hadrons, so a muon can't give rise to a wrongly identified charged hadron. Due to Bremsstrahlung photon emission, electrons are harder to reconstruct. For them a specific track reconstruction algorithm is used. After identifying charged hadrons, muons and electrons, all remaining clusters within the HCAL correspond to neutral hadrons and within ECAL to photons. If the list of particles and their corresponding deposits is established, it can be used to determine the particles four momenta. From that, the missing transverse energy can be calculated and tau particles can be reconstructed by their decay products.

3.3 Jet clustering

Because of the hadronisation it is not possible to uniquely identify the originating particle of a jet. Nonetheless, several algorithms exist to help with this problem. The algorithm used in this thesis is the anti-k_t clustering algorithm. It arises from a generalization of several other clustering algorithms, namely the k_t, Cambridge/Aachen and SISCone clustering algorithms.

The anti-k_t clustering algorithm associates hard particles with their soft particles surrounding them within a radius $R = \sqrt{\eta^2 - \phi^2}$ in the η - ϕ plane forming cone like jets. If two jets overlap, the jets shape is changed according to its hardness in regards to the transverse momentum. A softer particles jet will change its shape more than a harder particles. A visual comparison of four different clustering algorithms can be seen in fig. 6. For this analysis, a radius of 0.8 is used.

Furthermore, to approximate the mass of a heavy particle that caused a jet, the soft-drop mass can be used. It is calculated by removing wide angle soft particles from the jet to counter the effects of contamination from initial state radiation, underlying event and multiple hadron scattering. It therefore is more accurate in determining the mass of a particle causing a jet than taking the mass of all constituent particles of the jet combined.

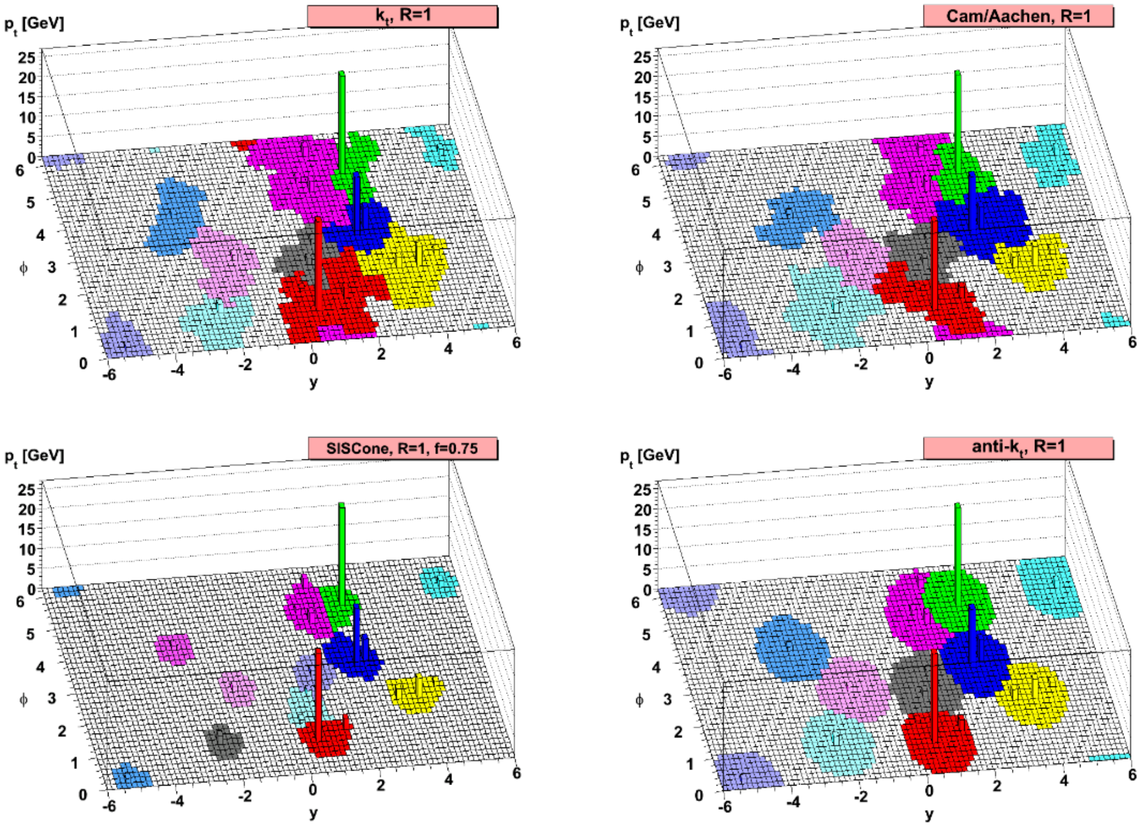

Figure 6: Comparison of the k_t, Cambridge/Aachen, SISCone and anti-k_t algorithms clustering a sample parton-level event with many random soft ‘ghosts’. Taken from [5]

Fig. 6 shows, that the jets reconstructed using the anti-k_t algorithm have the clearest cone like shape and is therefore chosen for this thesis.

4 Method of analysis

This section gives an overview over how the data gathered by the LHC and CMS is going to be analysed to be able to either exclude the q* particle to even higher masses than already done or maybe confirm its existence.

As described in sec. 2.2, the decay of the q* particle to a quark and a vector boson with the vector boson then decaying hadronically will be investigated. This is the second most probable decay of the q* particle and easier to analyse than the dominant decay to a quark and a gluon. Therefore it is a good choice for this research. The decay q* $\rightarrow$ qV + q $\rightarrow$ $q\bar{q}$ + q results in two jets, because the decay products of the heavy vector boson are highly boosted, causing them to be very close together and therefore be reconstructed as one jet. The dijet invariant mass of those two jets, which is identical to the mass of the q* particle, is reconstructed. The only background considered is the QCD background described in sec. 2.2.1. A selection using different kinematic variables as well as a tagger to identify jets from the decay of a vector boson is introduced to reduce the background and increase the sensitivity for the signal. After that, it will be looked for a peak in the dijet invariant mass distribution at the resonance mass of the q* particle.

The data studied were collected by the CMS experiment in the years 2016, 2017 and 2018. They are analysed with the Particle Flow algorithm to reconstruct jets and all the other particles forming during the collision. The jets are then clustered using the anti-k_t algorithm with the distance parameter R being 0.8.

The analysis will be conducted with two different sets of data. First, only the data collected by CMS in 2016 will be used to compare the results to the previous analysis [6]. Then the combined data from 2016, 2017 and 2018 will be used to improve the previously set limits for the mass of the q* particle. Also, two different V-tagging mechanisms will be used to compare their performance. One based on the N-subjettiness variable used in the previous research [6], the other being a novel approach using a deep neural network, that will be explained in the following.

4.1 Signal and Background modelling

To make sure the setup is working as intended, at first simulated samples of background and signal are used. In those Monte Carlo simulations, the different particle interactions that take place in a proton - proton collision are simulated using the probabilities provided by the Standard Model by calculating the cross sections of the different feynman diagrams. Later on, also detector effects (like its limited resolution) are applied to make sure, they look like real data coming from the CMS detector. The q* signal samples are simulated by the probabilities given by the q* theory [2] and assuming a cross section of 1 pb^{-1}. The simulation was done using MadGraph. Because of the expected high mass, the signal width will be dominated by the resolution of the detector, not by the natural resonance width.

The dijet invariant mass distribution of the QCD background is expected to smoothly fall with higher masses. It is therefore fitted using the following smooth falling function with three parameters p0, p1, p2:

$$\frac{dN}{dm_{jj}} = \frac{p_0 \cdot (1 - m_{jj}/\sqrt{s})^{p_2}}{(m_{jj}/\sqrt{s})^{p_1}} \tag{4}$$

Whereas m_{jj} is the invariant mass of the dijet and p_0 is a normalisation parameter. It is the same function as used in the previous research studying 2016 data only.

The signal is fitted using a double sided crystal ball function. It has six parameters:

- mean: the functions mean, in this case the resonance mass
- sigma: the functions width, in this case the resolution of the detector
- n1, n2, alpha1, alpha2: parameters influencing the shape of the left and right tail

A gaussian and a poisson function have also been studied but found to be not able to reproduce the signal shape as they couldn't model the tails on both sides of the peak.

A linear combination of the signal and background function is then fitted to a toy dataset with gaussian errors and a simulated signal cross section of 1 pb^{-1}. The resulting coefficients of said combination then show the expected signal rate for the simulated cross section. An example of such a fit can be seen in fig. 7. In this figure, a binning of 200 GeV is used. For the actual analysis a 1 GeV binning will be used. It can be seen that the fit works very well and therefore confirms the functions chosen to model signal and background. This is supported by a $\chi^2/$ ndof of 0.5 and a found mean for the signal at 2999 ± 23 GeV which is extremely close to the expected 3000 GeV mean. Those numbers clearly show that the method in use is able to successfully describe the data.

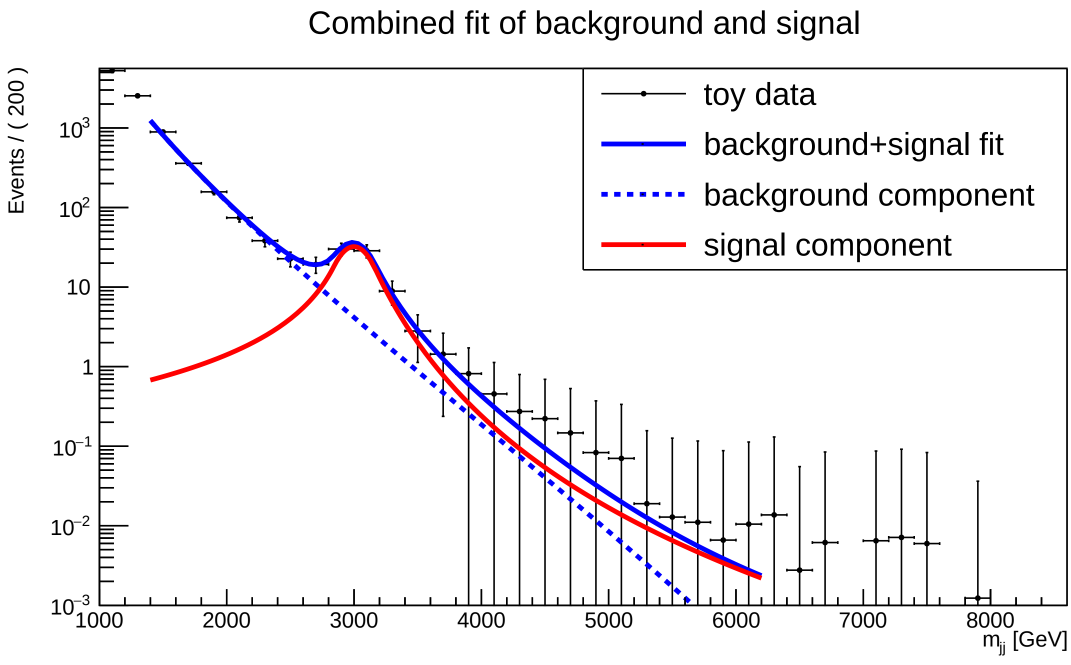

Figure 7: Combined fit of signal and background on a toy dataset with gaussian errors and a simulated resonance mass of 3 TeV.

5 Preselection and data quality

To reduce the background and increase the signal sensitivity, a selection of events by different variables is introduced. It is divided into two stages. The first one (the preselection) adds some general physics motivated selection using kinematic variables and is also used to ensure a high trigger efficiency. In the second part, different taggers will be used as a discriminator between QCD background and signal events. After the preselection, it is made sure, that the simulated samples represent the real data well by comparing the data with the simulation in the signal as well as a sideband region, where no signal events are expected.

5.1 Preselection

First, all events are cleaned of jets with a $p_t < 200\,\text{GeV}$ and a pseudorapidity $|\eta| > 2.4$. This is to discard soft background and to make sure the particles are in the barrel region of the detector for an optimal track reconstruction. Furthermore, all events with one of the two highest p_t jets having an angular separation smaller than 0.8 from any electron or muon are discarded to allow future use of the results in studies of the semi or all-leptonic decay channels.

From a decaying q* particle, two jets are expected in the final state. The dijet invariant mass of those two jets will be used to reconstruct the mass of the q* particle. Therefore a cut is added to have at least 2 jets, accounting for the possibility of more jets, for example caused by gluon radiation of a quark or other QCD effects. If this is the case, the two jets with the highest p_t are used for the reconstruction of the q* mass. The distributions of the number of jets before and after the selection can be seen in fig. 8.

The next selection is done using $\Delta\eta = |\eta_1 - \eta_2|$, with η_1 and η_2 being the η of the two jets with the highest transverse momentum. The q* particle is expected to be very heavy in regards to the center of mass energy of the collision and will therefore be almost stationary. Its decay products should therefore be close to back to back, which means the $\Delta\eta$ distribution is expected to peak at 0. At the same time, particles originating from QCD effects are expected to have a higher $\Delta\eta$ as they mainly form from less heavy resonances. To maintain comparability, the same selection as in previous research of $\Delta\eta \leq 1.3$ is used. A comparison of the $\Delta\eta$ distribution before and after the selection can be seen in fig. 9.

The last selection in the preselection is on the dijet invariant mass: $m_{jj} \geq 1050\,\text{GeV}$. It is important for a trigger efficiency higher than 99 % with a soft-drop mass cut of $m_{SDM} > 65$ GeV applied to the jet with the highest transverse momentum. A comparison of its distribution before and after the selection can be seen in fig. 10. Also, it has a huge impact on the background because it usually consists of way lighter particles. The q* on the other hand is expected to have a very high invariant mass of more than 1 TeV. The m_{jj} distribution should be a smoothly falling function for the QCD background and peak at the simulated resonance mass for the signal events.

After the preselection, the signal efficiency for q* decaying to qW of 2016 ranges from 48 %

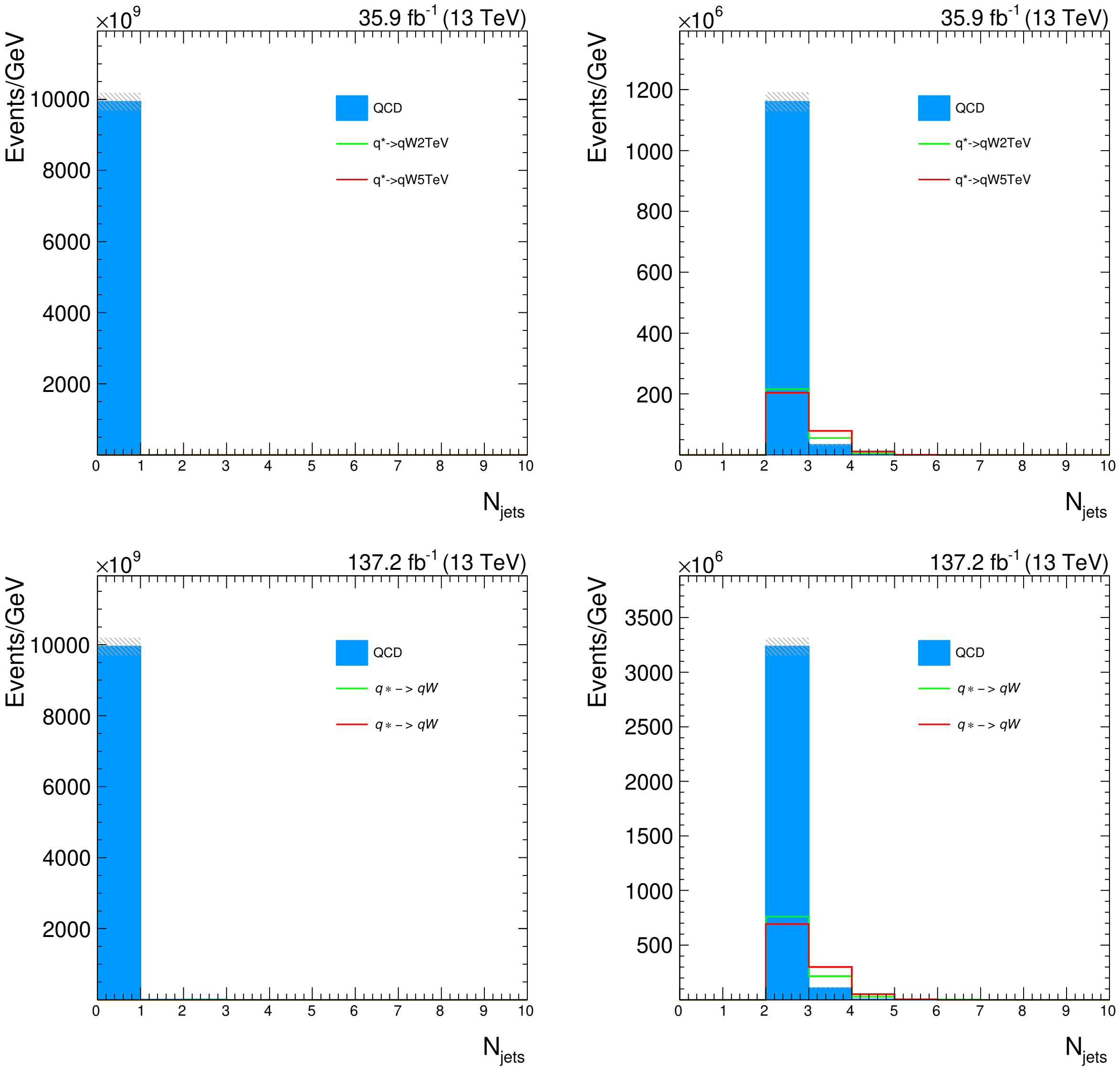

Figure 8: Comparison of the number of jet distribution before and after the cut at number of jets ≥ 2. Left: distribution before the cut. Right: distribution after the cut. 1st row: data from 2016. 2nd row: combined data from 2016, 2017 and 2018. The signal curves are amplified by a factor of 10,000, to be visible.

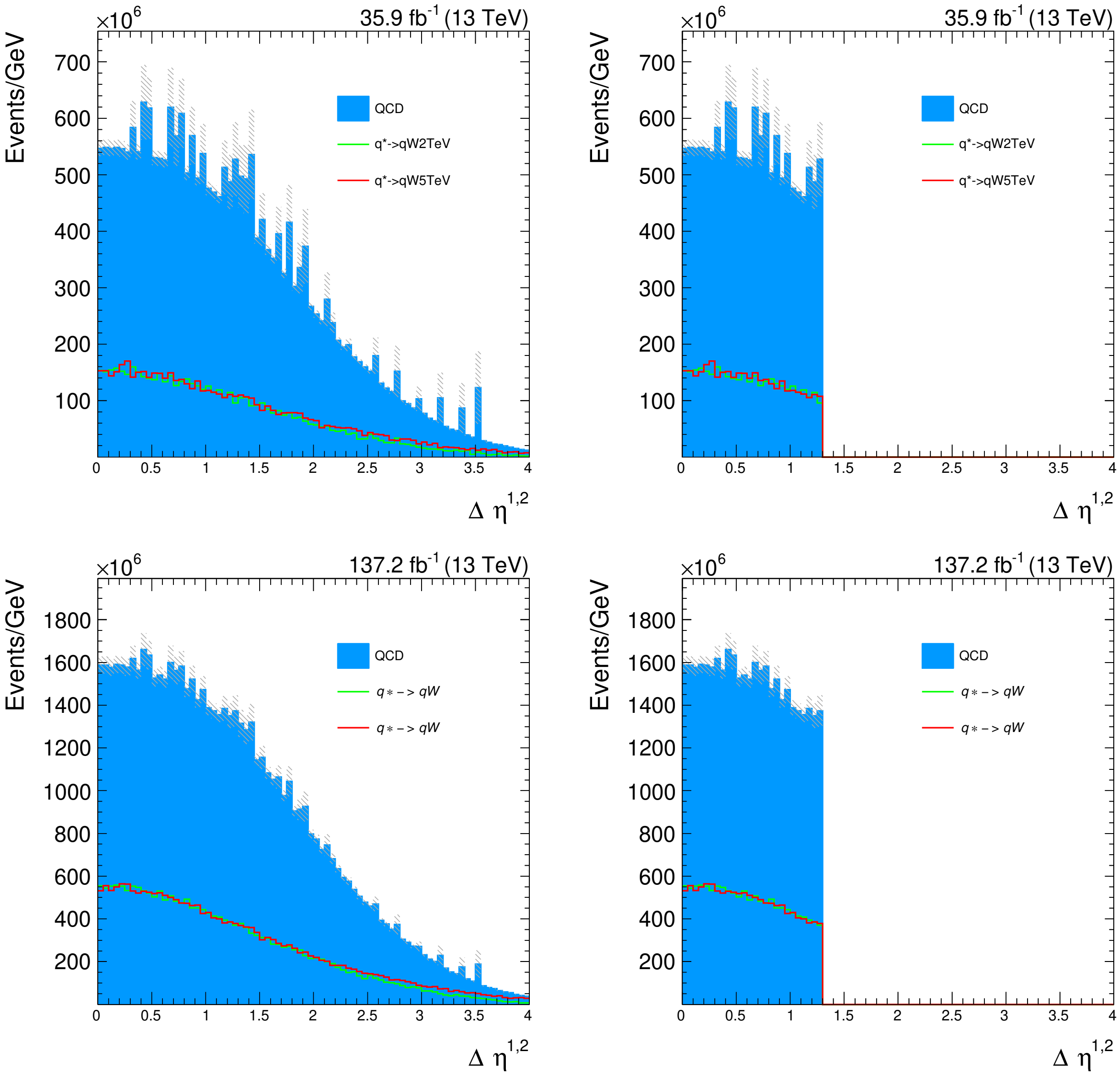

Figure 9: Comparison of the $\Delta\eta$ distribution before and after the cut at $\Delta\eta \leq 1.3$. Left: distribution before the cut. Right: distribution after the cut. 1st row: data from 2016. 2nd row: combined data from 2016, 2017 and 2018. The signal curves are amplified by a factor of 10,000, to be visible.

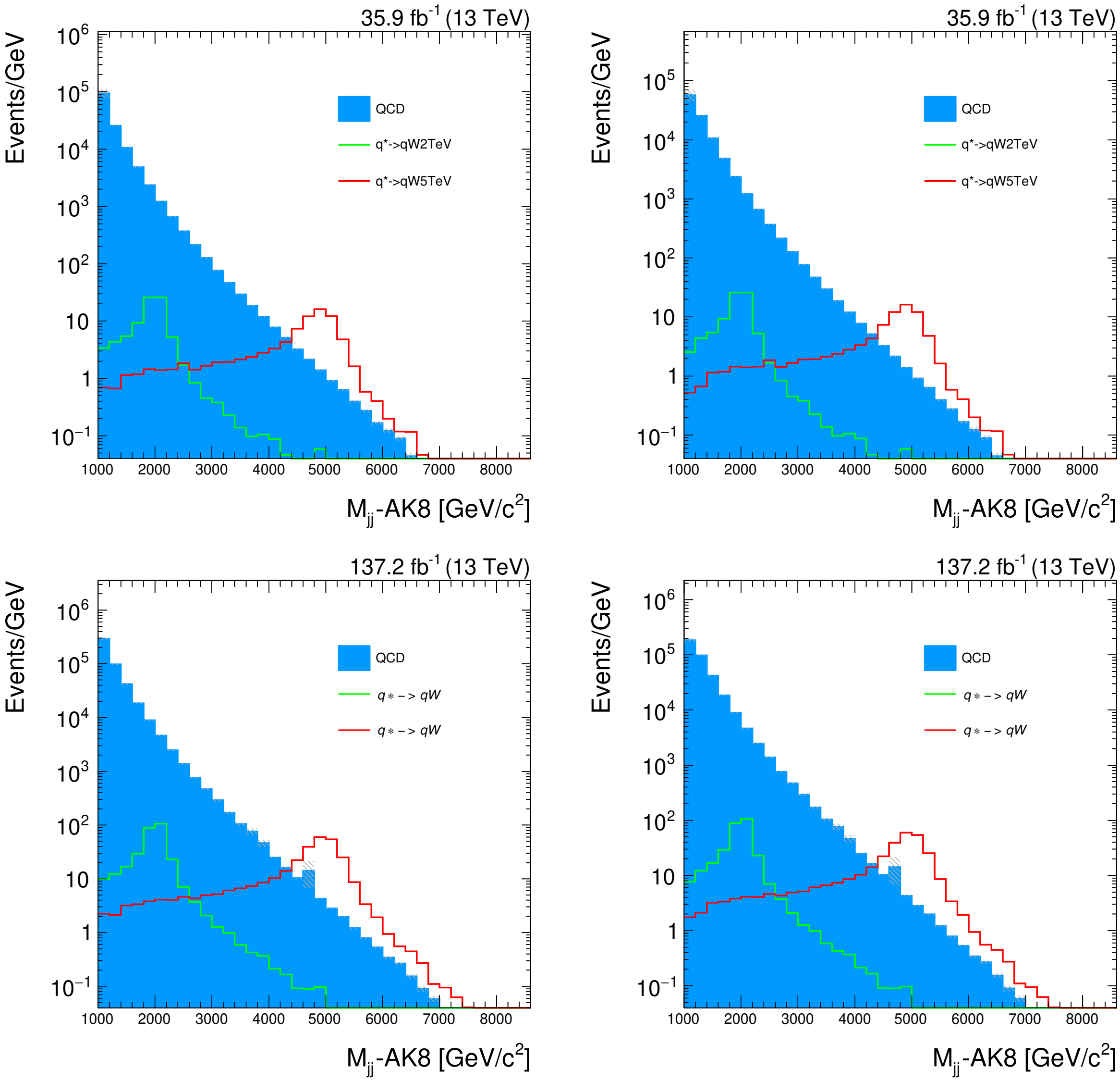

Figure 10: Comparison of the invariant mass distribution before and after the cut at $m_{jj} \geq 1050\,\text{GeV}$. It shows the expected smooth falling functions of the background whereas the signal peaks at the simulated resonance mass. Left: distribution before the cut. Right: distribution after the cut. 1st row: data from 2016. 2nd row: combined data from 2016, 2017 and 2018.

for 1.6 TeV to 49 % for 7 TeV. Decaying to qZ, the efficiencies are between 45 % (1.6 TeV) and 50 % (7 TeV). The amount of background after the preselection is reduced to 5 % of the original events. For the combined data of the three years those values look similar. Decaying to qW signal efficiencies between 49 % (1.6 TeV) and 56 % (7 TeV) are reached, wheres the efficiencies when decaying to qZ are in the range of 46 % (1.6 TeV) to 50 % (7 TeV). Here, the background could be reduced to 8 % of the original events. So while keeping around 50 % of the signal, the background was already reduced to less than a tenth.

5.2 Data - Monte Carlo Comparison

To ensure high data quality, the simulated QCD background sample is now being compared to the data of the corresponding year collected by the CMS detector. This is done for the year 2016 and for the combined data of years 2016, 2017 and 2018. The distributions are rescaled so the integral over the invariant mass distribution of data and simulation are the same. In fig. 11, the three distributions of the variables that were used for the preselection can be seen for year 2016 and the combined data of years 2016 to 2018. For analysing the data from the CMS, jet energy corrections have to be applied. Those are to calibrate the ECAL and HCAL parts of the CMS, so the energy of the detected particles can be measured correctly. The corrections used were published by the CMS group. [source needed, but not sure where to find it]

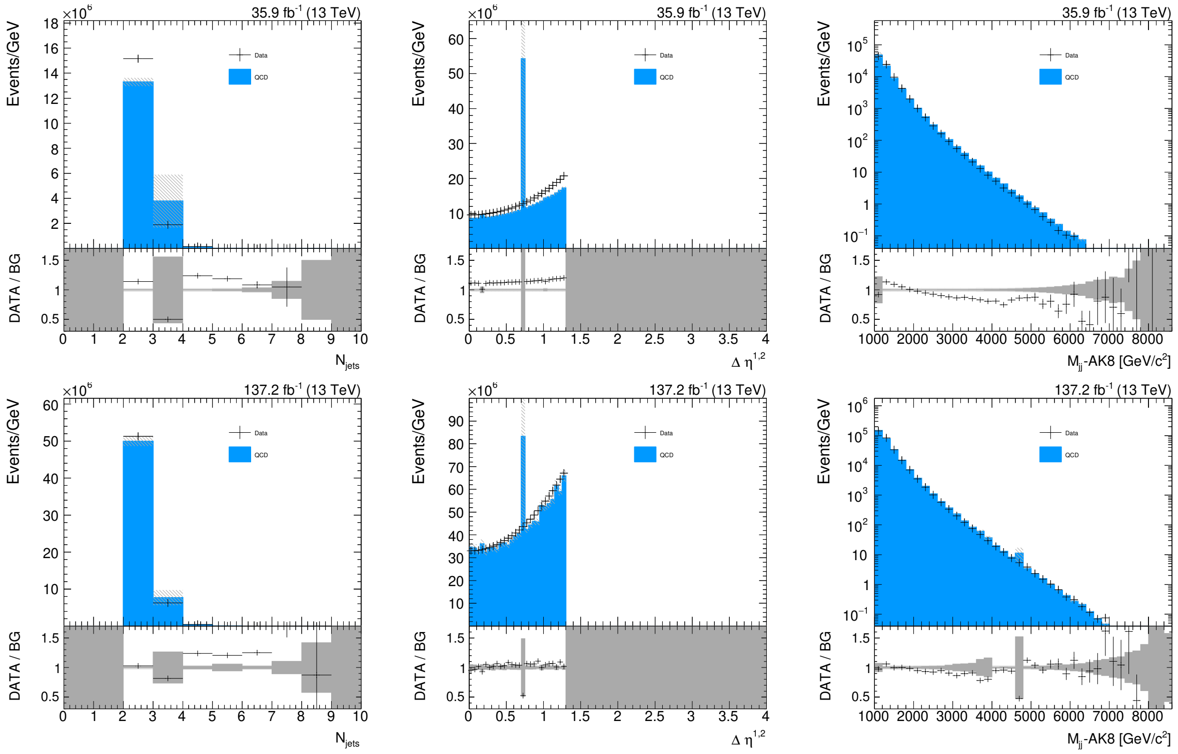

Figure 11: Comparision of data with the Monte Carlo simulation. 1st row: data from 2016. 2nd row: combined data from 2016, 2017 and 2018.

The shape of the real data matches the simulation well. The $\Delta\eta$ distributions shows some offset between data and simulation.

5.2.1 Sideband

The sideband is introduced to make sure no bias in the data and Monte Carlo simulation is introduced. It is a region in which no signal event is expected. Again, data and the Monte Carlo simulation are compared. For this analysis, the region where the soft-drop mass of both of the two jets with the highest transverse momentum (p_t) is more than 105 GeV was chosen. 105 GeV is well above the mass of 91 GeV of the Z boson, the heavier vector boson. Therefore it is very unlikely that a particle heavier than t In fig. 12, the comparison of data with simulation in the sideband region can be seen for the soft-drop mass distribution as well as the dijet invariant mass distribution. As in [fig:data-mc], the histograms are rescaled, so that the dijet invariant mass distributions of data and simulation have the same integral. It can be seen, that in the sideband region data and simulation match very well.

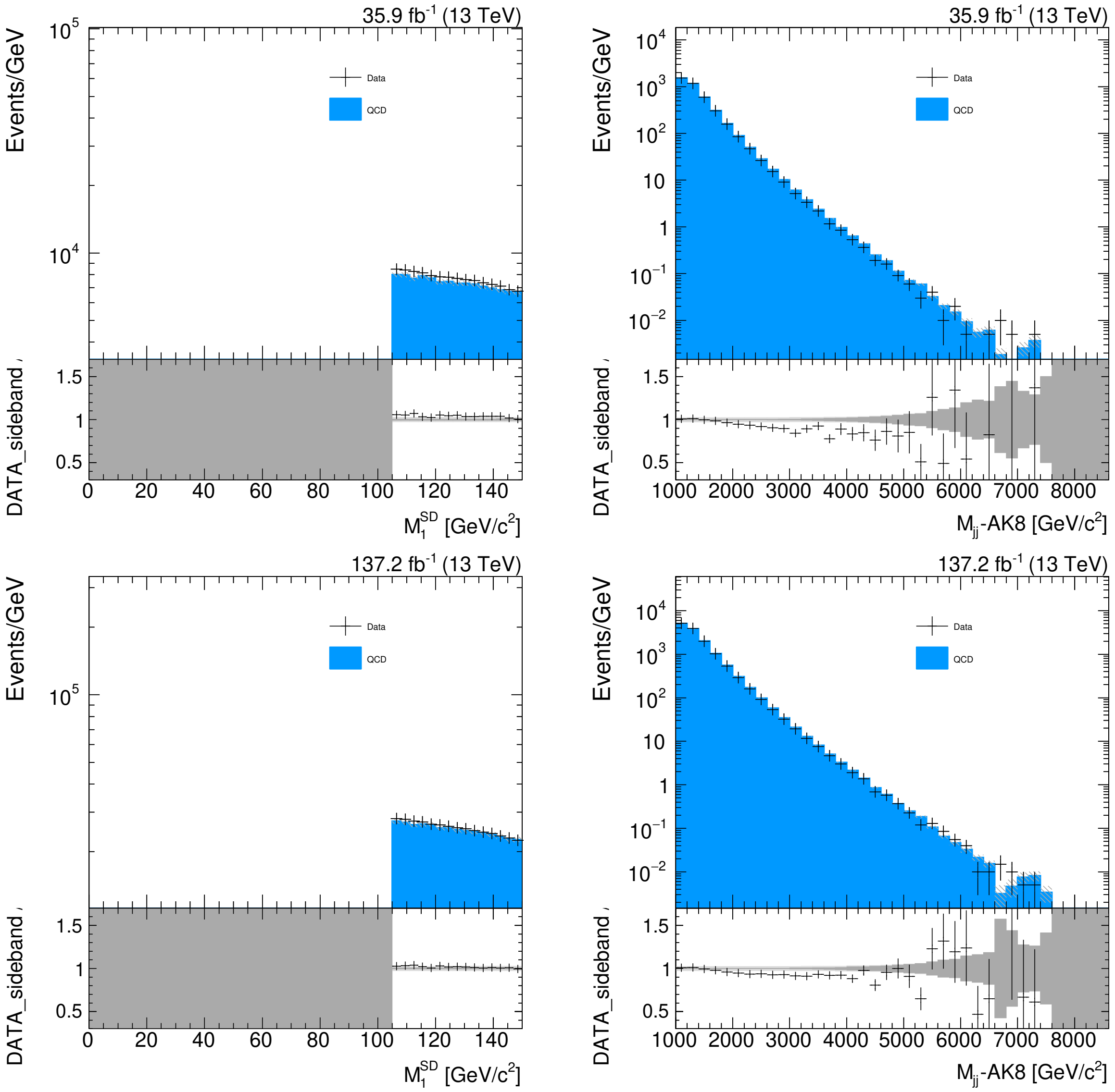

Figure 12: Comparison of data with the Monte Carlo simulation in the sideband region. 1st row: data from 2016. 2nd row: combined data from 2016, 2017 and 2018.

6 Jet substructure selection

So far it was made sure, that the data collected by the CMS and the simulation are in good agreement after the preselection and no unwanted side effects are introduced in the data by the used cuts. Now another selection has to be introduced, to further reduce the background to be able to look for the hypothetical signal events in the data.

This is done by distinguishing between QCD and signal events using a tagger to identify jets coming from a vector boson. Two different taggers will be used to later compare their performance. The decay analysed includes either a W or Z boson, which are, compared to the particles in QCD effects, very heavy. This can be used by adding a cut on the soft-drop mass of a jet. The soft-drop mass of at least one of the two leading jets is expected to be within 35 GeV and 105 GeV. This cut already provides a good separation of QCD and signal events, on which the two taggers presented next can build.

Both taggers provide a discriminator value to choose whether an event originates in the decay of a vector boson or from QCD effects. This value will be optimized afterwards to make sure the maximum efficiency possible is achieved.

6.1 N-Subjettiness

The N-subjettiness [15] τ_N is a jet shape parameter designed to identify boosted hadronically-decaying objects. When a vector boson decays hadronically, it produces two quarks each causing a jet. But because of the high mass of the vector bosons, the particles are highly boosted and appear, after applying a clustering algorithm, as just one. This algorithm now tries to figure out, whether one jet might consist of two subjets by using the kinematics and positions of the constituent particles of this jet. The N-subjettiness is defined as

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \cdot \min\{\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k}\} \tag{5}$$

with k going over the constituent particles in a given jet, $p_{T,k}$ being their transverse momenta and $\Delta R_{J,k} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ being the distance of a candidate subjet J and a constituent particle k in the η - ϕ plane. It quantifies to what degree a jet can be regarded as a jet composed of N subjets. Experiments showed, that rather than using τ_N directly, the ratio $\tau_{21} = \tau_2/\tau_1$ is a better discriminator between QCD events and events originating from the decay of a boosted vector boson.

The lower the τ_{21} is, the more likely a jet is caused by the decay of a vector boson. Therefore a selection will be introduced, so that τ_{21} of one candidate jet is smaller then some value that will be determined by an optimization process described in the next chapter. As candidate jet the one of the two highest p_t jets passing the soft-drop mass window is used. If both of them pass, the one with higher p_t is chosen.

6.2 DeepAK8

The DeepAK8 tagger [11] uses a deep neural network (DNN) to identify decays originating in a vector boson. It claims to reduce the background rate by up to a factor of ~10 with the same signal efficiency compared to non-machine-learning approaches like the N-Subjettiness method. This is supported by fig. 13, showing a comparision of background and signal efficiency of the DeepAK8 tagger, with, between others, the τ_{21} tagger also used in this analysis.

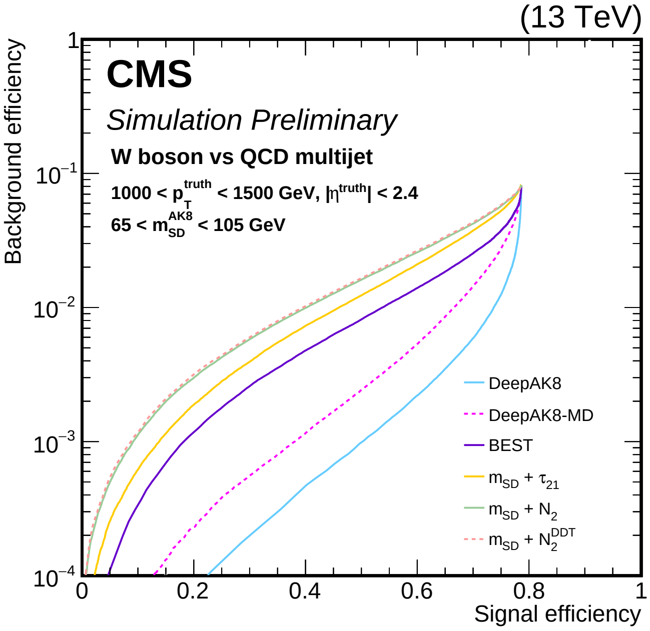

Figure 13: Comparison of tagger efficiencies, showing, between others, the DeepAK8 and τ_{21} tagger used in this analysis. Taken from [11]

The DNN has two input lists for each jet. The first is a list of up to 100 constituent particles of the jet, sorted by decreasing p_t. A total of 42 properties of the particles such es p_t, energy deposit, charge and the angular momentum between the particle and the jet or subjet axes are included. The second input list is a list of up to seven secondary vertices, each with 15 features, such as the kinematics, displacement and quality criteria. To process those inputs, a customised DNN architecture has been developed. It consists of two convolutional neural networks that each process one of the input lists. The outputs of the two CNNs are then combined and processed by a fully-connected network to identify the jet. The network was trained with a sample of 40 million jets, another 10 million jets were used for development and validation.

In this thesis, the mass decorrelated version of the DeepAK8 tagger is used. It adds an additional mass predictor layer, that is trained to quantify how strongly the output of the non-decorrelated tagger is correlated to the mass of a particle. Its output is fed back to the network as a penalty so it avoids using features of the particles correlated to their mass. The result is a largely mass decorrelated tagger of heavy resonances. As the mass variable is already in use for the soft-drop mass selection, this version of the tagger is to be preferred.

The higher the discriminator value of the deep boosted tagger, the more likely is the jet to be caused by decay of a vector boson. Therefore, using the same way to choose a candidate jet as for the N-subjettiness tagger, a selection is applied so that this candidate jet has a WvsQCD/ZvsQCD value greater than some value determined by the optimization presented next.

6.3 Optimization

To figure out the best value to cut on the discriminators introduced by the two taggers, a value to quantify how good a cut is has to be introduced. For that, the significance calculated by $\frac{S}{\sqrt{B}}$ will be used. S stands for the amount of signal events and B for the amount of background events in a given interval. This value assumes a gaussian error on the background so it will be calculated for the 2 TeV masspoint where enough background events exist to justify this assumption. It follows from the central limit theorem that states, that for identical distributed random variables, their sum converges to a gaussian distribution. The significance therefore represents how good the signal can be distinguished from the background in units of the standard deviation of the background. As interval, a 10 % margin around the resonance nominal mass is chosen. The significance is then calculated for different selections on the discriminant of the two taggers and then plotted in dependence on the minimum resp. maximum allowed value of the discriminant to pass the selection for the deep boosted resp. the N-subjettiness tagger.

The optimization process is done using only the data from year 2018, assuming the taggers have similar performances on the data of the different years.

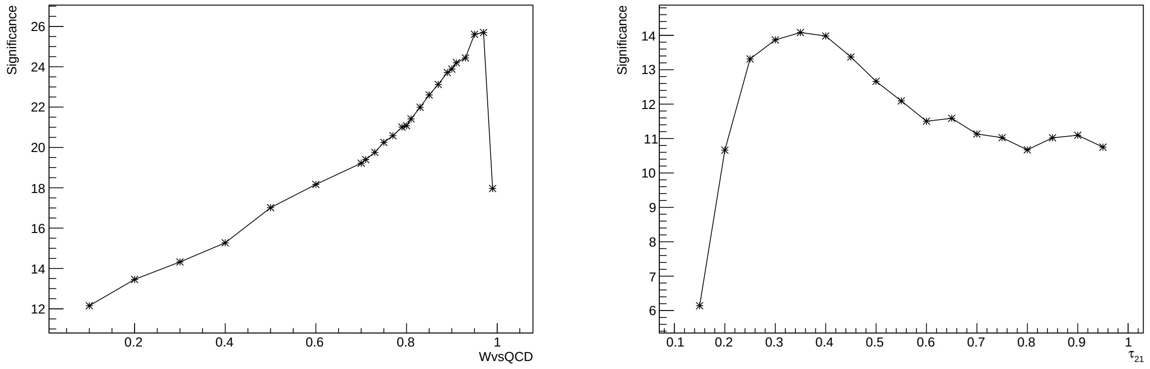

Figure 14: Significance plots for the deep boosted (left) and N-subjettiness (right) tagger at the 2 TeV masspoint.

As a result, the τ_{21} cut is placed at ≤ 0.35, confirming the value previous research chose and the deep boosted cut is placed at ≥ 0.95. For the deep boosted tagger, 0.97 would give a slightly higher significance but as it is very close to the edge where the significance drops very low and the higher the cut the less background will be left to calculate the cross section limits, especially at higher resonance masses, the slightly less strict cut is chosen. The significance for the τ_{21} cut is 14, and for the deep boosted tagger 26.

For both taggers also a low purity category is introduced for high TeV regions. Using the cuts optimized for 2 TeV, there are very few background events left for higher resonance masses, but to reliably calculate cross section limits, those are needed. As low purity category for the N-subjettiness tagger, a cut at $0.35 < \tau_{21} < 0.75$ is used. For the deep boosted tagger the opposite cut from the high purity category is used: $VvsQCD < 0.95$.

7 Signal extraction

After the optimization, now the optimal selection for the N-subjettiness as well as the deep boosted tagger is found and applied to the simulated samples as well as the data collected by the CMS. The fit described in sec. 4 is performed for all masspoints of the decay to qW and qZ and for both datasets used, the one from 2016 und the combined one of years 2016, 2017 and 2018.

To test for the presence of a resonance in the data, the cross section limits of the signal event are calculated using a frequentist asymptotic limit calculator described in [7]. Using the parameters and signal rate obtained using the method described in sec. 4 as well as a shape analysis on the data recorded by the CMS, it determines an expected and an observed cross section limit by doing a signal + background versus background-only hypothesis test. It also calculates upper and lower limits of the expected cross section corresponding to a confidence level of 95 %.

If there's no resonance of the q* particle in the data, the observed limit should lie within the 2σ environment, meaning a 95 % confidence level, of the expected limit. This observed limit is plotted together with a theory line, representing the cross section limits expected, if the q* predicted by [2] would exist. The crossing of the theory line with the observed limit is then calculated, to have a limit of mass up to which the existence of the q* particle can be excluded. To find the uncertainty of this result, the crossing of the theory line plus, respectively minus, its uncertainty with the observed limit is also calculated.

7.1 Uncertainties

For calculating the cross section of the signal, four sources of uncertainties are considered.

First, the uncertainty of the Jet Energy Corrections. When measuring a particle's energy with the ECAL or HCAL part of the CMS, the electronic signals send by the photodetectors in the calorimeters have to be converted to actual energy values. Therefore an error in this calibration causes the energy measured to be shifted to higher or lower values causing also the position of the signal peak in the m_{jj} distribution to vary. The uncertainty is approximated to be 2 %.

Second, the tagger is not perfect and therefore some events, that don't originate from a V boson are wrongly chosen and on the other hand sometimes events that do originate from one are not. It influences the events chose for analysis and is therefore also considered as an uncertainty, which is approximated to be 6 %.

Third, the uncertainty of the parameters of the background fit is also considered, as it might change the background shape a little and therefore influence how many signal and background events are reconstructed from the data.

Fourth, the uncertainty on the Luminosity of the LHC of 2.5 % is also taken into account for the final results.

8 Results

This chapter will start by presenting the results for the data of year 2016 using both taggers and comparing it to the previous research [6]. It will then go on showing the results for the combined dataset, again using both taggers comparing their performances.

8.1 2016

Using the data collected by the CMS experiment on 2016, the cross section limits seen in fig. 15 were obtained.

As described in sec. 7, the calculated cross section limits are used to then calculate a mass limit, meaning the lowest possible mass of the q* particle, by finding the crossing of the theory line with the observed cross section limit. In fig. 15 it can be seen, that the observed limit in the region where theory and observed limit cross is very high compared to when using the N-subjettiness tagger. Therefore the two lines cross earlier, which results in lower exclusion limits on the mass of the q* particle causing the deep boosted tagger to perform worse than the N-subjettiness tagger in regards of establishing those limits as can be seen in tbl. 2. The table also shows the upper and lower limits on the mass found by calculating the crossing of the theory plus resp. minus its uncertainty. Due to the theory and the observed limits line being very flat in the high TeV region, even a small uncertainty of the theory can cause a high difference of the mass limit.

Table 2: Mass limits found using the data collected in 2016

Decay	Tagger	Limit [TeV]	Upper Limit [TeV]	Lower Limit [TeV]
qW	τ_{21}	5.39	6.01	4.99
qW	deep boosted	4.96	5.19	4.84
qZ	τ_{21}	4.86	4.96	4.70
qZ	deep boosted	4.49	4.61	4.40

8.1.1 Previous research

The limit established by using the N-subjettiness tagger on the 2016 data is already slightly higher than the one from previous research, which was found to be 5 TeV for the decay to qW

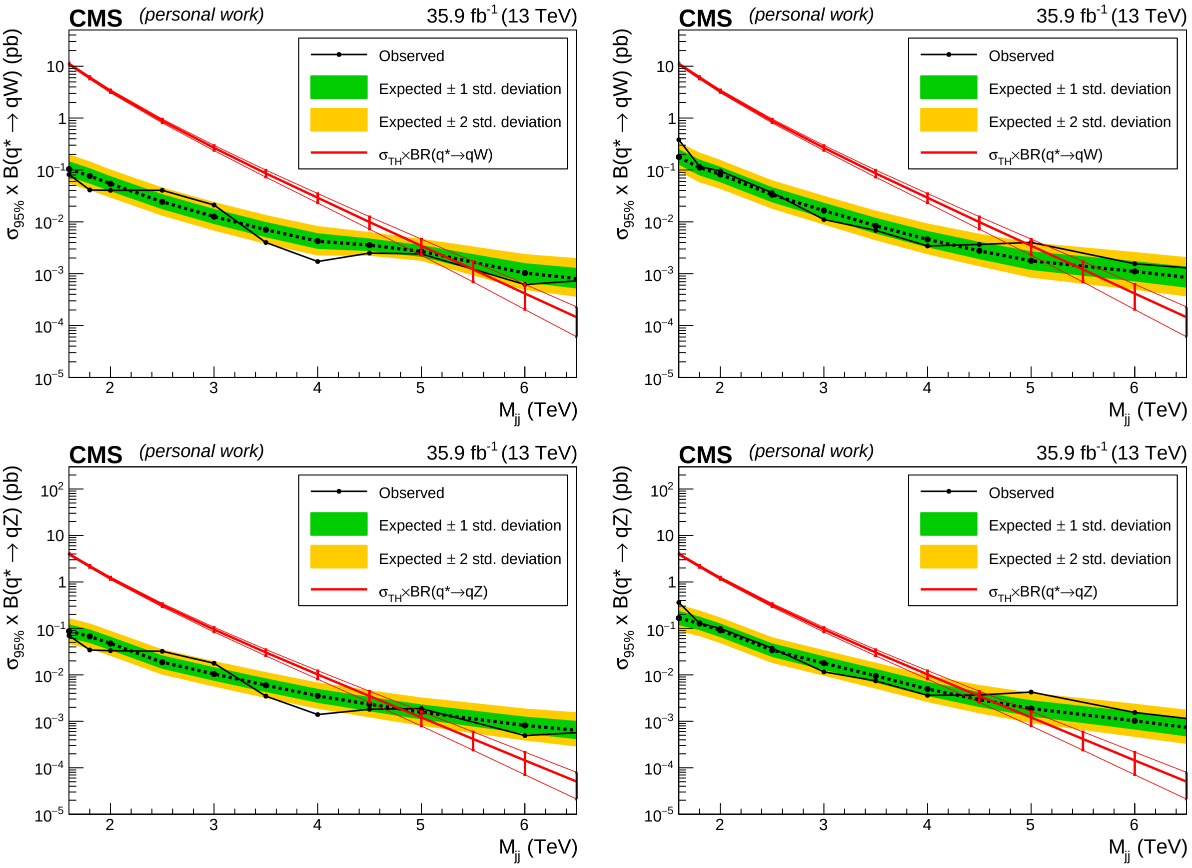

Figure 15: Results of the cross section limits for 2016 using the τ_{21} tagger (left) and the deep boosted tagger (right).

and 4.7 TeV for the decay to qZ. This is mainly due to the fact, that in our data, the observed limit at the intersection point happens to be in the lower region of the expected limit interval and therefore causing a very late crossing with the theory line when using the N-subjettiness tagger (as can be seen in fig. 15). This could be caused by small differences of the setup used or slightly differently processed data. Comparing the expected limits, there is a difference between 3 % and 30 %, between the values calculated by this thesis compared to the previous research. It is not, however, that one of the two results was constantly lower or higher but rather fluctuating. Therefore it can be said, that the results are in good agreement. The cross section limits of the previous research can be seen in fig. 16.

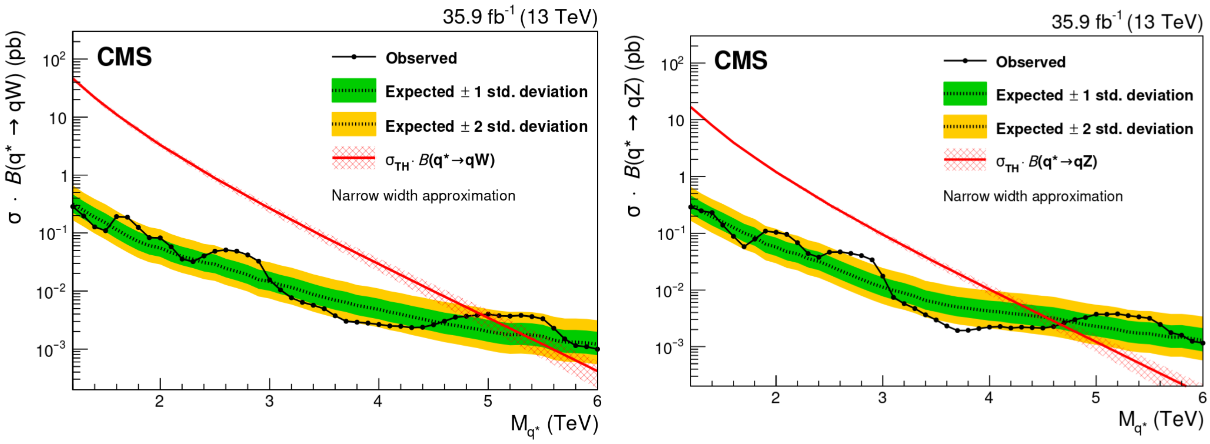

Figure 16: Previous results of the cross section limits for q decaying to qW (left) and q decaying to qZ (right). Taken from [6].

8.2 Combined dataset

Using the combined data, the cross section limits seen in fig. 17 were obtained. The cross section limits are, compared to only using the 2016 dataset, almost cut in half. This shows the big improvement achieved by using more than three times the amount of data.

The results for the mass limits of the combined years are as follows:

Table 3: Mass limits found using the data collected in 2016 - 2018

Decay	Tagger	Limit [TeV]	Upper Limit [TeV]	Lower Limit [TeV]
qW	τ_{21}	6.00	6.26	5.74
qW	deep boosted	6.11	6.31	5.39
qZ	τ_{21}	5.49	5.76	5.29
qZ	deep boosted	4.92	5.02	4.80

The combination of the three years not just improved the cross section limits, but also the limit for the mass of the q* particle. The final result is 1 TeV higher for the decay to qW and almost

0.8 TeV higher for the decay to qZ than what was concluded by the previous research [6].

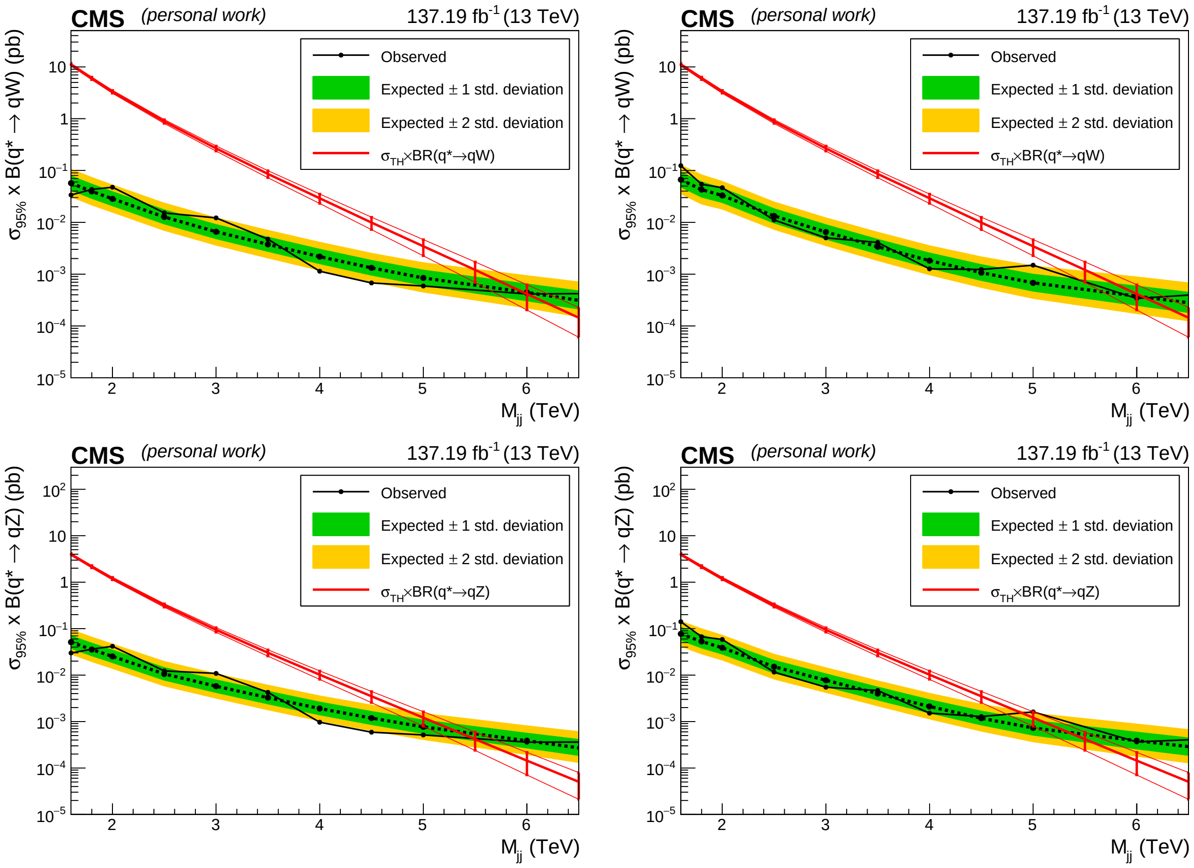

Figure 17: Results of the cross section limits for the three combined years using the τ_{21} tagger (left) and the deep boosted tagger (right).

8.3 Comparison of taggers

The previously shown results already show, that the deep boosted tagger was not able to significantly improve the results compared to the N-subjettiness tagger. For further comparison, in fig. 18 the expected limits of the different taggers for the $q^* \rightarrow qW$ and the $q^* \rightarrow qZ$ decay are shown. It can be seen, that the deep boosted is at best as good as the N-subjettiness tagger. This was not the expected result, as the deep neural network was already found to provide a higher significance in the optimisation done in sec. 6.3. The higher significance should also result in lower cross section limits. Apparently, doing the optimization only on data of the year 2018, was not the best choice. To make sure, there is no mistake in the setup, also the expected cross section limits using only the high purity category of the two taggers with 2018 data are compared in fig. 19. There, the cross section limits calculated using the deep boosted tagger are a bit lower than with the N-subjettiness tagger, showing, that the method used for optimisation was working but should have been applied to the combined dataset.

Recently, some issues with the training of the deep boosted tagger used in this analysis were also found, which might explain, why it didn't perform much better in general.

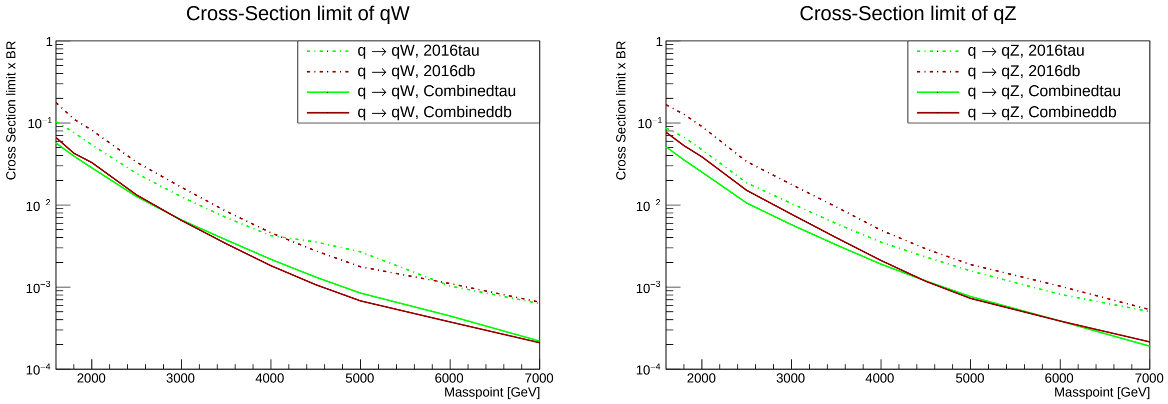

Figure 18: Comparison of expected limits of the different taggers using different datasets. Left: decay to qW. Right: decay to qZ

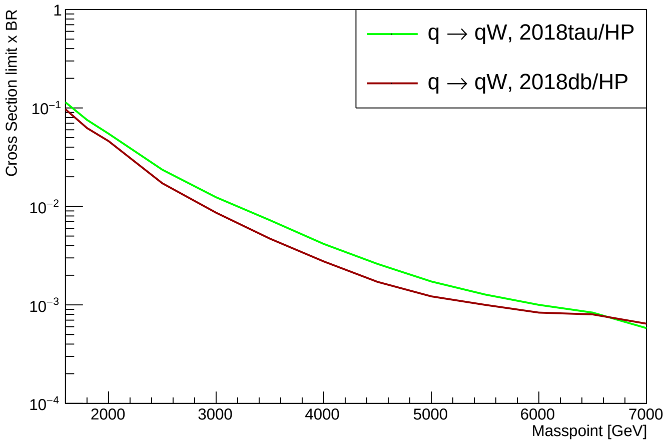

Figure 19: Comparision of deep boosted and N-subjettiness tagger in the high purity category using the data from year 2018.

9 Summary

In this thesis, a limit on the mass of the q* particle has been successfully established. By combining the data from the years 2016, 2017 and 2018, collected by the CMS experiment, the previously set limit could be significantly improved.

For the data analysis, the following selection was applied:

- #jets >= 2
- $\Delta\eta < 1.4$
- $m_{jj} >=$ 1050 GeV
- 35 GeV $< m_{SDM} <$ 105 GeV

For the deep boosted tagger, a high purity category of $VvsQCD > 0.95$ and a low purity category of $VvsQCD <= 0.95$ was used. For the N-subjettiness tagger the high purity category was $\tau_{21} < 0.35$ and the low purity category $0.35 < \tau_{21} < 0.75$. These values were found by optimizing for the highest possible significance of the signal.

After the selection, the cross section limits were extracted from the data and new exclusion limits for the mass of the q* particles set. These are 6.1 TeV by analyzing the decay to qW, respectively 5.5 TeV for the decay to qZ. Those limits are about 1 TeV higher than the ones found in previous research, that found them to be 5 TeV resp. 4.7 TeV.

Two different taggers were used to compare the result. The newer deep boosted tagger was found to not improve the result over the older N-subjettiness tagger. This was rather unexpected but might be caused by some training issues, that were identified lately.

This research can also be used to test other theories of the q* particle that predict its existence at lower masses, than the one used, by overlaying the different theory curves in the plots shown in fig. 15 and fig. 17.

The optimization process used to find the optimal values for the discriminant provided by the taggers, was found to not be optimal. It was only done using 2018 data, with which the deep boosted tagger showed a higher significance than the N-subjettiness tagger. Apparently, the assumption, that the same optimization would apply to the data of the other years as well, did not hold. Using the combined dataset, the deep boosted tagger showed no better cross section limits than the N-subjettiness tagger, which are directly related to the significance used for the optimization. Therefore, with a better optimization and the fixed training issues of the deep boosted tagger, it is very likely, that the result presented could be further improved.

References

[1] G. Apollinari et al. ‘High Luminosity Large Hadron Collider HL-LHC’. In: *arXiv e-prints*, arXiv:1705.08830 (May 2017), arXiv:1705.08830. arXiv: 1705.08830 `[physics.acc-ph]`.

[2] U. BAUR, I. HINCHLIFFE and D. ZEPPENFELD. ‘EXCITED QUARK PRODUCTION AT HADRON COLLIDERS’. In: *International Journal of Modern Physics A* 02.04 (1987), pp. 1285–1297. DOI: 10.1142/S0217751X87000661. eprint: https://doi.org/10.1142/S0217751X87000661. URL: https://doi.org/10.1142/S0217751X87000661.

[3] G L Bayatian et al. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. There is an error on cover due to a technical problem for some items. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/922757.

[4] Florian Beaudette. ‘The CMS Particle Flow Algorithm’. In: *arXiv e-prints*, arXiv:1401.8155 (Jan. 2014), arXiv:1401.8155. arXiv: 1401.8155 `[hep-ex]`.

[5] Matteo Cacciari, Gavin P. Salam and Gregory Soyez. ‘The anti-k$_t$ jet clustering algorithm’. In: *Journal of High Energy Physics* 2008.4, 063 (Apr. 2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 `[hep-ph]`.

[6] CMS Collaboration. ‘Search for massive resonances decaying into WW, WZ, ZZ, qW, and qZ with dijet final states at $\sqrt{s} = 13$ TeV’. In: *arXiv e-prints*, arXiv:1708.05379 (Aug. 2017), arXiv:1708.05379. arXiv: 1708.05379 `[hep-ex]`.

[7] Glen Cowan et al. ‘Asymptotic formulae for likelihood-based tests of new physics’. In: *European Physical Journal C* 71, 1554 (Feb. 2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 `[physics.data-an]`.

[8] Stephen D. Ellis and Davison E. Soper. ‘Successive combination jet algorithm for hadron collisions’. In: *arXiv* 48.7 (Oct. 1993), pp. 3160–3166. DOI: 10.1103/PhysRevD.48.3160. arXiv: hep-ph/9305266 `[hep-ph]`.

[9] Lyndon Evans and Philip Bryant. ‘LHC Machine’. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08001.

[10] Andrew J. Larkoski et al. ‘Soft drop’. In: *Journal of High Energy Physics* 2014, 146 (May 2014), p. 146. DOI: 10.1007/JHEP05(2014)146. arXiv: 1402.2657 `[hep-ph]`.

[11] *Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment*. Tech. rep. CMS-PAS-JME-18-002. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2683870.

[12] K. P. N. Murthy. ‘An Introduction to Monte Carlo Simulation of Statistical physics Problem’. In: *arXiv e-prints*, cond-mat/0104167 (Apr. 2001), cond–mat/0104167. arXiv: cond-mat/0104167 `[cond-mat.stat-mech]`.

[13] A.M. Sirunyan et al. ‘Particle-flow reconstruction and global event description with the CMS detector. Particle-flow reconstruction and global event description with the CMS

detector'. In: *JINST* 12.CMS-PRF-14-001. CMS-PRF-14-001-004. 10 (June 2017). Replaced with the published version. Added the journal reference and DOI. All the figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/PRF-14-001 (CMS Public Pages), P10003. 82 p. DOI: 10.1088/1748-0221/12/10/P10003. URL: https://cds.cern.ch/record/2270046.

[14] M. Tanabashi et al. 'Review of Particle Physics'. In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[15] Jesse Thaler and Ken Van Tilburg. 'Identifying boosted objects with N-subjettiness'. In: *Journal of High Energy Physics* 2011, 15 (Mar. 2011), p. 15. DOI: 10.1007/JHEP03(2011)015. arXiv: 1011.2268 `[hep-ph]`.

[16] Mia Tosi. *The CMS trigger in Run 2*. Tech. rep. CMS-CR-2017-340. Geneva: CERN, Oct. 2017. DOI: 10.22323/1.314.0523. URL: https://cds.cern.ch/record/2290106.

[17] Bryan Webber. 'Hadronization'. In: *arXiv e-prints*, hep-ph/9411384 (Nov. 1994), hep–ph/9411384. arXiv: hep-ph/9411384 `[hep-ph]`.

Appendix

Table 4: Cross Section limits using 2016 data and the N-subjettiness tagger for the decay to qW

Mass [TeV]	Exp. limit [pb]	Upper limit [pb]	Lower limit [pb]	Obs. limit [pb]
1.6	0.10406	0.14720	0.07371	0.08165
1.8	0.07656	0.10800	0.05441	0.04114
2.0	0.05422	0.07605	0.03879	0.04043
2.5	0.02430	0.03408	0.01747	0.04052
3.0	0.01262	0.01775	0.00904	0.02109
3.5	0.00703	0.00992	0.00502	0.00399
4.0	0.00424	0.00603	0.00300	0.00172
4.5	0.00355	0.00478	0.00273	0.00249
5.0	0.00269	0.00357	0.00211	0.00240
6.0	0.00103	0.00160	0.00068	0.00062
7.0	0.00063	0.00105	0.00039	0.00086

Table 5: Cross Section limits using 2016 data and the deep boosted tagger for the decay to qW

Mass [TeV]	Exp. limit [pb]	Upper limit [pb]	Lower limit [pb]	Obs. limit [pb]
1.6	0.17750	0.25179	0.12572	0.38242
1.8	0.11125	0.15870	0.07826	0.11692
2.0	0.08188	0.11549	0.05799	0.09528
2.5	0.03328	0.04668	0.02373	0.03653
3.0	0.01648	0.02338	0.01181	0.01108
3.5	0.00840	0.01195	0.00593	0.00683
4.0	0.00459	0.00666	0.00322	0.00342
4.5	0.00276	0.00412	0.00190	0.00366
5.0	0.00177	0.00271	0.00118	0.00401
6.0	0.00110	0.00175	0.00071	0.00155
7.0	0.00065	0.00108	0.00041	0.00108

Table 6: Cross Section limits using 2016 data and the N-subjettiness tagger for the decay to qZ

Mass [TeV]	Exp. limit [pb]	Upper limit [pb]	Lower limit [pb]	Obs. limit [pb]
1.6	0.08687	0.12254	0.06174	0.06987
1.8	0.06719	0.09477	0.04832	0.03424
2.0	0.04734	0.06640	0.03405	0.03310
2.5	0.01867	0.02619	0.01343	0.03214

Mass [TeV]	Exp. limit [pb]	Upper limit [pb]	Lower limit [pb]	Obs. limit [pb]
3.0	0.01043	0.01463	0.00744	0.01773
3.5	0.00596	0.00840	0.00426	0.00347
4.0	0.00353	0.00500	0.00250	0.00140
4.5	0.00233	0.00335	0.00164	0.00181
5.0	0.00157	0.00231	0.00110	0.00188
6.0	0.00082	0.00126	0.00054	0.00049
7.0	0.00050	0.00083	0.00031	0.00066

Table 7: Cross Section limits using 2016 data and deep boosted tagger for the decay to qZ

Mass [TeV]	Exp. limit [pb]	Upper limit [pb]	Lower limit [pb]	Obs. limit [pb]
1.6	0.16687	0.23805	0.11699	0.35999
1.8	0.12750	0.17934	0.09138	0.12891
2.0	0.09062	0.12783	0.06474	0.09977
2.5	0.03391	0.04783	0.02422	0.03754
3.0	0.01781	0.02513	0.01277	0.01159
3.5	0.00949	0.01346	0.00678	0.00741
4.0	0.00494	0.00711	0.00349	0.00362
4.5	0.00293	0.00429	0.00203	0.00368
5.0	0.00188	0.00284	0.00127	0.00426
6.0	0.00102	0.00161	0.00066	0.00155
7.0	0.00053	0.00085	0.00034	0.00085

Table 8: Cross Section limits using the combined data and the N-subjettiness tagger for the decay to qW

Mass [TeV]	Exp. limit [pb]	Upper limit [pb]	Lower limit [pb]	Obs. limit [pb]
1.6	0.05703	0.07999	0.04088	0.03366
1.8	0.03953	0.05576	0.02833	0.04319
2.0	0.02844	0.03989	0.02045	0.04755
2.5	0.01270	0.01781	0.00913	0.01519
3.0	0.00658	0.00923	0.00473	0.01218
3.5	0.00376	0.00529	0.00269	0.00474
4.0	0.00218	0.00309	0.00156	0.00114
4.5	0.00132	0.00188	0.00094	0.00068
5.0	0.00084	0.00122	0.00060	0.00059
6.0	0.00044	0.00066	0.00030	0.00041

Mass [TeV]	Exp. limit [pb]	Upper limit [pb]	Lower limit [pb]	Obs. limit [pb]
7.0	0.00022	0.00036	0.00014	0.00043

Table 9: Cross Section limits using the combined data and the deep boosted tagger for the decay to qW

Mass [TeV]	Exp. limit [pb]	Upper limit [pb]	Lower limit [pb]	Obs. limit [pb]
1.6	0.06656	0.09495	0.04698	0.12374
1.8	0.04281	0.06141	0.03001	0.05422
2.0	0.03297	0.04650	0.02363	0.04658
2.5	0.01328	0.01868	0.00950	0.01109
3.0	0.00650	0.00917	0.00464	0.00502
3.5	0.00338	0.00479	0.00241	0.00408
4.0	0.00182	0.00261	0.00129	0.00127
4.5	0.00107	0.00156	0.00074	0.00123
5.0	0.00068	0.00102	0.00046	0.00149
6.0	0.00038	0.00060	0.00024	0.00034
7.0	0.00021	0.00035	0.00013	0.00046

Table 10: Cross Section limits using the combined data and the N-subjettiness tagger for the decay to qZ

Mass [TeV]	Exp. limit [pb]	Upper limit [pb]	Lower limit [pb]	Obs. limit [pb]
1.6	0.05125	0.07188	0.03667	0.02993
1.8	0.03547	0.04989	0.02551	0.03614
2.0	0.02523	0.03539	0.01815	0.04177
2.5	0.01059	0.01485	0.00761	0.01230
3.0	0.00576	0.00808	0.00412	0.01087
3.5	0.00327	0.00460	0.00234	0.00425
4.0	0.00190	0.00269	0.00136	0.00097
4.5	0.00119	0.00168	0.00084	0.00059
5.0	0.00077	0.00110	0.00054	0.00051
6.0	0.00039	0.00057	0.00026	0.00036
7.0	0.00019	0.00031	0.00013	0.00036

Table 11: Cross Section limits using the combined data and deep boosted tagger for the decay to qZ

Mass [TeV]	Exp. limit [pb]	Upper limit [pb]	Lower limit [pb]	Obs. limit [pb]
1.6	0.07719	0.10949	0.05467	0.14090
1.8	0.05297	0.07493	0.03752	0.06690
2.0	0.03875	0.05466	0.02768	0.05855
2.5	0.01512	0.02126	0.01080	0.01160
3.0	0.00773	0.01088	0.00554	0.00548
3.5	0.00400	0.00565	0.00285	0.00465
4.0	0.00211	0.00301	0.00149	0.00152
4.5	0.00118	0.00172	0.00082	0.00128
5.0	0.00073	0.00108	0.00050	0.00161
6.0	0.00039	0.00060	0.00025	0.00036
7.0	0.00021	0.00034	0.00013	0.00045