Karlsruhe Institute of Technology

Measurement of the t-channel single top quark production cross section in the electron and muon decay channel with the CMS detector at the LHC

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik des
Karlsruher Institut für Technologie genehmigte

DISSERTATION

von

Dipl.-Phys. Daniel Michael Martschei
aus Engelskirchen

Tag der mündlichen Prüfung: 06.07.2012

Referent:	Prof. Dr. M. Feindt,	Institut für Experimentelle Kernphysik
Korreferent:	Prof. Dr. Th. Müller,	Institut für Experimentelle Kernphysik

Für meinen Vater

Contents

1. Introduction

In 1995 the top quark was observed by the CDF and DØ experiments at the Tevatron [10], in events from proton-antiproton collisions. It is still the last newly discovered elementary particle. The top quark was predicted to exist in 1973 by Kobayashi and Maskawa [77] together with its lighter partner the bottom quark, which was discovered in 1977. But only the Tevatron collider could provide enough energy to produce the top quark and enough luminosity to detect it. This observation of a long predicted particle proved once again the correctness of the standard model of particle physics.

The first observed events contained pairs of top and anti-top quarks, produced via the strong interaction. Additionally the production of single top or anti-top quarks via the electroweak interaction were expected to be possible, as well. It was again at the Tevatron in 2009, when single top quark events were observed by the CDF [3] and the DØ [9] collaborations, yet the delivered amount of events at the Tevatron is not sufficient for deeper studies of this process.

When in 2010 the Large Hadron Collider (LHC) at CERN finally started running with a centre of mass energy of $\sqrt{s} = 7\,\text{TeV}$ and produced proton-proton collisions, it was not long until the top and anti-top quark pair production also was observed in the early data by the ATLAS [1] and CMS [73] collaborations. Also the electroweak production of single top quarks has been observed in early data by both experiments [35, 99]. This early success was possible due to the higher center of mass energy and the higher luminosity at the LHC than at the Tevatron.

The measurment of the single top cross section is especially of interest since the top quarks are produced via an electroweak process via a Wtb vertex. Therefore the cross section is proportional the CKM-matrix element $|V_{tb}|^2$. This provides a direct measurement of $|V_{tb}|$ in this process. Additionally the left-handed structure of the weak interaction can be studied in single top processes.

A precise measurement of the single top production cross section is presented in this thesis. The events utilised were collected by the CMS detector at the LHC in 2011. For a pure event selection and reconstruction, the semileptonic decay channel is used, where a muon or an electron are found in the final state. Several discriminating input variables are reconstructed in the candidate events and a neural network is trained in either channel to separate signal from background events. The dataset is split into several different categories which differ in their composition of different processes. In these categories discriminator distribution are derived and are used in the statistical inference to extract the signal yield and the residual uncertainty.

In the following, first an overview of the standard model of particle physics with a special focus on the top quark is given in chapter 2. In chapter 3, the LHC together with the CMS detector and all its components is presented, since a thorough understanding of the data taking process is crucial for a successful event selection and reconstruction.

The CMS detector measures the events produced by LHC proton-proton collisions. Additionally it is necessary to generate pure Monte Carlo samples for all physics processes that are expected to be present in the selected data, in order to study them separately and compare them to data. Therefore in chapter 4 the utilised event generators as well as the needed steps to correctly reconstruct all needed aspects of an event to use it for the analysis is presented.

Several different statistical methods and algorithms are needed to extract the signal yield from data. These methods and the software packages used are described in chapter 5.

In chapter 6 the measurement of the single top cross section is presented.

Finally, the results are discussed in chapter 7 and an outlook on possible improvements for future measurements of this cross section are given.

2. Theory

2.1. The standard model of particle physics

The standard model of particle physics (SM) is a quantum field theory which describes the elementary particles and the interactions among them. It will be briefly presented in the following.

2.1.1. Particles

The particles of the SM are :

- six quarks (up (u), down (d), charm (c), strange (c), top (t), bottom (b)) each carrying one of three colour charge types (blue, red and green) [63] and also electromagnetic charge of either $+\frac{2}{3}\,|e|$ ($u,\ c,\ t$) or $-\frac{1}{3}\,|e|$ ($d,\ s,\ b$), where e is the elementary charge

- six leptons (electron (e), muon (μ) and tau (τ) and one neutrino for each), which have negative electromagnetic charge except the neutrinos, which are electromagnetic neutral.

- four gauge bosons mediating forces (γ, gluon, Z^0, W^+ and W^-)

An overview of the particles and is given in figure 2.1.

Quarks and leptons are fermions which means they carry spin $\frac{1}{2}$ and have to obey Pauli's exclusion principle ("There must never be two fermions in the same quantum mechanical state"), whereas the force carrying particles are bosons, i.e. they have an integer spin of 1. For the fermions[1] there are additionally also an anti-particle for every one of them with the opposite charges. Shortly after the big bang there must have been the same amount of matter and antimatter since they were equally produced from energy. It is one of today's mysteries why we live in a "matter"-world where there is so much more matter than antimatter left.

Another yet unresolved question is why the standard model's fermions have mass. For the time being the most favoured theory to explain this is the Higgs mechanism, and a large effort is made to search for the Higgs boson. In the standard model the fermion masses are generated via the so called Yukawa coupling. Since the top quark is the heaviest known particle today, it must have a strong coupling to the Higgs boson, if it exists.

[1] It is not yet clear whether neutrinos are their own anti-particles (Majorana-Neutrino) or not.

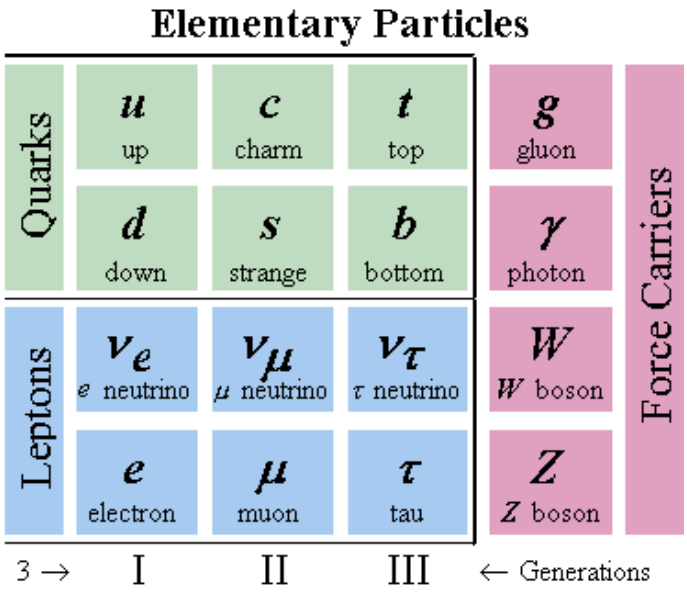


Figure 2.1.: The elementary particles of the standard model, taken from [102].

2.1.2. Interactions

Between the particles presented above there are the following three interactions[2]described in the SM. They are mediated via the bosons:

The electromagnetic interaction is probably the longest known standard model interaction. It is responsible for light, electron binding in atoms and the formation of molecules as well as all kinds of electrical and magnetic fields. In the standard model this interaction is described by the quantum electro dynamics (QED) [62], which is a Abelian gauge theory with $U(1)$ symmetry. The acceleration of protons in the LHC with electric fields, as well as keeping them on track with magnetic fields is both due to this force. But the more important effect is the interaction of high energetic particles with detector material. In fact, most of the interaction which lead to detectable signals in the detector are electromagnetic, as is described in 3.2.

The strong interaction is the dominant force during the collision particles like the proton. It is responsible for the parton interaction and the evolution of parton showers as well as for the hadronisation of the partons into mesons (pair of quark and antiquark) and baryons (triple of quarks). This force is mediated by gluons, which all carry colour and anticolour. With three different colour charges for the strong interaction, there are nine different gluons, an octet and a singlet. But only the octet gluons actually mediate the strong force, while the singlet is colour-neutral. An overview of these processes will be given in section 2.2.

[2] Even if gravity seems to be the most present force in everyday's life it can not be described in the standard model so far, but it is well described in the theory of general relativity.

The weak interaction is for example responsible for the decay of quarks into lighter ones and thus their limited lifetimes. It is the only force which can change the flavour of a quark and is therefore very important for the generation and the decay of quarks in LHC collisions. More details about this quark decay will be given in section 2.3.

An overview of the three standard model interactions is given in table 2.1.

Table 2.1.: Overview of the standard model interactions

interaction	mediator(s)	involved particles	specialities
electromagnetic	photon (γ)	quarks, e, μ, τ, $W^{\pm}$	photon is massless and has an infinite lifetime
weak	$W^{\pm}$, Z^0	$W^{\pm}$ couple only with left handed fermions and right handed antifermions	heavy mediators, $W^{\pm}$ maximal parity violation
strong	8 gluons	quarks and gluons	confinement, asymptotic freedom, gluons are massless

2.2. Hard scattering, hadronisation and parton showers

In a proton-proton collision with a centre of mass energy of $\sqrt{s} = 7\,\text{TeV}$, which was the default during the LHC data taking in 2011, it is not the proton as a whole which interacts with the other proton but their constituents, the partons. Parton is a collective term for particles inside a baryon or meson. It was proposed by Murray Gell-Mann [66] at a time, when it was not yet known out of which particles these hadrons consist, but it was already certain that they have some substructure.

Today it is known that hadrons are build up of three valence quarks. Gluons bind them together and lead to sea quarks inside a hadron. The sea quarks are quantum mechanical vacuum fluctuations which are generated from gluon splittings. Only the valence quarks can be considered as "real" particles in a sense, that their energy and momentum relation satisfies

$$E = \sqrt{(m_0 \cdot c^2)^2 + (c \cdot p)^2}, \tag{2.1}$$

where m_0 is the rest mass of the particle. Such particles are also called to be "on the mass shell", since equation (2.1) describes a three dimensional hyperboloid in the for dimensional energy-momentum phase space. All possible combinations of energy and momentum for a real particle are on this plane.

In contrast the sea quarks can only exist due to Heisenberg's uncertainty principle for energy and time

$$\Delta E \Delta t \geq \frac{\hbar}{2} \approx 3.291 \cdot 10^{-16}\,\text{eVs}. \tag{2.2}$$

It says that energy and the lifetime of a particle can only be known to a finite precision simultaneously. Hence a pair of heavy particles can be generated from vacuum, if the uncertainty on their lifetime remains large, which is especially the case for very short lifetimes.

2.2.1. Parton distribution functions

The total proton momentum is split up among its constituents. In order to calculate proton-proton collisions it is important to know with which probability a certain quark of flavour i will interact with another quark of flavour j and how large there momentum fractions x_i and x_j are. These probabilities are given by the *parton distribution functions* (PDF). These functions also depend on the momentum exchange $\mu = Q^2$ in the interaction, $f(x, Q^2)$.

They have been measured e.g. at the HERA experiments H1 and Zeus at DESY where electrons and protons were collided to measure the deep inelastic scattering electrons with partons inside the proton. The latest combined result of H1 and Zeus for the PDF of valence quarks xu_v, xd_v, as well as for gluons xg and sea partons xS is given in figure 2.2.

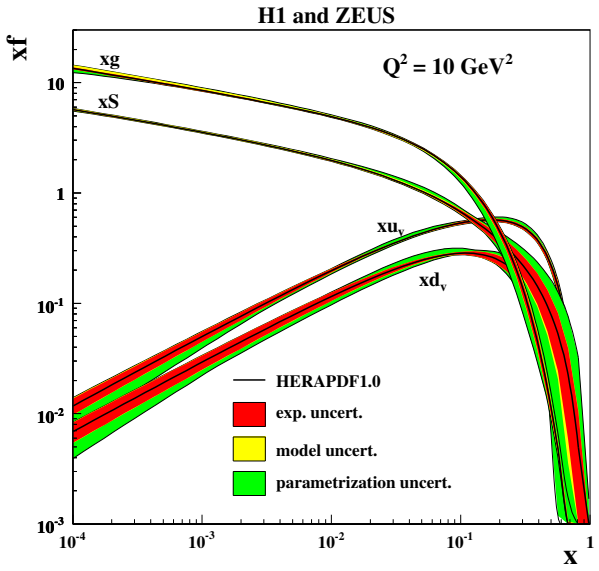


Figure 2.2.: The parton distribution functions from HERAPDF1.0, xu_v , xd_v , $xS = 2\text{x}(\text{U} + \text{D})$, xg at Q = 10 GeV. [5]

There are several sets of PDFs available to be used in Monte Carlo generators. The CTEQ2008 next-to next-to leading order PDFs for two different scale parameters Q^2 are shown figure 2.3.

During the collision of two partons energy and momentum are exchanged. They will be deflected and gain momentum transversal to the proton flight direction and will separate

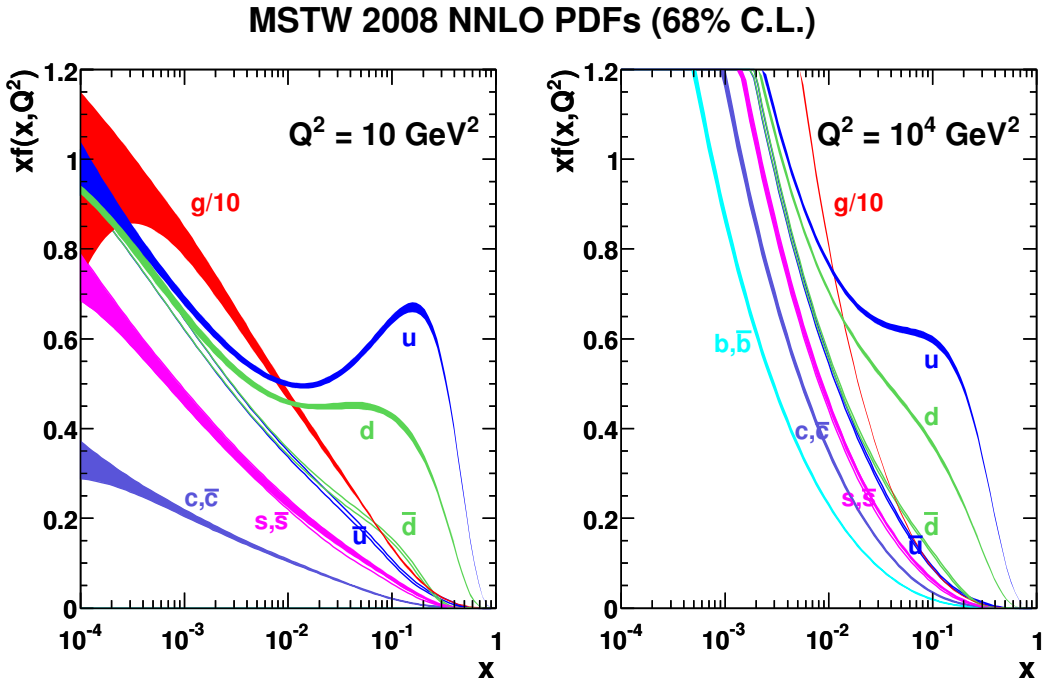


Figure 2.3.: MSTW Parton distribution functions for all quarks in protons relevant for the LHC.[83]

from the proton remnants. At this step *confinement* starts to play a role. Since gluons carry colour charge, they do interact with each other. If two quarks are separated, a gluon string is build between them. This string becomes stronger with larger distances, similar to a rubber band. At some point it becomes energetic favourable to form a new quark antiquark pair out of the energy stored in the gluon string than letting its field increase furthermore. This process will occur several times for each quark moving away from the collision point. It stops when a group of quarks, which has in sum a neutral colour charge (so called "white"), is bound together and moves in one direction and forming a stable particle. Because there are three different types of (anti)colour charge a (anti)quark can have, the simplest colour charge free objects are a pair of quark and antiquark (meson) and a triple of (anti)quarks (baryon).

These groups of hadrons, and possibly some leptons originating from weak quark decays, form a so called "*jet*", which is a collimated spray of particles. There are several jet reconstruction algorithms available in CMS reconstruction software, which are described in 4.2.4. Most of them will start identifying jets in the calorimeter entries and then make use of additional information from the tracking system.

2.3. The weak interaction and the decay of quarks

Mediators of the weak force are $W^{\pm}$ and Z^0 bosons which have relatively high masses ($m(W^{\pm}) = 80.398 \pm 0.025$ GeV and $m(Z) = 91.1876 \pm 0.0021$ GeV [69]). This is quite a difference between the weak force and the electromagnetic or the strong force because photons and gluons are massless. The W^+ and the W^- bosons only interact with left handed fermions and right handed anti-fermions. Therefore their coupling is maximally violating the P and C parity. Still the combined CP parity in most of the cases. The Z^0 boson interacts with left and right handed particles, but with different coupling strength.

Due to their heaviness the W and Z bosons have a very large decay phase space as they can decay into basically all leptons and quarks, except the top quark. Hence their lifetime is about 10^{-24}s, which is too short to bind quarks together.

On the other hand it is the only one which does not conserve fermion flavours and thus has the ability to change the flavour of a quark into another and is responsible for the decay of the μ and τ leptons. Within a quark flavour change the quark's electric charge has to change as well, i.e. it has to happen via a charged W boson. Decays between two up-type or down-type quarks like $c \to u$ or $b \to d$ which could in principle be mediated via a Z boson have not been observed at tree-level, since they are suppressed by the so called GIM mechanism. However they happen via higher order processes, like W box-diagrams, e.g. in $B^0 \leftrightarrow \overline{B^0}$ oscillation.

The most prominent example for a weak interaction is the β-decay of a radioactive nucleus which is the conversion of a down quark into an up quark or vice versa. This happens via the emission of a virtual W boson which decays into an electron or positron plus a neutrino. The Feynman graph of such a decay can be seen in figure 2.4.

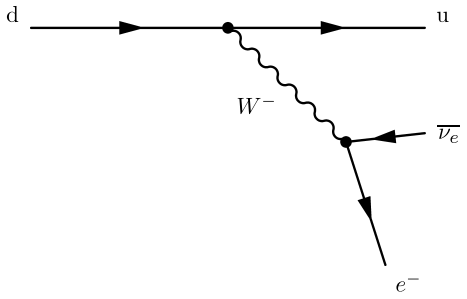


Figure 2.4.: Down quark decays into an up quark, an electron and an electron-antineutrino like in neutron β-decays

A weak decay across generations is only possible because weak-force eigenstates are not the same as mass eigenstates but they are superpositions as depicted in 2.3 where q' is a weak-force eigenstate while q is a strong force eigenstate. The mixing amplitudes corresponding to the strength of the flavour-changing weak decay are given by the Cabibbo-Kobayashi-Maskawa(CKM) matrix, which is shown in (2.4).

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.3}$$

The matrix elements $|V_{q,q'}|$ are being measured with more and more precision and have currently the following world average value [92]:

$$V_{CKM} = \begin{pmatrix} 0.97419 \pm 0.00022 & 0.2257 \pm 0.0010 & 0.00359 \pm 0.00016 \\ 0.2256 \pm 0.0010 & 0.97334 \pm 0.00023 & 0.0415^{+0.0010}_{-0.0011} \\ 0.00874^{+0.00026}_{-0.00037} & 0.0407 \pm 0.0010 & 0.999133^{+0.000044}_{-0.000043} \end{pmatrix} \tag{2.4}$$

The matrix is nearly diagonal which means that transitions between quarks of the same generation are more favoured than transitions between different generations. This explains the relatively long lifetime of the b quark. Since it is the lighter particle in the third generation, it has to decay into a charm or an up quark. But the matrix elements $|V_{cb}|$ and $|V_{ub}|$ are only 0.0415 and 0.00359 respectively. Thus the b-lifetime is way higher than one would expect it to be by just considering its mass.

On the other hand the matrix element $|V_{tb}|$ is close to one, while $|V_{ts}|$ and $|V_{td}|$ are rather small. In return this means that the top quark will decay in almost 100% of the cases into a b quark and very seldom into an s or d quark.

2.4. The top quark

The top quark is currently the heaviest known particle in the standard model. It was first discovered in 1995 by the CDF [10] and DØ [6] experiments at the Tevatron collider at Fermilab (Chicago, USA). In the following the measured and or inferred quantities of the top quark are listed.

mass $\mathrm{m} = (173.2 \pm 0.9)\,\mathrm{GeV}$ [103]

width $\Gamma = 1.99^{+0.69}_{-0.55}\,\mathrm{GeV}$ [8]

charge $+ 2/3\,|e|$

spin $\mathrm{I}(\mathrm{J}^{\mathrm{P}}) = 0(1/2^{+})$

relative branching ratio $\Gamma(Wb)/\Gamma(Wq(q = b, s, d)) = 0.99^{+0.09}_{-0.08}$ [69]

Since the mass of the top quark of $173.2 \pm 0.9\ \mathrm{GeV}/c^2$[103] is higher than the W boson's mass, it will decay almost immediately into a real W boson and a quark. The short lifetime of only about $10^{-25}\,\mathrm{s}$ leads to a relatively high width of about 2.0 GeV, which is also due to Heisenberg energy-time uncertainty principle as given in equation (2.2). The t quark is especially not able to form bound states with other quarks via QCD interaction and hence there are no known mesons or baryons containing t quarks.

The relative branching fraction of decays into $W + b$ is $0.99^{+0.09}_{-0.08}$ [69], which leads to a CKM-matrix element $|V_{tb}|$ of essentially 1 as can be seen in 2.4.

In the following the dominating generation processes for single top quarks in LHC collisions will be described, as well as its semileptonic decay.

2.4.1. Top quark pair production at the LHC

With a centre of mass energy of $\sqrt{s} = 7\,\text{TeV}$, many of the events generated in LHC collisions contain a pair of top quarks. The dominant leading order production modes are depicted in figure 2.5. All of these processes involve gluons but no W or Z bosons and are therefore mediated via the strong interaction. Generating $t\bar{t}$ events via the weak interaction is possible as well, but the cross section of these processes is negligible compared to the strong interaction [79].

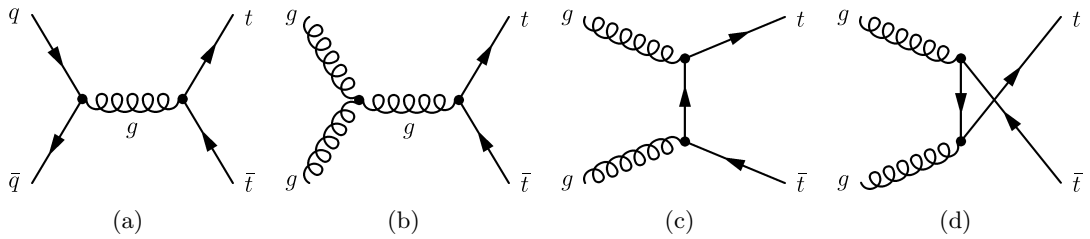


Figure 2.5.: Leading order Feynman diagrams for the production of $t\bar{t}$ pairs via $q\bar{q}$ annihilation (a) and gluon-gluon fusion (b), (c) and (d). At the LHC, the gluon-gluon fusion processes contribute $\approx 80\%$ to the total cross section for the $t\bar{t}$ production.

Comparing the production processes of top quark pairs at the Tevatron and the LHC one finds an important difference in the relative cross sections for quark-antiquark annihilation and gluon fusion processes. In proton-proton collisions at the LHC, there are no antiquarks as permanent constituents in the colliding particles, but only as sea quarks. Therefore the cross-section for annihilation processes is rather low, compared to the Tevatron, were protons and anti-protons are collided, and the gluon-gluon fusion processes play a more important role in LHC collisions. Yet the overall cross sections for producing t and $\bar{t}$ quarks at the LHC are higher than at the Tevatron due to the increased centre of mass energy.

Theory predictions from [76] at a next to next to leading order (NNLO) approximation yield for the Tevatron at $\sqrt{s} = 1.96\,\text{TeV}$ and a top quark mass of $m_t = 173\,\text{GeV}$ a $t\bar{t}$ cross section of

$$\sigma_{t\bar{t}} = 7.08^{+0.00\,+0.36}_{-0.24\,-0.27}\,\text{pb}$$

where the first error is from scale variations and the second from PDF uncertainties.

The predicted cross section for LHC collision at $\sqrt{s} = 7\,\text{TeV}$ is

$$\sigma_{t\bar{t}} = 163^{+7\,+9}_{-5\,-9}\,\text{pb}$$

and therefore more than 23 time higher than at the Tevatron (uncertainties have the same sources as before).

2.4.2. Single top quark production at the LHC

The production of single top quarks is a pure electroweak process which involves a Wtb vertex. It is therefore of special interest since the cross sections σ_t are directly proportional to measure the CKM-matrix element $|V_{tb}|^2$.

There are three different processes for the production of a single top quarks in proton-proton collisions at the LHC. They are distinguished by the virtuality of the involved W boson.

s-channel ($q^2 > M_W^2$) The W boson is time-like virtual

t-channel ($q^2 < M_W^2$) The W boson is space-like virtual

associated production ($q^2 = M_W^2$) also called "tW-channel". The W boson is not virtual but real (on shell). It therefore contributes to the final state particles and will produce detectable particles. This leads to a different event signature, than in the s- and t-channel.

In figure 2.6 the leading order (LO) Feynman diagrams for the single top quark production are shown. Due to the flavour conservation in the strong interaction, only the

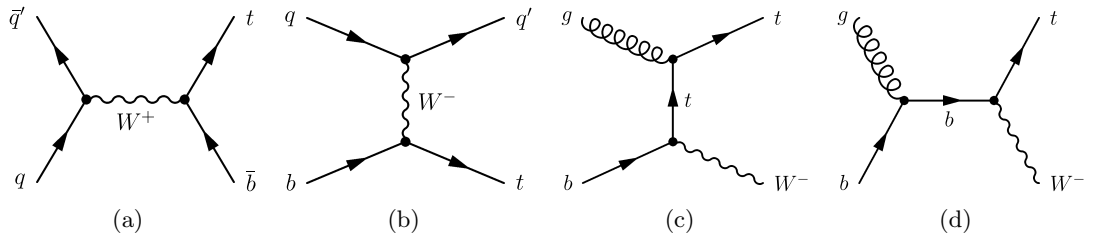


Figure 2.6.: LO Feynman diagrams for the electroweak production of single top quarks via the s-channel (a), the t-channel (b), and associated tW production (c) and (d).

weak interaction can produce single top quarks by transforming an initial b quark into a top quark. The t-channel and the associated tW production have in common, that an initial b sea quark is transformed into a t quark via W boson emission. In the t-channel the W boson is exchanged with another quark, while in the associated production the W boson is emitted.

Each of these three different processes will be explained in more detail in the following.

s-channel single top quark production

Besides the LO process in figure 2.6(a) there are NLO single top generating processes which involve a time-like virtual W boson. Three of them are depicted in figures 2.7(a), 2.7(b) and 2.7(c).

t-channel single top quark production

Additionally to the s-channel, with the time-like W exchange, there is the t-channel with the space-like W boson exchange. Besides the LO process which is shown in figure 2.6(b) there are NLO processes which play an important role and need to be taken into account. The according Feynman diagrams for the t-channel are shown in figure 2.8.

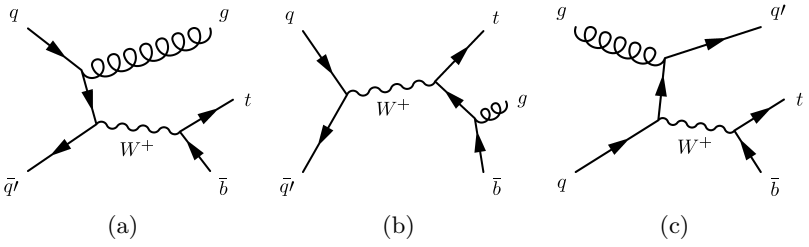


Figure 2.7.: NLO Feynman diagrams for the s-channel production of single top quarks via initial and final state gluon radiation (a), (b) and initial gluon splitting (c).

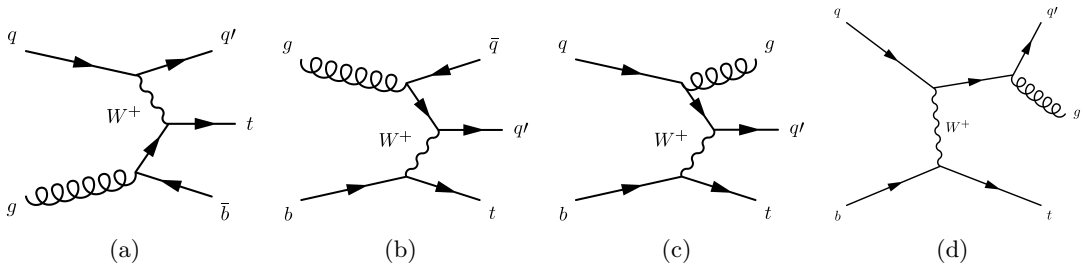


Figure 2.8.: NLO Feynman diagrams for the t-channel production of single top quarks via initial state gluon splitting (a) (b) and via gluon radiaton in the initial (c) and final (d) state.

associated production tW production

In the associated production, the involved W boson is emitted as a real particle in the single top production. Three NLO production Feynman graphs for this process is shown in figure 2.9.

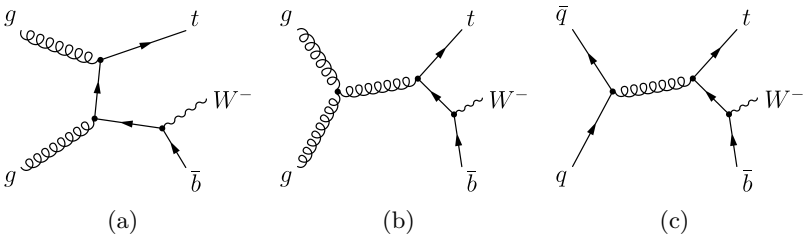


Figure 2.9.: NLO Feynman diagrams for the associated production of single top quarks and a W boson via gluon fusion (a), (b) and $q\bar{q}$ fusion (c).

Expected cross section for single top quark events at hadron colliders

The expected cross section for these processes are quite different, and they also differ from what was measured at the Tevatron due to the different colliding particles, as was

explained before in 2.4.1. The individual theoretically predicted cross-sections are listed in table 2.2 for the LHC and, for comparison also, for the Tevatron.

Table 2.2.: Theoretically predicted single top quark production cross sections for the LHC and the Tevatron as published in [76]

	LHC at 7 TeV	Tevatron at 1.96 TeV
s-channel	$\sigma_t = 3.19 \pm 0.06^{+0.13}_{-0.10}$ pb $\sigma_{\bar{t}} = 1.44 \pm 0.01^{+0.06}_{-0.07}$ pb	$0.523^{+0.001\,+0.030}_{-0.005\,-0.028}$ pb
t-channel	$\sigma_t = 41.92^{+1.59}_{-0.21} \pm 0.83$ pb $\sigma_{\bar{t}} = 22.65 \pm 0.50^{+0.68}_{-0.91}$ pb	$1.04^{+0.00}_{-0.02} \pm 0.06$ pb
associated production	$7.87 \pm 0.20^{+0.57}_{-0.55}$ pb	negligible

As can be seen, the predicted cross sections for all single top processes are much higher at the LHC than they were at the Tevatron which is due to the higher centre of mass energy of $\sqrt{s} = 7\,\text{TeV}$ compared to $\sqrt{s} = 1.96\,\text{TeV}$. Also the cross section for top and anti-top production are the same at the Tevatron while they are different by approximately a factor of two at the LHC. This is explained by the different content of the colliding particles. Since the Tevatron collided protons with anti-protons, the probability for quarks and antiquarks in the initial state are the same, while at the LHC protons are collided with protons. There it is less likely to have anti quarks in the initial state, since they only exist as sea quarks in the proton. Now considering the LO t-channel Feynman graph in figure 2.6(b), the chance that a u and b quark are converted into a b and t quark is double as high than the process $d + \bar{b} \to u + \bar{t}$, because there are two u quarks and one d quark a proton, while the chances to find a b or $\bar{b}$ sea quark are equal.

The latest measurement from CDF yields a cross section for the single top quark cross section in the s- and t-channel combined of $3.04^{+0.57}_{-0.53}\,\text{pb}$ [31], while the DØ collaboration measured this quantity to be $3.43^{+0.73}_{-0.74}\,\text{pb}$ [7]. Both measurement are in good agreement with the predicted values quoted above.

2.4.3. Top quark decay modes

The most recent combined measurement for the top quark mass of both Tevatron experiments CDF and DØ is $173.2 \pm 0.9\ \text{GeV}/c^2$[103], which is much larger than the W boson mass of $(80.399 \pm 0.023)\,\text{GeV}$ [69]. As was already discussed, a top quark can always decay via the weak interaction into a real W boson and a b quark and therefore has a very short lifetime of only about $10^{-25}\,\text{s}$. The Feynman graph for the top quark, together with the subsequent decay of the W boson is shown in figure 2.10.

It has been measured that in $(67.60 \pm 0.27)\,\%$ [69] of the cases, the W boson decays hadronically while the branching fractions for the leptonical decays are around 10 % for each lepton type, as can be seen in table 2.3

Table 2.3.: W-boson branching fractions as published in [69]

mode	branching fraction (Γ/Γ_i)
$e^+\nu$	$(10.75 \pm 0.13) \times 10^{-2}$
$\mu^+\nu$	$(10.57 \pm 0.15) \times 10^{-2}$
$\tau^+\nu$	$(11.25 \pm 0.20) \times 10^{-2}$
hadrons	$(67.60 \pm 0.27) \times 10^{-2}$

The analysis in chapter 6 focuses on the leptonic decay modes into a muon or an electron, since they produce event topologies which can be detected and reconstructed more clear in measured CMS data, as will be explained in section 4.5.

2.4.4. Semileptonic top quark decays

The semileptonic decay of a top quark actually means that the W boson decays leptonically into a lepton and a corresponding neutrino. This leads to some special features in the final state particles of the single top quark production and decay, which are exploited during the event selection and the reconstruction of top quark candidates, as will be explained in section 4.5.

Another important feature of the single top quark production and decay chain is the polarisation which leads to a certain angular distribution of the decay products. This angular distribution can be used for an advanced event selection.

Spin polarisation in single top quark decays

As stated above, the top quark has a negligible lifetime and will decay immediately after its generation. Since this time span is too short for QCD interactions, which would randomise its spin, it will decay with the same spin it was produced with and will pass its spin information to its decay products. This together with the standard model $V - A$ coupling of the W boson with the top quark leads to a strong angular correlation among

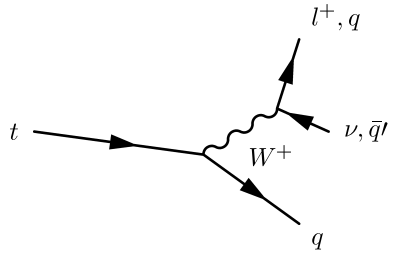


Figure 2.10.: The top quark decays into a b quark and a real W boson, which then decays leptonically or hadronically.

the top quark decay products. The decay angle distribution is given in equation (2.5) [81].

$$\frac{1}{\sigma_T}\frac{d\sigma}{d\cos\Theta} = \frac{1}{2}\left[1 + \frac{N_\uparrow - N_\downarrow}{N_\uparrow + N_\downarrow}\cos\Theta\right]. \tag{2.5}$$

In this differential equation Θ is the angle between the charged lepton, coming from the top quark decay, and the down type quark from the top quark production, while $N_\uparrow$ and $N_\downarrow$ are the number of up and down polarised top quarks.

In order to exploit this angular dependence for the event selection, some choices concerning the rest frame and the definition of axes in the reconstructed events have to be made. These details will be explained in section 4.5.1.

3. The Large Hadron Collider and the Compact Muon Solenoid

The Large Hadron Collider (LHC) is a collider hosted and operated at CERN[1] near Geneva (Switzerland). It is one of the biggest and most complex machines ever built by mankind. During the most time of the year protons are collided, but in some weeks heavy ions (lead) are accelerated for collisions.

There are six experiments installed around the ring at the for collision points. They are dedicated to the precise measurement of the collision events.

ATLAS: **A** **T**oroidal **L**HC Apparatu**S**[38]
A general purpose detector. One of its aims is to search for the Higgs boson. It is also searching for physics beyond the standard model.

CMS: **C**ompact **M**uon **S**olenoid[40]
Also a general purpose detector but with a different configuration than ATLAS. Both experiments have the same physiscs program and they shall mutually cross-check the results they retrieve.

ALICE: **A** **L**arge **I**on **C**ollider **E**xperiment[4]
This detector is specialised on measuring the fragments of heavy ion collisions. Its main data taking period is one month per year (usually november), when the LHC is operated with lead ions instead of protons.

LHCb: **L**arge **H**adron **C**ollider **b**eauty[53]
This detector focuses on b and $\bar{b}$ events. It is the only detector of these four which does not cover the whole spatial range. Instead it has a a forward geometry and is thus specialised on those jets which have a high boost. One of the main physics objectives of this experiment is to find out, why today the universe is dominated by matter and why the antimatter vanished after the big bang.

LHCf: **L**arge **H**adron **C**ollider **f**orward experiment[54]
It is uses forward particles created inside the LHC as a source to simulate cosmic rays in laboratory conditions.

TOTEM: **TOT**al **E**lastic and diffractive cross section **M**easurement[55]
This experiment studies forward particles created at LHC collisions in the CMS interaction point.

They four largest experiments are shown in the figure 3.1.

[1] Conseil Européen pour la Recherche Nucléaire, or European Council for Nuclear Research

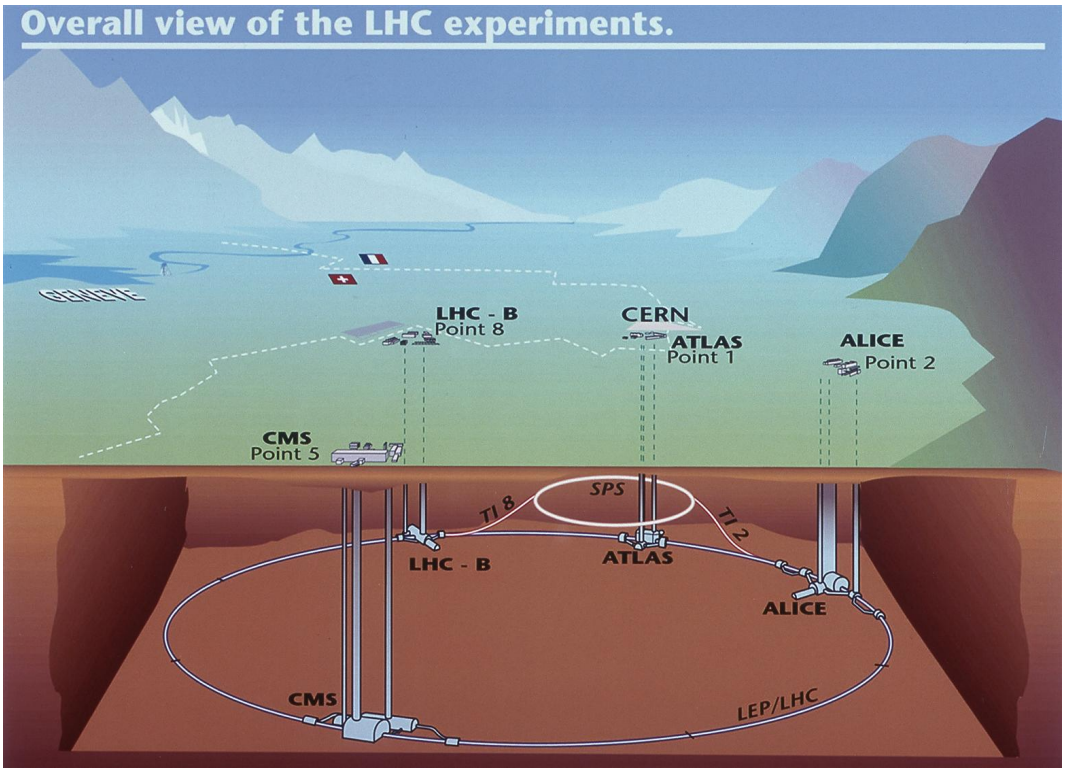


Figure 3.1.: The four large-scale LHC experiments and the accelerator tunnel underground[28].

3.1. The Large Hadron Collider

The main part of the Large Hadron Collider (LHC) are the 1232 superconducting dipole magnets, which are installed in a ring tunnel with a circumference of 26 659 m. The tunnel was built for the LEP (Large Electron-Positron Collider) accelerator, which was operated at CERN until 2000.

The dipole magnets keep the protons on track in the LHC beam pipes. They need to produce a field of up to 8.36 Tesla for 7 TeV beams. To achieve such a high dipole field, the LHC uses superconducting magnets which are cooled with superfluid, liquid helium to a temperature of 1.9 K. Additionally a few quadrupole and sextupole magnets for beam focussing and orbital corrections are installed as well at several points around the accelerator ring.

For the time being, this beam energy has not yet been reached. The sudden loss of superconductivity ("quenching") in one of the connections between two magnets lead to an incident in 2008, resulting in several damaged magnets which needed to be replaced. To reduce the risk of operation, the beam energy has been decreased for the 2010 and 2011 data taking period to $\sqrt{s} = 7\,\text{TeV}$, i.e. each beam had an energy of 3.5 TeV.

The protons are obtained by removing the electrons from hydrogen atoms and accelerating them in four steps (LINAC2, BOOSTER, PS, SPS) to 450 GeV before they are

injected into the LHC ring where they reach the projected energy of up to 7 TeV within 20 minutes. The accelerator chain can be seen in figure 3.2.

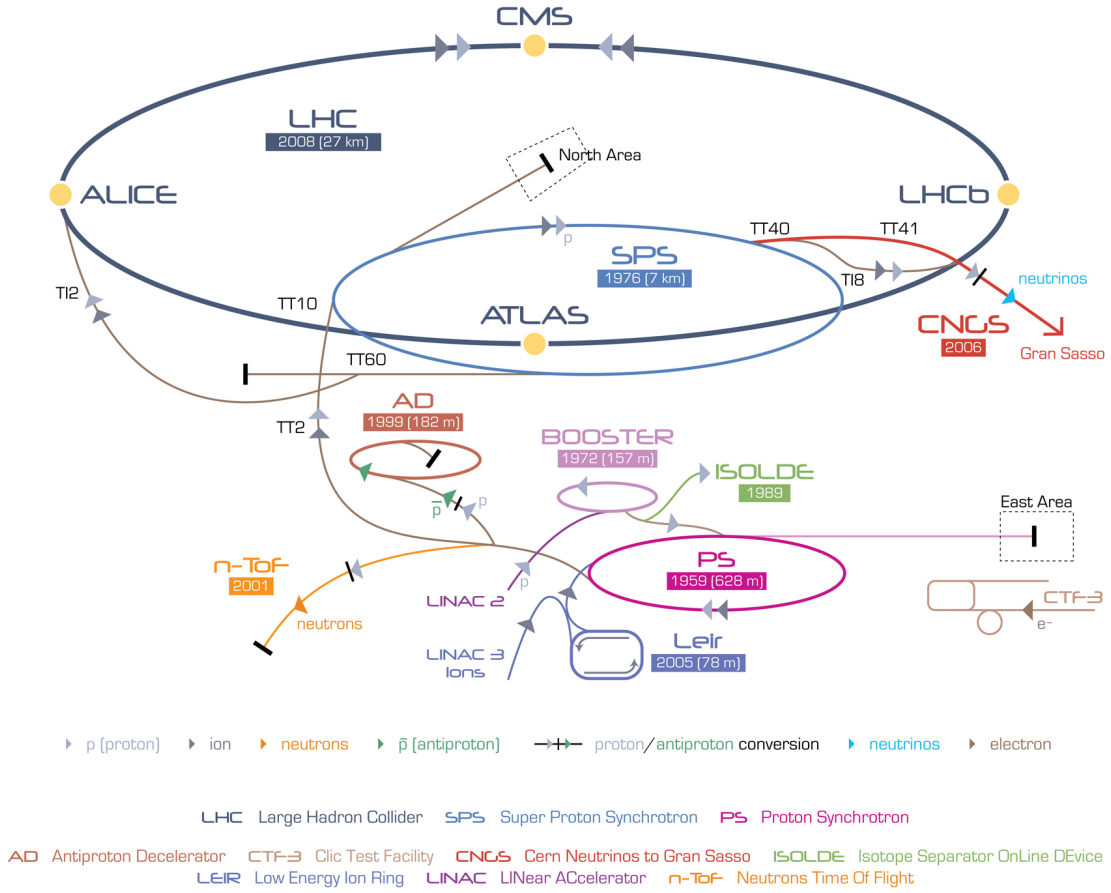


Figure 3.2.: CERN accelerator complex[33]. The protons for the LHC are started in the LINAC2 and get accelerated to 30 % c. The next step is the BOOSTER which puts them at an energy of 450 GeV. Other accelerator facilities are shown as well, which provide different particles for other experiments at CERN or remote, like CNGS (CERN Neutrinos to Gran Sasso).

At four intersection points, the two beams are collided. Each of these collision points is surrounded by one of the four big LHC experiments to measure the particles generated at these collisions.

3.1.1. Luminosity

In collider experiments the probability for a certain process to happen is usually given as an (imaginary) cross section, which is in analogy to classical mechanical collisions between two hard objects. The unit for these cross sections σ is a "barn", which is defined as $1\,\mathrm{b} = 10^{-28}\,\mathrm{m}^2$.

The instantaneous luminosity L denotes the number of events per barn and second. The event rate for a certain process can then be calculated as

$$\dot{N}_{exp} = \frac{\partial N}{\partial t} = L[b^{-1}s^{-1}] \cdot \sigma_{prod}[b]. \tag{3.1}$$

In this sense it can be understood as a stream of events with current intensity L, hitting an imaginary area, where the different process types are randomly distributed according to their probability, i.e. their cross section.

The instantaneous luminosity only depends on beam parameters and for Gaussian distributed beam particles inside a beam bunch it is given by

$$L = \frac{N_a N_b f_{\mathrm{rev}} \gamma_{\mathrm{rel}}}{4\pi \epsilon_{\mathrm{norm}} \beta^*} F. \tag{3.2}$$

N_a and N_b are the number of protons in either beam, f_{rev} the revolution frequency and γ_{rel} the relativistic gamma factor. β^* and ϵ_{norm} denote the beta function and the normalised beam emittance. They are both tunable machine parameters and need to be adjusted by the machine operators to maximise the luminosity.

The geometric luminosity reduction factor F accounts for the fact that both beams are crossing each other with a finite angle θ_c.

$$F = \left(1 + \left(\frac{\sigma_z \theta_c}{2\sigma^*}\right)\right) \tag{3.3}$$

Is is assumed here beams are symmetric, i.e. there transvers quadratic mean size is σ^* in x and y direction, and that the longitudinal quadratic mean beam size σ_z is much smaller than β^*.

Since the number of protons in the beam decreases during a run, and because the beam parameters may be adjusted, the instantaneous is not constant over time but needs to be continuously monitored. In CMS, this is done by using four rings in the forward hadron calorimeter [34, 52]. Their occupancy is linearly correlated to the luminosity.

Still this only allows a relative measurement and the absolute normalisation needs to be retrieved, which is done with van der Meer scans[104]. These scans are done in separate fills of the LHC. One beam is vertically displaced with respect to the other one and the number of interactions is monitored in dependence of the displacement.

A measure for the total number of data taken is the (time) integrated luminosity

$$\mathcal{L} = \int L \, dt. \tag{3.4}$$

The effective integrated luminosity for a data sample also depends on the used triggers, as will be discussed in section 6.2.1.

3.2. The CMS experiment

As a general purpose detector at the LHC, CMS' main task is the precise measurement of all particles produced in the proton-proton collisions. These collisions are the most

energetic ones ever made in a particle collider so far, and explore an unprecedented energy scale, the TeV scale.

At this energy scale and with its high luminosity of up to $L = 2 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, the LHC is a W and Z boson as well as a bottom and top quark factory. It will also produce the Higgs boson, SUSY sparticles and yet unknown types of other particles, if they exist on the TeV scale.

The CMS detector is designed to measure a wide range of particles produced in the decay of the particles mentioned before. The data taken are used to perform precise measurements of standard model parameters as well as to search for evidence of physics beyond the standard model. It is also expected that the data taken at the LHC will give an explanation for the electroweak symmetry breaking mechanism. In the following the overall properties of the detector are introduced.

The acronym CMS stands for "Compact Muon Solenoid", which already describes its main properties. It is, compared to ATLAS, relatively compact with a length of only 21.5 m and a overall diameter of 15.0 m while ATLAS is 46.0 m long and measures 25.0 m in diameter. The compactness of CMS was not the design priority itself but to have the tracker, the Electromagnetic Calorimeter (ECAL) and if possible also some parts of the Hadron Calorimeter (HCAL) inside of the solenoid. This was desired for the following reasons: If the calorimeters were outside of the solenoid, the electromagnetic particles would have had to traverse a lot of very dense material (the solenoid) before reaching the calorimeter. Thus they would have already produced a broad shower and would have deposited some of their energy inside of the solenoid. This would have led to two negative effects. The energy resolution would have been worse and the superconducting solenoid would have got warmed up by the particles energy, making it more difficult to cool it below its superconducting transition temperature.

But a larger solenoid in which the tracker together with the ECAL and the HCAL fit in easily is quite expensive and it would be difficult to control the magnetic field it produces. To achieve the goal of having the calorimeters inside of the solenoid and still have an affordable detector made it necassary to make it all very compact.

However the CMS detector is much heavier than the ATLAS detector. This weighs only 7000 metric tons while CMS is with 12500 metric tons nearly double as heavy. The enormous weight of the CMS detector is mostly due to the iron return yoke, which are the red parts in the expanded view of the CMS dectector in figure 3.3 and the segment view in figure 3.4.

The detector is built up in the classic cylindrical principle, with different layers of material and detection equipment. Their task is to identify the different kind of particles, which are produced in the collision and precisely measure their properties like energy, momentum and charge. The large muon chambers (see section 3.2.6), which are interleaved with the iron return yoke allow for a clean muon identification. In combination with the silicon-based tracking system it also yields a high momentum resolution for muons, hence the experiment's name.

In the following the CMC coordinate system will be introduced. Afterwards the different CMS detector layers will be presented going from the centre to the outside. For more detailed information, please refer to the Technical Design Report of CMS [39].

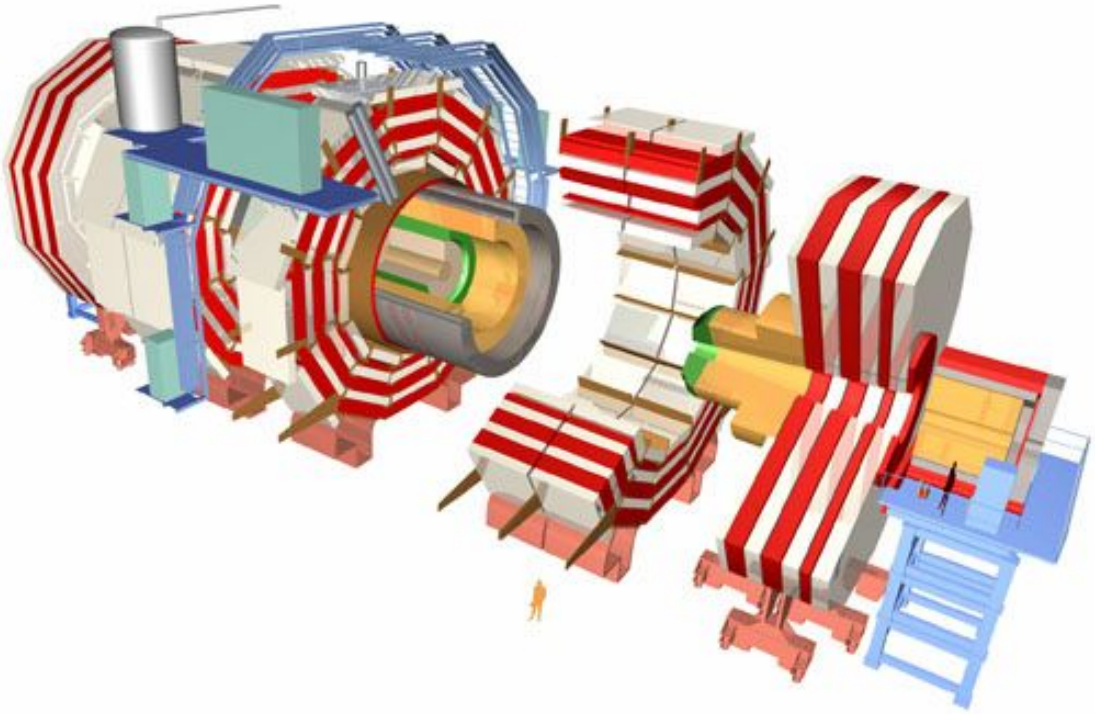


Figure 3.3.: Expanded view of the CMS detector

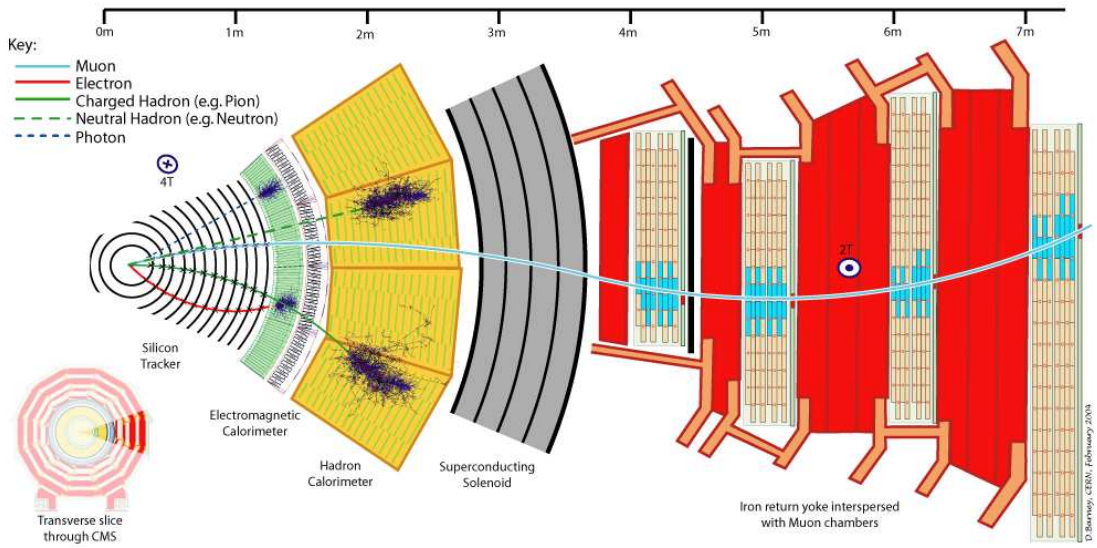


Figure 3.4.: Slice of the CMS detector

3.2.1. Coordinate conventions

The CMS coordinate systems origin is the nominal center of the CMS detector. The direction of the anti clock-wise circulating beam defines the positive z direction, while x and y span perpendicular plane in the right handed coordinate system. The positive y axis points towards the surface. Within the transversal x, y plane, the azimuthal angle φ is measured to the positive x axis. The polar angle θ is defined with respect to the positive z direction.

Besides the Cartesian and polar coordinates, it is useful to define two additional quantities to express the direction of a physics object momenum within the detector. These are the rapidity y and the pseudorapidity η.

Rapidity

The rapidity is defined as

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \tag{3.5}$$

It is additative for Lorentz transormations along the beam axis. Also it has the nice feature, that the particle flux per rapidity interval is constant for hadron-hadron colliders.

Pseudorapidity

The pseudorapidity is defined as

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right). \tag{3.6}$$

For massless particles or in the high energy limit, where $E \approx p$, the pseudorapidity and the rapidity are the same. The pseudorapidity has the advantage in this case, that only the polar angle θ is needed, but no information about the particles rest mass, and therefore also not about the particle type. Hence it is used quite often to specify a track or jet direction, together with the transversal momentum p_T.

3.2.2. Silicon tracker

A reliable and precise track measurement with a high spatial resolution is crucial for the reconstruction of an event. Without precisely reconstructed tracks it is not possible to measure the particles' momenta. Also finding the decay point of long-lived particles into several daughter particles, e.g. the decay of B hadrons, relies on a precise track reconstruction.

Because charged particles lose energy and get deflected when traversing material, the tracker has to be the innermost detector layer. Otherwise the initial particle's momentum would be altered by the interaction with material before reaching the detector. In previous detectors like DELPHI or CDF, drift chambers were used in addition to silicon detectors. However drift chambers need typically in the order of 100 ns until the charged particles have drifted to cathodes and anodes. If new particles cross the drift chamber during this time the pileup of events would lead to a mix-up of signals. Thus this technique is too slow to be used with the high collision rate of the LHC of 40 MHz, which corresponds to

a bunch spacing of only 25 ns. Instead silicon detectors are used for the CMS tracker. They make use of the fact that whenever a charged particle traverses such a detector it ionises the silicon which is equivalent to a deposition of charge. This charge is then amplified and can be measured as a current.

In order to measure the exact track of each particle as precisely as possible and still keep the number of readout channels low, the CMS tracking system is split up into three layers with different granularity.

- Closest to the interaction point at radii of 4 cm, 7 cm and 11 cm there are cylindrical layers of pixel detectors and disks at both ends. Each pixel has the size of about $(100 \times 150)\,\mu\text{m}^2$. Such small pixels are needed due to the high particle flux. Furthermore one needs this high spatial resolution to distinguish the different tracks from another.

- In the intermediate region ($20\,\text{cm} < r < 55\,\text{cm}$) silicon microstrip detectors are used. Their cell size is $10\,\text{cm} \times 80\,\mu\text{m}$. Their advantages in comparison to pixel detectors are: they are cheaper and they have less read out channels. This means in return less read out electronics which deflect radiation. The strip direction of sequent layers are tilted, which gives again a three dimensional space resolution by combining two subsequent layers. Since the particle flux is lower at this distance to the beam pipe, the advantages outbalance the disadvantages.

- At radii $55\,\text{cm} < r < 110\,\text{cm}$, there are silicon microstrips with a larger pitch and a maximum cell size of $25\,\text{cm} \times 180\,\mu\text{m}$.

The different tracker elements are arranged in a way, that the overall place and momentum resolution of a track is better than the single resolution of a pixel or a strip. This means the different layers are shifted against each other and are not radially and parallel aligned. The combined spatial resolution of the pixel tracking is measured to be about $10\,\mu\text{m}$ for the $r - \phi$ measurement and about $20\,\mu\text{m}$ for z measurement.

However a silicon vertex detector has one major disadvantage against drift chambers. Because the tracker material has a much higher density, the particles get multiply scattered and, especially electrons, radiate already inside of the tracker, producing photons and thus losing energy. These photons either reach the ECAL and produce an energy hit without an associated track, or they can convert into electron-positron pairs which will lead to two additional tracks and subsequently also to ECAL hits [19].

An important requirement for precise track reconstruction is a good described of the detector alignment. The actual position of a certain detector cell is only known to some very finite precision after the detector assembling. But especially for the very small pixel detectors the position and angles have to be known very precisely to benefit from their high resolution, otherwise merging hits of a particle in subsequent elements would not be possible. The alignment of CMS tracker has been performed using cosmic muons and data events from early runs [85].

3.2.3. Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) is designed to measure the energy of electrons, positrons or photons. Charged hadrons and muons deposit only a small part of

their energy, but are not fully stopped. The energy measurement is based on the fact, that electromagnetic showers evolve through bremsstrahlung and electron-positron pair production when electromagnetic particles traverse the detector. Since the calorimeter consists of a scintillating material, the showering induces photons and the number of emitted photons is a direct measure for the energy of the original particle. The scintillator gets activated via Compton scattering and photo effect, and emits light which is measured by photo detectors and transformed into a signal.

The ECAL contains 61200 lead tungstate ($PbWO_4$) crystals in the central barrel part and 7324 crystals in each of the 2 end caps. The main advantages of lead tungsten crystal as electromagnetic calorimeters are the short radiation length ($X_0 = 0.89\,\text{cm}$) and the small Molière radius (2.2 cm), the fast emission of light (80 % is emitted within 25 ns) and the radiation hardness (up to 10 Mrad). Because of the relative low light yield (30 γ / MeV) photo detectors with intrinsic gain are needed which can also be operated within high magnetic fields.

For the barrel silicon avalanche photo diodes are used, since they are insensitive to the high axial magnetic field. For the end caps vacuum photo triodes are used because they withstand the high radiation in this region. In front of the end caps there is an additional pre-shower detector installed, which is made of lead absorbers and silicon detectors. It separates high energetic single photons from photon pairs originating from π^0-decays.

The crystals used for the barrel part have a front face cross-section of $\approx 22 \times 22\,\text{mm}^2$ and are 230 mm (= 25.8 X_0) long. In the end caps the crystals are a little bit larger. They have a front-face cross-section of $28.6 \times 28.6\,\text{mm}^2$ and a length of 220 mm.

The spatial resolution of the crystals also allows a measurement of the shower shapes. This is helpful for the identification of the particle which produced the shower since electron-showers have a significantly different shape than those of muons or hadrons.

The intrinsic energy resolution of the ECAL is below 0.5 % for energy deposits above 120 GeV. Since many neighbouring crystals can be involved for measuring high energetic electrons or photons, the inter-calibration of the different channels of the ECAL has to be even better than 0.5 %. Otherwise the overall measurement uncertainties would be dominated by this effect.

3.2.4. Hadron Calorimeter

Hadrons, like protons, neutrons and pions, pass the electromagnetic calorimeter without much energy loss which makes them quite different from electrons and photons which hardly reach the end of the ECAL. This is due to the fact that hadrons do not interact with electrons but only with the absorber's nuclei. Hence a second calorimeter has to be installed just after the ECAL which is dedicated to hadrons. The hadron calorimeter (HCAL) has to be made of even more dense material than the ECAL to give the hadrons enough nuclei to interact and distribute their energy to showers.

The CMS HCAL uses brass as absorber which is easy to fabricate and non-magnetic. It is arranged in plates with a thickness of about 5 cm interleaved with scintillators.

There are still be some non-interacting particles, like muons and neutrinos or some yet unknown particles, which will not deposit energy in the ECAL or HCAL, as will also be explained in 4.5. Since the sum of the energy and momentum transversal to the beam has to be zero, such a particle might still be measured indirectly as missing energy and

momentum. This requires the HCAL to cover the whole solid angle (η, ϕ) and of course provide spatial resolution. The η-coverage is limited because of the beam-pipe which has to traverse the HCAL. To still get the best possible spatial coverage, the CMS HCAL is divided into four subsystems. These are the hadron barrel (HB), hadron outer (HO), hadron endcap (HE) and hadron forward (HF). Together they cover the pseudorapidites range $-5 < \eta < 5$. This means, only a small angle of 2 degrees to the beam pipe remains uncovered.

Compared to the ECAL, the HCAL has a much worse resolution. This makes it necessary to calibrate the HCAL on measured events to retrieve the Jet Energy Correction (JEC) information, which will be explained in 4.3.1.

3.2.5. Superconducting solenoid

Due to the Lorentz force, the track of a charged particle with charge q is bent when it moves in a magnetic field. The bending direction is orthogonal to the plane spanned by the magnetic field-vector $\vec{B}$ and the particle's velocity vector $\vec{v}$,

$$\vec{F} = q\left(\vec{v} \times \vec{B}\right). \tag{3.7}$$

The track curvature is used to measure the momentum of a charged particle within a detector. The CMS detector is designed to measure especially the properties of muons very precisely. This can be very difficult for muons with extreme high momenta of $\approx 1\,\mathrm{TeV/c}$. In order to fulfil this task and measure the charge sign of even high energetic muons unambiguously, the momentum has to be measured with a resolution of $\Delta p/p \approx 10\,\%$. This demands a very high magnetic field to get a measurable curved track. CMS uses a superconducting solenoid which produces a field of 3.8 T in the barrel region. Additionally the iron return yoke (see 3.2.6) bends the magnetic flux on the solenoids outside and leads to a field of 1.9 T in the opposite direction.

3.2.6. Muon system

As already mentioned the iron return yoke catches the solenoids outer flux and provides a magnetic field of 1.9 T for the outer part of the CMS detector. It is interleaved with different kinds of muon chambers. In the barrel region ($|\eta| < 1.2$) there are drift tubes (DT). Here, the neutron induced background is small, the muon rate is low and the residual magnetic field in the chambers is low. The situation in the two end caps is quite the opposite. Here cathode strip chambers (CSC) are deployed which cover the region up to $|\eta| < 2.4$. Additionally in both, the end caps and the barrel, resistive plate chambers (RPCs) are used which are operated in avalanche mode to ensure good operation at high rates. To accurately measure the muon momenta, a combination of muon chamber and tracker information is used. A profile of one quarter of the CMS muon system can be seen in figure 3.5

3.2.7. Trigger and data acquisition

At design luminosity 40 million bunch crossings per seconds are forseen, which lead to $\sim 10^9$ pp collisions per second in the CMS detector. Each of these events produces signal in some of the detector elements presented in this section.

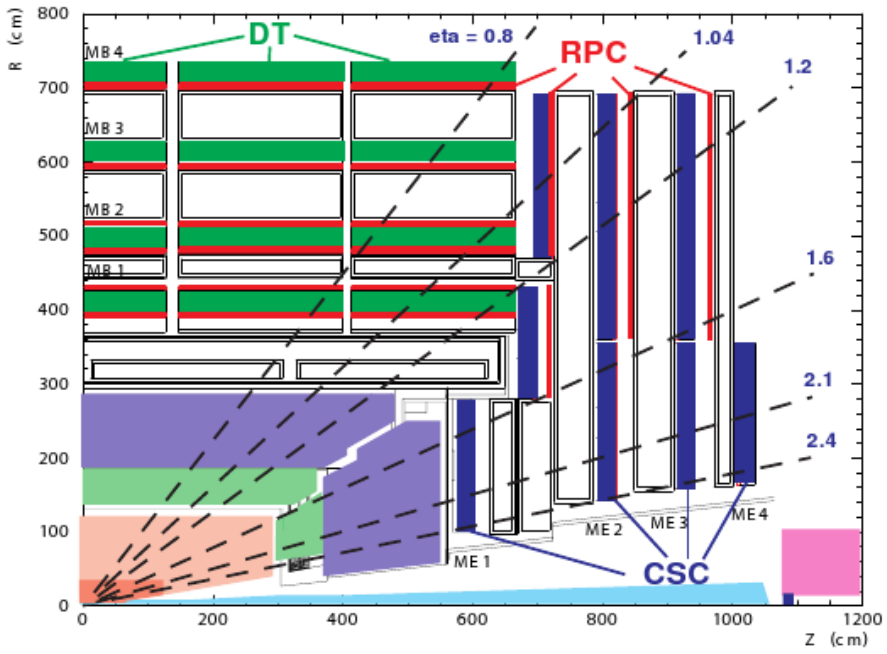


Figure 3.5.: One quarter of the CMS muon system. Taken from [39]

With todays technology it is not possible to collect all these information properly in time and store them in any kind of mass storage. Instead it is necesary to discard most of the events during the data taking. In order to still select those events, which are likely to contain physics processes of interest, the concept of triggering was introduced, which partially reconstructs and evaluates the events online.

The design of triggers requires the trade off between precission and speed. While a fast trigger is not able to consider every aspect of the event, a decent trigger takes too long to examine the event. Hence the CMS triggering is done in two steps, the Level-1 and the High Level Triggers, which are explained below.

During the design of CMS and the readout electronics it was estimated to be able to collect just about 150 events per second. Due to faster hardware, this could already be increased to about 300 events during the 2011 data taking.

Level-1 triggers are implemented in specialised hardware, to get a very fast response. They have to reduce the number of events by a factor of ∼1000 taking only rough information into account. These information are the presence of trigger primitives like electrons, muons and jets over a certain threshold as well as raw global sums of E_t and $E_\mathrm{t}^\mathrm{miss}$ based on reduced granularity. While the triggers "decide" data is being stored in hardware buffers. Those events passing Level-1 trigger are passed to High Level Triggers, the others are discarded.

High Level Triggers (HLT) further reduce the number of events to be stored using more sophisticated algorithms and partially reconstructing the whole event. This is done on a dedicated computing cluster using special HLT algorithms. However the gran-

ularity for event reconstruction is lowered to increase computing speed. The HLT cluster is installed close-by to the CMS detector for short latencies.

Since every event uses approximately 1.5 MByte for its raw data, about 450 MByte/s of data is stored. This induces enormous requirements to storage as well as computing. Because no single computing centre could fulfil these requirements properly, and the very expensive production of these data demands delocalised backup, the Grid was invented which will be explained later in section 4.7.

There are a lot of different triggers running in CMS and it is not possible to store all events which fire any given trigger. Hence some of the triggers are "prescaled", i.e. only a certain fraction of events they trigger will be stored in the end. The individual prescaling of each trigger is negoatiated among the Physics Analysis Groups and can change from one run to another, or even within one run e.g. if pileup decreases.

Since the triggers have a high impact on the effective integrated luminosity the selected dataset will have, it is very important to pay attention to all these trigger effects. Especially in cross section measurements like this analysis, calculating the wrong integrated luminosity for the dataset used, will lead to a false result.

4. Data acquisition and event reconstruction

The physical detection of particles produced in proton-proton collisions at the LHC is only the first step of the data taking process. The raw electrical signals read out of all detector components are evaluated in several steps of filtering and reconstruction, until the underlying physics process can be explored.

The data samples from the LHC collisions are usually compared to generated and simulated events from Monte Carlo generators. These generated events make it possible to study each process individually and the true particle information like the flavour are accessible. An overview of the MC generators utilised in this thesis is given in section 4.1.

The reconstruction of physical objects like tracks and jets is explained in section 4.2. It is also necessary to identify the type of particle which produced a track or a jet. The identification of leptons is discussed in section 4.2.3, while section 4.4 is devoted to the different b jet identification algorithms in CMS.

The reconstructed objects raw properties need to be corrected to reduce noise and to calibrate the measured signal. This is explained in section 4.3.

After these corrections are done, the reconstructed physics objects are used to reconstruct the t and W candidate. The necessary steps are described in section 4.5.

The technical part of the data processing is briefly described in sections 4.6 and 4.7.

4.1. Monte Carlo generators and detector simulation

Each analysis requires generated events to optimise selection cuts, train multivariate methods or check selection efficiencies. These events are produced by Monte Carlo (MC) generators. The phrase "Monte Carlo" refers to the fact, that they make use of random numbers to simulate the quantum mechanical processes

The full simulation from the proton collision to the detector read-out is usually done in three steps.

1. Hard process generation.
   Simulate the parton interaction in the collision and the generation of new partons in the inelastic scattering process. The simulation uses parton distribution functions as presented in section 2.2.1 in a random process and predicted probability densities for the specified process. The latter are usually derived from the evaluation of the Feynman diagrams that need to be taken into account for the process that is to be generated. Both probability densities are then folded to determine the actual probability for the process in the collision events.

2. Showering and hadronisation.
   The hard process generators mentioned before are well suited to simulate particle interactions with high momentum transfer Q, but the calculations do not converge if Q is small, which is the case for gluon and photon radiation of the generated quarks and leptons. Hence a second simulations step is needed to account for these radiations and the confinement in the strong interaction to form hadrons. This is usually done in a dedicated showering and hadronisation simulation. The gluon radiation's structure is given in terms of branchings, which can be described using Dokshitzer-Gibov-Lipatov-Altarelli-Parisi (DGLAP) [17, 58, 68] equation evolution. The probability for a gluon radiation is then given by the Altarelli-Parisi splitting functions.

3. Detector simulation
   The interaction of stable particles and the detector material is simulated. This includes trajectories in the magnetic field as well as calorimeter showering. In the end the event should produce a similar readout picture from the detector elements as real data does, which can than be used in the original event reconstruction algorithms, as they are described in sections 4.2.

Generated datasets for CMS analyses use all three types of MC generators. First the hard initial process is simulated. These programs use perturbative QCD calculations, which do not work for softer processes, i.e. for quark gluon radiation processes which happen at a lower Q^2 scale. Hence all particles which carry less than a minimum momentum Q_{min} are not further treated by the matrix element generators. Instead their evolvement is simulated by the second type of MC generator. The matching between these two simulation steps, the hard matrix element and the soft showering and hadronisation process, is usually described by the MLM or the CKKW scheme [30, 70].

In the last step the on shell particles are processed in the detector simulation.

These processes are needed to have a full mapping between the particles generated in the hard process and the final reconstructed event. In addition reconstruction algorithms are optimised with simulated events, using the true information of the generated particles.

In the following only the small subset of the most important MC generators for this thesis are discussed. A more complete overview can be found at [56]. Between different generator steps particles are usually forwarded in the Les Houches file format [18], which is a standardised format for this purpose.

4.1.1. MadGraph

MadGraph [101] is a matrix element generator, which is used in the multi-purpose tree level event generator MadEvent [82]. MadGraph automatically generates the amplitudes for all relevant subprocesses and produces the mapping for the phase space integration. It does not produce any particle showering. Therefore the MadGraph output has to be further processed by general purpose generators like Pythia or Herwig(++) [24].

4.1.2. POWHEG

The Positive Weight Hardest Emission Generator method [16, 64, 89] is a general method for interfacing parton-shower generators with NLO computations. A general problem when doing this interfacing between NLO hard interaction simulation and LO showering is the potential double counting of events, because the showering algorithms also implement approximate NLO corrections [64]. The POWHEG method uses a special technique to solve this overcounting issue but still using only positive events weights.

It can simulate certain process, among which there are also the single top production in the t- and s-channel and the associated tW production [15, 95] and can be interfaced to Pythia, Herwig and the like for showering. In this analysis it is used to produce all three kinds of single top events.

4.1.3. Pythia

Pythia is a multi-purpose generator for particle collisions at high energies. It "contains a subprocess library and generation machinery, initial- and final-state parton showers, underlying event, hadronization and decays, and analysis tools" [56]. Hence it is widely used to generate MC events and also to simulate the showering and hadronisation of events which where produced in other hard process generators.

4.1.4. CompHEP

The CompHEP generator [23, 94] is a meta-generator which does not come with a library of various precalculated matrix elements, but calculates symbolically the matrix elements for any given process defined by the user. The process definition is done in terms Feynman rules for a gauge model Lagrangian.

An extensive study of different MC generators and their suitability for the generation of single top events and the most important backgrounds has been done in [22]. Also the most discriminating variables were studied there. POWHEG was taken into account in [96] and turned out to be the best generator for the electroweak single top quark production. Hence it is used for the simulation of these events, while MadGraph is utilised for heavy boson and $t\bar{t}$ events. Pythia is used for the showering and hadronisation of the generated events mentioned before.

4.1.5. Detector simulation

The showered and hadronised particles are processed in a detector simulation in order emulate the detector's response and to be able to apply the event reconstruction algorithms. This has to take into account all kind of physical processes that happen with high energetic particles in matter, like multiple scattering, ionisation, bremsstrahlung and hadronic and electromagnetic interactions. A detailed model of the CMS detector is used in the Geant4 Toolkit [13].

The simulated events can then be treated just like real measured events and all reconstruction algorithms can be applied to them. Also triggers are simulated to measure their efficiencies and turn-on curves.

4.2. Physics object reconstruction

Events are reconstructed in several steps starting with raw detector signals and transforming them into physics objects which can be used for physics analysis. A general overview of the whole CMS data processing can be found in [39]. The most important physics objects and the reconstruction of certain properties, which are used in the following chapters, are briefly described in the following.

4.2.1. Track reconstruction

As was described in section 3.2.2, CMS utilises a very precise tracking system in its inner layer. Charged particles that are produced in LHC collisions traverse the tracker and produce hits. Inside the CMS detector and its magnetic field, the tracks of charged particles are curved due to the magnetic field and the Lorentz force.

The tracker hits are read out as isolated dots on the tracker layers. There are several track reconstruction algorithms which try to fit tracks through the set of three dimensional coordinates. They take the curved shape and potential energy and momentum loss due to interactions with the tracker material into account.

The two track reconstruction algorithms used in CMS are:

Kalman-Filter (KF): This algorithm [12] takes a group of at least three pixel hits from the inner most tracker layer as seed to start from. These hits are used to calculate a coarse flight direction of the particle. From this track seed the filter proceeds iteratively to the following layers along the approximated trajectory and searches for additional tracker hits, which are compatible with the current track. For the step from one to the next layer the tracks curvature in the magnetic field as well as the energy loss in the material are taken into account. Since there might be several hits in the subsequent layer, which are compatible with the extrapolated trajectory, the algorithm follows several track candidates. The ones with the worst fit result are truncated later on, to avoid an exponential growth of the candidate list.

As a least squares estimator, the Kalman filter algorithm has one major disadvantage. To take radiative energy loss into account, the momentum is corrected by the mean value of the energy loss in each propagation step and the momentum's variance is increased by the variance of the energy loss distribution. Thus the Kalman filter is only optimal when all probability distributions are Gaussian. This makes it necessary to use a modified algorithm for electrons, since they radiate a lot of their energy as bremsstrahlung. This radiative energy loss can not be described by a single Gauss perfectly, which leads to the Gaussian Sum Filter, that is briefly explained next.

Gaussian Sum Filter (GSF): To be more precise with the energy loss probability distributions than the Kalman-Filter, the *Gaussian Sum Filter* (GSF) [11] approximates the energy loss distribution of electrons with a weighted sum of Gaussians instead of a single Gaussian as in the KF. This improves its ability to find the correct hits in the subsequent tracker layer and leads to more correctly reconstructed electron tracks. c

4.2.2. Vertex reconstruction

Once all track candidates are reconstructed, they can be used to search for vertices. In general a vertex is the intersection point of two or more track trajectories. In the event reconstruction they are of interest because they indicate the place of a particles decay. This is especially the case for the points where a hard proton-proton interaction happened, the so called "primary vertices". The "secondary vertices" usually indicate the displaced decay of a b quark, which is used in some of the b tagging algorithms (see section 4.4).

In CMS the vertices are fitted by the Adaptive Vertex Fitting method (AVF) [65], which is a modified version of the Kalman-Filter, that was described above. All tracks are assigned a weight which depends on their compatibility χ^2 with the vertex candidate. The vertex candidates are then iteratively refitted and the weights are updated after each step. The final set of vertices is sorted by the weighted sum of p_T^2 of all associated tracks.

Due to pile-up interactions there are usually several primary vertices in an event. The one with the highest weighted p_T^2 sum is considered to be the point of the hard interaction.

Detailed results on the performance of track reconstruction and vertex finding can be found in [46].

4.2.3. Lepton reconstruction and identification

The most important track-like objects for this analysis are the isolated electrons and muons. Their individual reconstruction and identification algorithms are described in the following.

Muons

There are three types of muons reconstructed in CMS events

1. **Standalone muons,** which are reconstructed using only muon system information.
2. **Tracker muons,** which are reconstructed using tracker information only.
3. **Global muons,** that are formed by muon system and tracker information.

In both cases a track in the muon system in reconstructed using the Kalman-Filter algorithm. The global muon is then extended by searching for a matching tracker track which is combined with the standalone part. This matching takes into account the energy loss and the multiple scattering in the detector material, especially the solenoid and the calorimeter. If a matching tracker track is found, the associated hits in both detector components are used to fit the global muon, starting with the tracker hits [39]. At this stage the reconstructed objects are only muon candidates. Further refinement is needed to suppress hadronic punch throughs, i.e. hadrons that are not fully stopped in the hadron calorimeter and enter the muon system. Hence the candidates need to pass a muon identification before being considered a physical/good muon. The recommended muon selection requirements for 2011 data analysis are [88]:

- The candidate is reconstructed as a global muon, producing hits in the tracker and the muon system.

- The ratio between the χ^2 of the global muon track fit and the number degrees of freedom has to be smaller than ten; $\frac{\chi^2}{\text{ndof}} < 10$.
- There has to be at least one muon chamber hit.
- There are muon segments in at least two muon stations.
- The transverse impact parameter of the tracker track is smaller than $d_{xy} < 2\,\text{mm}$ with respect to the primary vertex.
- There must be hits already in the pixel detector to suppress muons from decays in flight
- At least 8 tracker layers must contain hits from the candidate.

Electrons

Electron reconstruction in CMS is closely related to the reconstruction of photons, since both produce showers in the electromagnetic calorimeter (ECAL). Because the tracker poses a lot of material budget in front of the calorimeter, many photons will convert into electron-positron pairs even before reaching the ECAL. For the same reason electrons and positrons radiate bremsstrahlung in the tracker. This produces additional photons that are reach the ECAL and are spread in φ direction due to the track curvature. This energy and momentum loss changes the electron/positron track curvature in the magnetic field, hence the GSF algorithm for the reconstruction of these particles' tracks.

The electrons in CMS are composed of a single track from the primary vertex and a matching ECAL supercluster. A "supercluster" is a cluster of ECAL clusters which is spread out in φ to retrieve bremsstrahlung showers. Details on the identification of super clusters and the matching of tracks to these clusters can be found in [39].

4.2.4. Jet reconstruction algorithms

The strong interaction forbids the existence of isolated colour charged particles, i.e. quarks never exist isolated but they build mesons and baryons, which are colour neutral. As was described in 2.2.1, the process of hadronisation leads not only to isolated particles but to sprays of collimated particles, called "jets", hitting the detector material. The event reconstruction software at hadron colliders use dedicated algorithms to identify these jets in data. The goal of these algorithms is to correctly cluster entries in the calorimeters and potentially also the tracker and the muon system, which originate from the same mother particle, e.g. a b hadron. Energy and momentum for each jet have to be reconstructed with high precision since they are the most important information about the underlying parton.

Two pitfalls have to be considered by the jet algorithms in order to give stable results and be compatible with theory calculations.

infrared safety: Additional soft radiation should not alter the number of reconstructed jets.

collinear safety: The collinear radiation of gluons could lead to cluster seeds that are below the required threshold for jet seeds. In this case the jet would not be reconstructed while it would be reconstructed without the radiation. This change in the jet algorithm's result due to collinear radiation should not happen. This is especially important for comparison with theoretical calculations.

There are two major types of jet algorithms which are used in CMS event reconstruction, the cone algorithms and the sequential clustering algorithms:

(iterative) Cone algorithms as the name suggests, work with cones of a given radius in the $\eta - \phi$ plane. It starts with a list of input objects (particles or calorimeter entries) fulfilling certain quality criteria which are ordered by transverse energy E_T. The first object from this list, i.e. the one with the highest transverse energy is taken as seed for a proto-jet. Other objects within the selected cone are added one by one while recalculating the jet's energy and direction after each step. The process is stopped when a certain criteria is reached. This could for example be the maximal number of iteration steps or the relative change for the jet energy or direction between two steps has dropped below a certain threshold. The proto-jet is then considered a stable object and its constituents are removed from the list of objects. The algorithm is then repeated with the remaining objects until none are left in the list. [39]

This algorithm is not collinear safe since only candidates above a certain threshold are used. One could solve this problem by identifying all stable cones without using seeds. The computing effort for this is $\mathrm{O}(N2^N)$ for N particles, which is not possible at hadron colliders due to their high particle multiplicity per event. But there are seedless infrared save cone algorithms (SISCone) which use split and merge methods that can guarantee infrared and collinear safety while being computable in reasonable time with a complexity of $\mathrm{O}(N^2 \ln N)$ [98].

Sequential clustering algorithms in contrast to cone algorithms do not yield a specific jet shape. For each input particle i the distance to any other particle j $d_{i,j}$ is calculated as well as the distance to the beam axis $d_{i,B}$. The distance between two objects is defined as

$$d_{i,j} = \min\left(p_{T_i}^{2n}, p_{T_j}^{2n}\right) \cdot \Delta R_{i,j}^2 \tag{4.1}$$

while the distance between any object and the beam is calculated by

$$d_{i,B} = p_{T_i} \cdot D^2 \tag{4.2}$$

The resolution parameter D is introduced to control the size of the jets and depends on the energy of the object and the geometrical distance ΔR. There are three different sequential clustering algorithms which differ in the value for n:

n=1 standard k_T algorithm [29]

n=0 Cambridge-Aachen algorithm [57, 108]

n=-1 anti-k_T algorithm [26]

The algorithmic prescription for all three algorithms is the same and depends on $d_{i,j}$ and $d_{i,B}$. If the smallest distance of all found is between two objects, they are merged into one and the original objects are removed from the list. This is repeated until the smallest distance found is between a object and the beam. When this is the case the object is removed from the list of objects and considered a reconstructed jet. The above steps are repeated until all objects are included in reconstructed jets.

4.2.5. Missing transverse energy

In collider experiments, like the LHC, were the flight direction of the colliding particles are parallel in the interaction point, it is expected that the deposited energies should be balanced in the plane transverse to the beamline, due to momentum conservation.

However this is not the case if not all escaping particles are detected. From the standard model particles this only happens for the neutrinos, but other theories beyond the standard model also predict barely interacting particles which can escape the detector without leaving a direct trace.

These kind or particles are indirectly detectable as missing transverse energy and momentum, making the reconstructed event imbalanced in the transverse plane. It is calculated as the negative vectorial sum over all N transverse energy deposits.

$$\vec{\not{E}}_{\mathrm{T}} = -\sum_{n=1}^{N} E_n \left(\sin\theta_n \cos\varphi_n \hat{x} + \sin\theta_n \sin\varphi_n \hat{y}\right) \tag{4.3}$$

$\hat{x}$ and $\hat{y}$ are the unit vectors for the x and y direction, while θ and φ denote the polar angles as defined in section 3.2.1.

Potential inaccuracies for this calculation arise from particles escaping the detector with a high rapidity close to the beamline and a not well calibrated calorimeter response. The latter is taken care of by the jet energy correction, as described in section 4.3.1.

4.2.6. Particle Flow objects

Particle Flow [41] objects are more elaborated objects, reconstructed simultaneously from all CMS subdetectors. Each object is then classified as a stable particle like electron, muon, pion or photon depending on the pattern they produce in the detector.

This classification at an early stage makes it possible to apply specialised recipes for the energy calibration, e.g. electrons are known to radiate a lot of energy as bremsstrahlung while traversing the tracker. The particle flow reconstruction tries to find these tangentially radiated photons in the ECAL and adds their energy to the original electron. This energy would otherwise be missing and could, on the other hand, be interpreted as additional photons from another process. The calibrated objects are then used for the jet clustering which in turn yields better calibrated jets. This improves the performance for jets, taus and missing transverse energy compared to the generic object reconstruction. For more details on the performance of Particle Flow in the CMS event reconstruction refer to [43–45].

In the following jets reconstructed with the anti-k_T algorithm with a distance parameter $R = 0.5$ and particle flow objects as input objects are used.

4.3. Physics object corrections

4.3.1. Jet energy corrections

CMS uses a factorised multi-level *Jet Energy Corrections* (JEC) system. They are needed to calibrate reconstructed jets and to guarantee consistency with the energy of the original parton [47].

L1FastJet: These corrections are applied to account for pile-up (PU) and the electronic noise in the detector.

L2Relative: The non-uniformities in the CMS detector system lead to different jet response in different η regions, which are accounted for by this correction.

L3Absolute: p_T dependent correction account for the non-linear response for particles with different energies.

Also the L2L3Residual corrections are applied which take into account the fact, that the L2 and L3 corrections are derived from MC only. There are more levels (L3 - L7) which are however currently not used as defaults within CMS.

4.3.2. Pile-up and underlying events

During the 2011 LHC run the pile-up of events was already an important issue for the CMS event reconstruction and will get even more important in future runs with higher instantaneous luminosities. Pile-up refers to the effect of having additional signals in the detector which are not directly related to the hard interaction of this event. There are two major groups of pile-up.

1. In-time pile-up: During a bunch cross not only one pair of protons will interact but it is estimated that up to 25 collisions can happen. This will typically increase the number of primary vertices in an event.

2. Out-of-time pile-up: Detector signals from neighbouring events are present in the detector. This can either be remaining signals from the previous event or newly added from the following, which hit the detector before the current event is fully read out. The severity of out of time pile-up is closely dependent on the time gaps between two bunch crossing (also called "spacing") which were 50 ns in early 2011 runs.

MC samples are generated with the `flat10` pile-up scenario in 2011, i.e. the probability to have zero to ten pile-up events in a generated event is flat and decreases smoothly for higher pile-up rates. The events need to be reweighted according to the pile-up of the data sample one is studying, as is explained in 6.2.6.

The effect of pile-up is mitigated as much as possible in data and MC. There is on the other hand the L1 FastJet Pile-up subtraction which estimates pile-up as a diffuse noise in the detector and subtracts the added energy from the detector cells on a per event basis. On the other hand there are the PFnoPileUp algorithms for ParticleFlow

objects, which subtract charged tracks from the event which are not compatible with any reconstructed vertex.

Another process that needs to be considered in event simulations besides the hard interaction is the underlying event. This means the interaction between the partons taking part in the hard interaction and the proton remnants, and the interaction of the proton remnants with each other. These interactions need to be considered for example in the hadronisation process, because they lead to additional gluon strings that may produce additional particles.

4.4. b Tagging

b Tagging refers to the identification of jets which originate from the hadronisation of a b quark. This is of special interest, since b quarks are often produced in the decay of heavier particles, like W, Z bosons and the top quark. Also, several hypothetical heavy particles from physics processes beyond the standard model are expected to produce b quarks in their decay chain. Hence it is important to identify those jets stemming from a b quark with a high precision since it is one of the keys to identify these physics processes in the vast amount of events. With a reliable b tagging at hand it is easier to identify events that include the heavy particles listed above. Therefore b tagging is even utilised in the high level triggers (HLT), one of which is used in this thesis (section 3.2.7).

There are several b tagging algorithms available in CMS software, which exploit different features of the weak b quark decay in order to identify these jets [42].

The most important feature is the relatively long lifetime of b quarks which is currently measured as $\tau = 1.32 \pm 0.04 \cdot 10^{-12}\,\mathrm{s}$ [69]. At a velocity close to the speed of light, a b hadron may travel about $c\tau \approx 480\,\mu\mathrm{m}$ before decaying into lighter particles. This travel distance has a measurable impact on the trajectory of tracks originating from the b decay. When extrapolating the tracks to the primary vertex, tracks stemming from a distant b decay will have a rather large impact parameter, i.e. the smallest distance between the track trajectory and the primary vertex.

In the CMS software multiple algorithms make use of the impact parameter (IP) or the impact parameter significance IP/σ_{IP}. The IP is signed based on the scalar product of the IP segment and the jet direction. Negative signs occur, if the reconstructed jet axis differs from the real B-hadron flight direction. One group of taggers using this effect are the TrackCounting taggers, which only utilise one of the tracks inside a jet at a time. The jet probability taggers in contrast take into account several tracks in a jet simultaneously and combine their information into one b tag discriminator.

The so called "secondary vertex" taggers make use of several tracks information at a time, too. A secondary vertex can be found if there are several tracks which are incompatible with the primary vertex, because they have a high impact parameter (significance) inside a jet. If at least three track trajectories have of common intersection point, this point can then be identified as a secondary vertex.

There are currently two algorithms available in CMS software, which exploit information extracted from the reconstructed secondary vertex. These are the simple secondary vertex (SSV) and the combined secondary vertex (CSV) tagger. The latter differs from

the first by returning a meaningful result for all jets, even if no good secondary vertex could be fitted, while the SSV only can tag jets with vertices.

The third class of b tagging algorithms utilises information about the soft lepton which is produced in the b decay. Since the b quark decays in about 20 % of the cases into an electron or a muon, roughly 40 % of all b decays qualify for this tagger category at all. They are however of interest for the calibration of b taggers among each other, since they provide a way of tagging b jets which is independent of the information lifetime based taggers use.

The track counting algorithms are used in the following. They are rather simple taggers, which just compute the impact parameter significance for the tracks associated to the jet and order them by decreasingly. The actual discriminator for the jet is the significance of the Nth track, where N depends on the specific tagger type. The *track counting high efficiency (TCHE)* tagger uses the impact parameter significance of the second track, hence in this case $N = 2$. The *track counting high purity (TCHP)* in contrast uses the third track from the list ($N = 3$). The reason for connecting these N's with the characteristics 'efficient' and 'pure' is that there are more jets which have two tracks with high impact parameter significance, but not all of them are real b jets. On the other hand it is more likely that b jets also have a third track with a high IP_{sig}.

The b tagging and vertexing Physical Objects Group at CMS supports three working points for each tagger and provides b tagging efficiencies, mistag rates and data/MC scale factors for these three working points. The working points, loose (L), medium (M) and tight (T) are those discriminator values at which the light flavour mistag rate is measured in MC to be 10 %, 1 % and 0.1 % respectively. The b tagging efficiencies and mistag rates are listed in tables 4.1 and 4.2.

Table 4.1.: Measured b tagging efficiencies and data/MC scale factors for several b tagging algorithms and operating points. Results for the muon-jet p_T between 50 and 80 GeV are indicated for the PtRel and System8 methods. Uncertainties are statistical for ε_b^{tag} and statistical and systematic for SF_b. [48]

b tagger	PtRel		System8	
50-80 GeV	ε_b^{tag}	SF_b	ε_b^{tag}	SF_b
TCHEL	0.76 ± 0.01	$0.95 \pm 0.01 \pm 0.05$	0.77 ± 0.01	$0.96 \pm 0.02 \pm 0.05$
TCHEM	0.63 ± 0.01	$0.93 \pm 0.02 \pm 0.06$	0.63 ± 0.02	$0.93 \pm 0.02 \pm 0.07$
TCHPM	0.48 ± 0.01	$0.92 \pm 0.02 \pm 0.05$	0.49 ± 0.01	$0.93 \pm 0.03 \pm 0.09$
SSVHEM	0.62 ± 0.01	$0.95 \pm 0.02 \pm 0.07$	0.60 ± 0.01	$0.94 \pm 0.02 \pm 0.06$
SSVHPT	0.38 ± 0.01	$0.89 \pm 0.02 \pm 0.06$	0.37 ± 0.01	$0.90 \pm 0.03 \pm 0.05$
TCHPT	0.36 ± 0.01	$0.88 \pm 0.02 \pm 0.05$	0.37 ± 0.01	$0.88 \pm 0.03 \pm 0.07$

On trigger level a *TCHE* tagger is used. It is not exactly the same as the offline tagger, since it uses trigger objects as input. These are not reconstructed as precise as offline objects due to the lack of computing time until a trigger decision has to be delivered. For the offline event selection the *TCHP* tagger is used at the tight working point (discriminator threshold is 3.41 for a jet to be tagged as b jet).

Table 4.2.: Mistag rate and data/MC scale factor for different b taggers and operating points for jets with p_T between 50 and 80 GeV. The statistical and systematic uncertainties are quoted. [48]

b tagger	mistag rate	SF_light
TCHEL	$0.128 \pm 0.001 \pm 0.026$	$1.11 \pm 0.01 \pm 0.12$
TCHEM	$0.0175 \pm 0.0003 \pm 0.0038$	$1.21 \pm 0.02 \pm 0.17$
TCHPM	$0.0177 \pm 0.0002 \pm 0.0036$	$1.27 \pm 0.02 \pm 0.15$
SSVHEM	$0.0144 \pm 0.0003 \pm 0.0029$	$0.91 \pm 0.02 \pm 0.10$
SSVHPT	$0.0012 \pm 0.0001 \pm 0.0002$	$0.93 \pm 0.09 \pm 0.12$
TCHPT	$0.0017 \pm 0.0001 \pm 0.0004$	$1.21 \pm 0.10 \pm 0.18$

The b tagging POG published an efficiency of $\approx 36\,\%$ and a mistag rate of only $0.17\,\%$ for this working point, as listed in tables 4.1, 4.2.

4.5. Reconstruction of t and W candidates

In the single top quark analysis it is important to reconstruct the top quark candidates as precisely as possible. This is only achievable by selecting the correct final state particles from the detector objects. As described in 2.4.3 and depicted in figure 2.10, the semileptonic decay of a top quark has the following event signature. These are:

- One isolated high energetic lepton
- Two or three jets, depending on whether the spectator b jet enters the detector or not.
- At least one of the jets can be identified to stem from a b quark ("b tagged") 4.4
- Missing transverse energy due to the undetected escaping neutrino.

These components need to be reconstructed and identified in each event under consideration and combined in such a way, that it is possible to form a W candidate and further on a t candidate.

This is done by walking the t quark decay Feynman graph 2.10 backwards and first combining the missing transverse energy ($\vec{\not{E}}_{\mathrm{T}}$) and the required lepton to a W candidate, which implies the calculation of momentum and energy of the escaping neutrino. These are constrained by the W mass and have to obey the relativistic energy and momentum conservation for the lepton and the neutrino, which is given in formula (4.4).

$$M_W^2 = (E_{lep} + \sqrt{\vec{\not{E}}_{\mathrm{T}}^{\,2} + P_{z,\nu}^2})^2 - (\mathbf{P}_{\mathrm{T},lep} + \vec{\not{E}}_{\mathrm{T}})^2 - (P_{z,lep} + P_{z,\nu})^2 \tag{4.4}$$

Solving this quadratic equation for the z component of the neutrino momentum $P_{z,\nu}$, yields in general two solutions.

$$P_{z,\nu}^{A,B} = \frac{\mu \cdot P_{z,lep}}{P_{\mathrm{T},lep}^2} \pm \sqrt{\frac{\mu^2 \cdot P_{z,lep}^2}{P_{\mathrm{T},lep}^4} - \frac{E_{lep}^2 \cdot \vec{\not{E}}_{\mathrm{T}}^{\,2} - \mu^2}{P_{\mathrm{T},lep}^2}}\,, \tag{4.5}$$

with

$$\mu = \frac{M_W^2}{2} + \mathbf{P}_{\mathrm{T},lep} \cdot \vec{\not{E}}_{\mathrm{T}}\,. \tag{4.6}$$

The discriminant in equation (4.5) turned out to be positive in more than 60 % of the performed reconstructions of generated single top events in the signal regions with two or three jets of which one is b tagged. The precise numbers are given in table 4.3. In this case the solution with the smallest $P_{z,\nu}$ is chosen. In 23.1 % of the cases the discriminant is negative and the $P_{z,\nu}$ has an imaginary part. Instead of just ignoring the imaginary part or modifying the W mass, it is assumed in this analysis, that this misbehaviour of the solution in equation (4.5) is due to a finite resolution of missing E_T ($\vec{\not{E}}_{\mathrm{T}}$). Therefore the $\vec{\not{E}}_{\mathrm{T}}$ components are tuned such, that the discriminant vanishes ($M_{T,W} = M_W$) while complying with equation (4.4) [37]. This leads to an ambiguity for the x and y components of the neutrino momentum since they are then related via a quadratic relation. In this case the distance between the transverse neutrino momentum and the missing transverse energy is calculated for both cases as given in equation (4.7) and the solution with the smaller distance is selected.

$$\delta_{1,2}(P_{x,\nu}) = \sqrt{(P_{x,\nu} - (\vec{\not{E}}_{\mathrm{T}})_x)^2 + (P_{y_{1,2},\nu}(P_{x,\nu}) - (\vec{\not{E}}_{\mathrm{T}})_y)^2} \tag{4.7}$$

Having the W boson reconstructed out of the lepton and missing transverse energy in the next step, the b jet from the top decay has to be selected and added to the W boson. These three particles are then expected to be the t candidate. Depending on how many jets are required and how many of them are considered to be a b candidate (b tagged) different selection criteria have to be applied in order to identify the correct b jet. Also the light quark jet from the t production has to be identified correctly since it is needed to calculate the spin correlation in the single top quark production ($\cos\theta$, see section 4.5.1). These criteria are:

2 jets: In the 2 jet region either one can have a b tag or not. The region where we expect the most single top t-channel events has exactly two jets of which only one is b tagged and the other one not.

- **1 jet b tagged:** the tagged jet is selected as b jet while the other one is the light quark jet.
- **0 or 2 jets b tagged:** The pseudorapidity η of the jets is compared and the more central jet is considered to be the b jet from the top decay while the more forward jet is taken as light quark jet.

3 or 4 jets: In the case where more than two jets are present, it is more difficult to assign the parton candidates to the jets correctly.

no jets b tagged: As in the case for two untagged jets, the jet with most forward jet, i.e. the one with the highest pseudorapidity, is taken as light quark jet. The others are sorted by the b tag discriminator and the one with the highest value is taken as b jet from the t decay. The second highest b tagged is considered to be the spectator b jet from the top quark production.

at least 1 jet b tagged: Here the jets are sorted by the b tag discriminator, again. The one with the highest value is considered to be the b jet from the top decay while the one with the lowest discriminant is used as light quark jet candidate.

The matching recipe given above was tested on MC samples and the correctness of the assignments was retrieved. In table 4.3 the results are listed.

Table 4.3.: Single top event reconstruction statistics for the correctness of the chosen b jet hypothesis and the ν reconstruction

	muon		electron	
	2j1t	3j1t	2j1t	3j1t
b tagged jet is b from top	87,82%	64,95%	86,42%	64,58%
b tagged jet is 2nd b	9,02%	19,77%	10,58%	19,82%
b tagged jet is light quark	0,34%	5,77%	0,35%	5,37%
b tagged jet is none of above	2,81%	9,50%	2,65%	10,23%
non b tagged jet is b from top	4,64%	6,57%	4,82%	6,91%
non b tagged jet is 2nd b	5,24%	10,64%	5,44%	9,93%
non b tagged jet is light quark	82,72%	60,86%	82,46%	59,98%
non b tagged jet is none of above	7,39%	21,93%	7,27%	23,17%
Neutrino solution is complex	27,94%	29,65%	37,23%	37,68%
Neutrino solution is real	72,06%	70,35%	62,77%	62,32%

It is found that in the muon channel for events with two jets of which one is b tagged, the b tagged jet is correctly assumed to be the b from the top decay in more than 87 % of the cases. For events with three jets the matching is correct in almost 65 % of the cases.

In the electron channel the derived efficiencies are above 86 % and 64 % for events with two or three jets of which one is b tagged.

4.5.1. Reconstruction of the spin polarisation $\cos\Theta$

The correct assignment of the light quark jet is especially important to exploit the spin polarisation in the production and the decay of single top quarks (see section 2.4.4). This spoing polarisation is a result of the $V - A$ coupling of the weak interaction. It can be measured as a specific shape for the distribution of the cosine of the angle θ between the lepton from the W boson decay and the light quark jet in the t quark rest frame. The normalised shape distributions for this observable, together with the pseudo rapidity of

the light quark η_{lq}, which has a unique shape as well, are shown exemplary for events with a muon and two jets of which one is b tagged in figure 4.1.

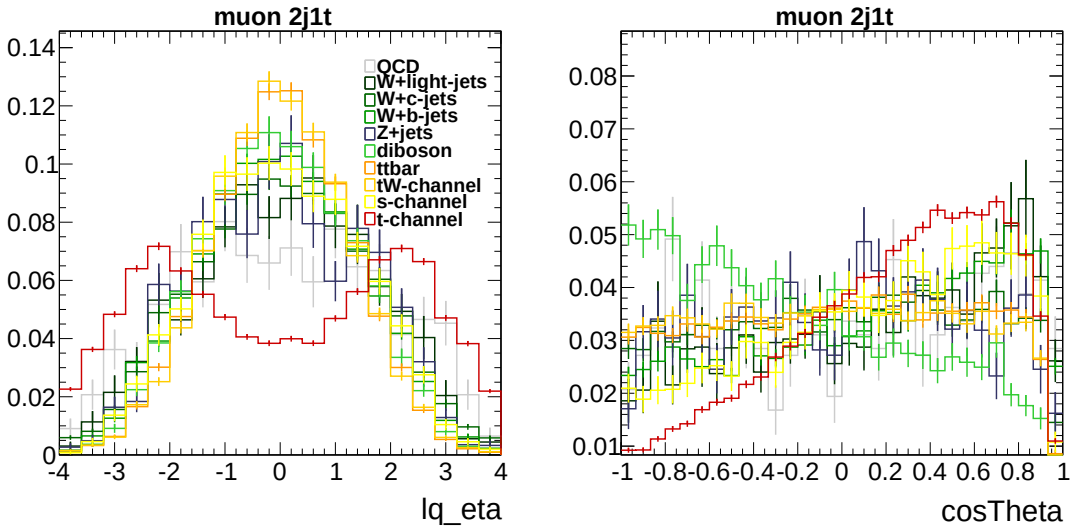


Figure 4.1.: Shapes of η_{lq} and $\cos\theta$ in 2j1t muon channel

The $\cos\theta$ is calculated using the momenta of the light quark $\overrightarrow{p}_{lq}$ and the lepton $\overrightarrow{p}_{lep}$ in the t quark candidate rest frame as

$$\cos\theta = \frac{\overrightarrow{p}_{lep} \cdot \overrightarrow{p}_{lq}}{|\overrightarrow{p}_{lep}| \cdot |\overrightarrow{p}_{lq}|}. \tag{4.8}$$

Remark on the usability of $\cos\theta$ for this analysis

Unfortunately it turned out, that the spin correlation seems to be not well modelled in the W+jets MC samples generated with MadGraph, which lead to a bad shape agreement between the stack of generated samples and the data distribution for $\cos\theta$. For the time being, this issue was not resolved, which made it necessary to exclude $\cos\theta$ from the list of potential NeuroBayes input variables. However tests have shown, that the loss in significance when removing $\cos\theta$ from the NeuroBayes input variables, is below 1 %.

For the time being this issue is under investigation and future measurements a different generator might be used for these samples.

4.6. Data processing software

The original files as they are distributed in GRID sites are ROOT files, which store all information in the so called **E**vent **D**ata **M**odel. It stores all relevant information about the CMS events in several C++ classes. There are different data tiers for CMS samples [39].

FEVT contains the full event information. Also the raw detector readout may be stored in these files.

RECO contains only the reconstructed event objects. RAW information are dropped, since they are not needed after the event reconstruction anymore and consume a lot of storage space.

AOD (Analysis Object Data) is the refined version of RECO, which only contains those information which are essentially needed for analysis. The samples used for this thesis were derived from this data tier.

The simulated MC events are of data tier AODSIM, which means, additional to the AOD objects also information about the event simulation in the generator, showering and detector simulation are included. This is necessary for matching reconstructed objects to generated particles and measuring identification efficiencies, e.g. for b tagging.

4.6.1. (S)KITA

The official AOD(SIM) samples are processed in the GRID using the `KITA` framework. This framework utilises its own set of data objects to represent the event content. The development was a common effort of several analysts in Karlsruhe in order to have a shared data format for t quark analysis which requires less storage space than AOD samples while containing all necessary information for these kind of analysis.

The actual event reconstruction of single top candidates is then done in the `SKITA` frameworks which takes `KITA` files as input and writes out flat nTuples as well as histograms of several control variables. Also all the necessary event weights are calculated in this step and the final event selection cuts are applied. The flat nTuples are the basis for all NeuroBayes trainings and the production of all histograms which are used as templates for the statistical inference.

4.7. GRID-Computing

As it was already stated in section 3.2.7, the enormous amount of more than 15 Petabyte of data, which are produced at the LHC by all four detectors per year, cannot be stored and processed at one single computing centre with today's technology. Instead all the data produced is distributed around the world over the LHC Computing Grid (LCG), which is a hierarchical network.

From the data delivered by the detectors, a primary backup is stored at the CERN computing centre (Tier-0). After a first processing step the data is copied to the ten Tier-1 computing centres around the world, where every Tier-1 should get only a fraction of the whole data. Altogether the Tier-1 centres should mirror the whole data taken. These computing centre have to be always available and provide sufficient amount of network bandwidth, storage and computing capacities. Their task is to reconstruct the events and make the data available to Tier-2 centres. A pictogram of this tier model is shown in figure 4.2.

The advantages of such a tier model are the following:

- Redundantly mirroring the data on the Tier-1 centres reduces the risk of data loss.

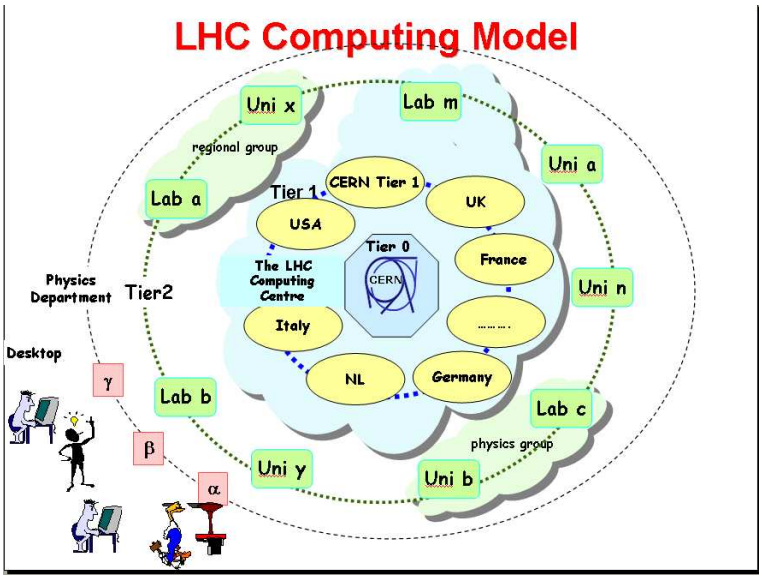


Figure 4.2.: The LCG tier-model. The Tier-1 centre for Germany resides in Karlsruhe (GridKa)

- The user can send his program to the data, instead of downloading the data to his own computer. Thus one does not need to have the storage capacity for all the events.
- The load balancing between all computing farms around the world optimally uses the available capacity and reduces the waiting time for the user.

Originally it was foreseen, that the individual scientist will access data only by sending GRID-jobs to Tier-2 centres and copying the slimmed and skimmed output to his local Tier-3 computing centres where he can run his physics analysis. However these restrictions where a little bit loosened. German grid users can login in interactively to the National Analysis Facility (NAF) at DESY in Hamburg and use GRID resources via the local batch system and have direct access to Tier-2 data residing at DESY. The same is possible in Aachen, where another big German Tier-2 centre is hosted. With some more restrictions it is even possible to submit user jobs to the German Tier-1 centre at GridKa and process data which are stored on disk. This was allowed in order to use computing resources which would otherwise be idling and hence use them more efficiently.

5. Tools and techniques

In this chapter several statistical methods and their software implementations are introduced. They play an important role for this analysis during the measurement of the cross section of single top quark t-channel events.

First the concept of multivariate analysis (MVA) is presented in section 5.1. The NeuroBayes package [60, 61] is used as MVA tool in this analysis and is described in section 5.1.4. It is utilised to combine many variables into one discriminator which is used in the statistical inference to produce fit templates for all processes.

In the second section template fitting procedures are discussed. They are implemented in the statistics package `theta` [87], which is used for fitting the different process templates to data and derive the cross section taking uncertainties into account.

5.1. Multivariate analysis

Multivariate analysis in general means the concept of considering several observables at a time and drawing conclusions about some unknown other variable. This usually includes the utilisation of statistical methods like probability density and likelihood functions to quantify the variable of interest.

5.1.1. Definition

In high energy physics MVA methods are most commonly used to classify objects, i.e. to make a decision based on measured quantities whether the object is of type A or B. Examples would be a jet is a b jet or a light quark jet [84], or whether an event is a single top t-channel event or not.

From the mathematical point of view a classifying MVA is a transformation of a vector to a number,

$$f(\overrightarrow{x} \in \mathbb{R}^n) \rightarrow y \in \mathbb{R}. \tag{5.1}$$

There are various types of algorithms implementing this behaviour, like (Boosted) Decision Trees, Support Vector Machines and Neural Networks. But following the above definition even the rather simple track counting algorithms (see section 4.4) are multivariate since they use the information of several tracks simultaneously. In the following the focus is on neural networks, since the core functionality of NeuroBayes is a neural network, which is utilised in this analysis.

Neural networks have been used in high energy physics already for decades, like in the `BSAURUS`[14] package. But they have the bad reputation of consuming a lot of computing time during the training phase and being very at risk of learning insignificant fluctuations from the training sample. This effect is called "over-training" and can happen to any complex machine learning algorithm. Both issues have been solved in NeuroBayes [61].

5.1.2. Artificial neural networks

Artificial neural networks try to model the working principle of biological neural networks like the brain. The key ingredient of such networks are the neurons, which follow a rather simple working principle. A neuron retrieves n input values and yields one output value. The output value is proportional to the sum of the inputs but usually not linearly dependent. Instead each input value is weighted and the sum is processed in an activation function (figure 5.1).

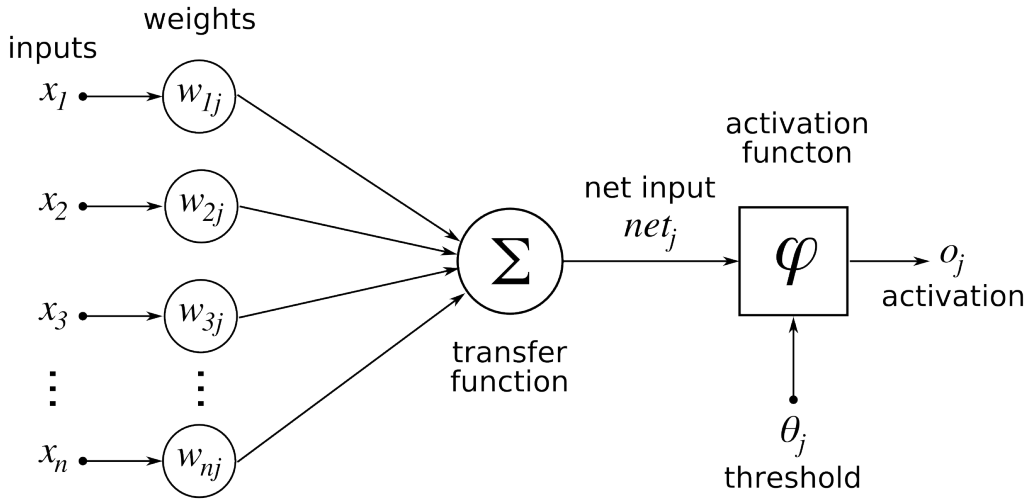


Figure 5.1.: Schematic view of an artificial neuron. The weighted sum of input values is processed in the activation function, which yields the neuron output. [36]

The activation function can be any monotonically increasing function. In practise the sigmoid function is often used:

$$\mathrm{sig}(x) = \frac{1}{1 + e^{-t \cdot x}} - 1. \tag{5.2}$$

This function has the good features of being two times continuously differentiable, which becomes important during the training with back-propagation in which the input weights are adjusted, as is explained in section 5.1.3. Values from the range $[-\infty, \infty]$ are mapped to $[-1, 1]$ but still the relation is almost linear in the region around zero (figure 5.2). Together with the preprocessing, which is explained in section 5.1.4, this leads to a strong response for the most populated input value range.

These artificial neurons alone do not provide much intelligence. However by connecting several neurons in a network an algorithm is retrieved that can model any $\mathbb{R}^n \to \mathbb{R}$ relation, if a sufficient amount of neurons and connections is available.

These networks usually consist of several (at least three) layers of neurons where each neuron from layer i is connected to each neuron in layer $i + 1$ (figure 5.3).

Like in the biological neural network, the free parameters on ANNs are the weights of the connecting edges. These are adjusted during the training process to yield the desired network behaviour as will be explained in the next section.

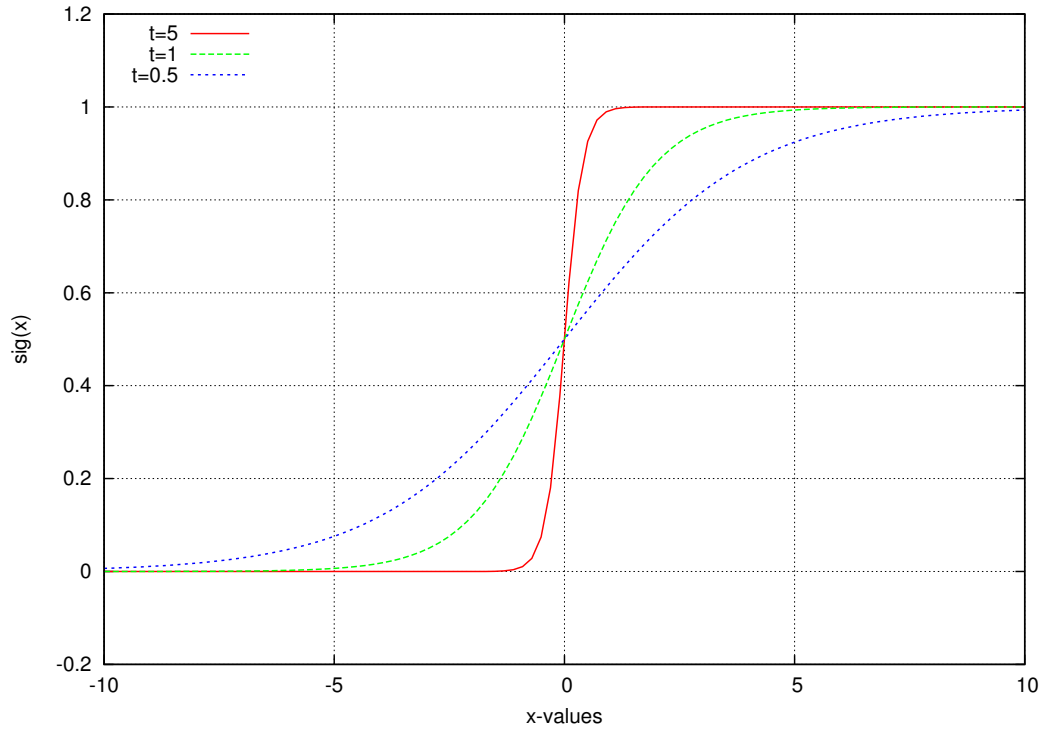


Figure 5.2.: The sigmoid function for the different values of parameters t.

5.1.3. Training of neural networks

Neural networks and other MVA methods parametrise a transformation function, which is not known a priori, by using a complex, non-linear method to weight and transform the input values. Before these tools can be used for the actual data classification, they need to be trained. This training is usually done on simulated or historical samples, where the event type is known. Many simulated events are needed for the training in order to scan the phase space and let the MVA method adjust its internal parameters to yield the correct output. The correctness of an MVA output is quantified in a loss function E that "compares" the prediction value y to the target value t and is summed up over all events indexed by i

$$\mathrm{err}(y) = \sum_i E(y_i, t_i) \ , \tag{5.3}$$

where $E(y,t)$ could be e.g. the quadratic loss function

$$E(y,t) = (y-t)^2 \tag{5.4}$$

or the entropic loss function

$$E(y,t) = \log\left(\frac{1+y\cdot t}{2}\right) \ . \tag{5.5}$$

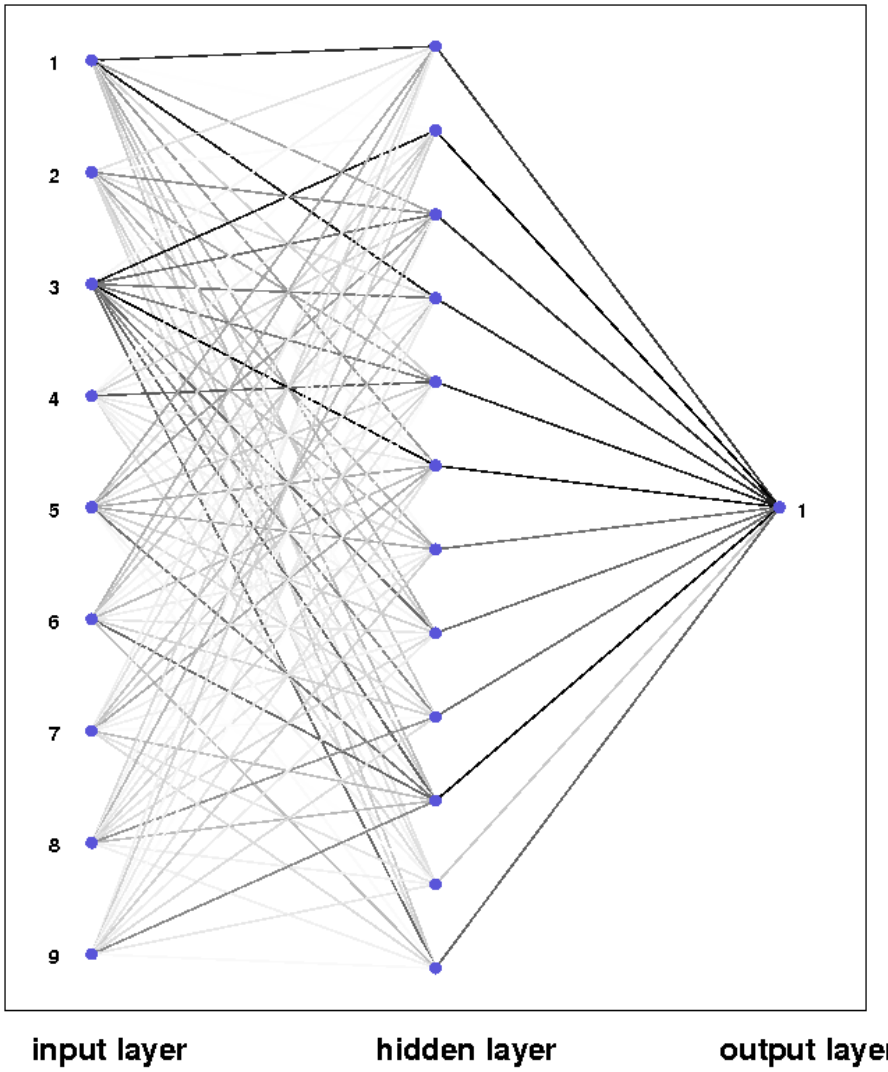


Figure 5.3.: Example of a feed forward neural network. The thickness of the connecting lines between the neurons is proportional to the weight on this edge.

Finding the minimum of (5.3) can be difficult since y usually depends non-linearly on a lot of free MVA parameters $\vec{w}$. Thus $\mathrm{err}(y(\vec{w}))$ is a hyperplane in a high dimensional space of which one needs to find a minimum. Although any minimisation algorithm could be used to solve this problem, two special algorithms are presented, that are well suited to optimise neural network parameters.

Neural network training - back propagation and BFGS training

Neural networks are usually trained using the back propagation algorithm [97]. As the name suggests this algorithm propagates the result of the network backwards in order to

adjust the connection weight between the neurons. In each step the weights are updated according to the current value in the loss function.

In NeuroBayes there is additionally the BroydenFletcherGoldfarbShannon (BFGS) method [25] available for optimising the network weights. It is a general algorithm for solving large non-linear optimisation problems and works in a similar way as the Newton algorithm does to seek a stationary point of the error function. This algorithm requires that the error function is two times differentiable, which is the case for the above presented loss functions and the sigmoid function.

Training with event weights

During the training processes many events are processed. There are several use cases where it is necessary to change the impact of an event in the training.

- In order to achieve a well calibrated network it is important to use a training sample where the relative normalisation of all processes is similar to what is expected in real data. Since generic MC samples, in which all possible processes are generated, are not available for LHC collisions, one has to blend the training sample from several special MC samples (also see section 6.1.1). Since the number of events in each sample is not normalised to the integrated luminosity $\mathcal{L}$ and cross section σ, this has to be done by weighting all events with their expected cross section:

  $$w = \frac{\sigma}{\mathrm{N_{events}}} \cdot \mathcal{L}.$$

  It is convenient to postpone the scaling by luminosity to the last step in the analysis, since this might change once in a while due to new information or a different selection of data samples.

- Studies might show that events need to be reweigted in MC depending on certain properties like missing transverse energy, number of b tagged jets, etc. to achieve a better data/MC agreement in observable distributions.

- MC events are generated with a flat distribution in pileup to be usable for several pileup scenarios. Thus MC events need to be reweighted depending on their number of pileup events to fit a specific data sample.

- The signal process may be very rare even after applying cuts to enrich it. But MVAs will also learn the prior probability of signal events. Thus calibrated MVAs like NeuroBayes yield quite different discriminator shapes depending on the signal fraction in the training. It may be useful to artificially enhance the signal ratio in analysis like this to derive a certain discriminator behaviour.

  One way to derive the signal and background weights to shift the signal fraction from p_1 to p_2 is the following. Given the number of signal events N_S and the number of background events N_B in the training sample[1], the signal fraction is $p_1 = \frac{N_S}{N_S+N_B}$.

[1] N_S and N_B may already be the weighted sum of events

To achieve the desired signal rate $p_2 = \frac{w_S N_S}{w_S N_S + w_B N_B}$ the ansatz $w_S = \frac{1}{w_B}$ can be used and the weight which has to be applied for each signal event turns out as

$$w_S = \sqrt{\frac{\frac{1}{p_1} - 1}{\frac{1}{p_2} - 1}}. \tag{5.6}$$

So there are several sources which will introduce a weight for each event, which we need to take into account during the training. This is done by modifying the calculation of the loss function such, that each event is weighted individually.

$$\mathrm{err}(y) = \sum_i w_i \cdot g(y_i, t_i)$$

Modern MVA methods like NeuroBayes are able on the one hand to use the weights of each event in the loss function, while on the other hand also considering the real given statistics. Without this it would make a difference on the result if all events would be weighted up by a common factor, since it might look like the number of events has increased and statical fluctuations are taken too much into account.

5.1.4. NeuroBayes

NeuroBayes [60, 61] is an advanced implementation of an artificial neural network.

It uses a three layer feed forward neural network and features a powerful and adjustable preprocessing of input variables. NeuroBayes cannot only be used for event classification but also to predict full probability densities of measurable values [60].

Its predecessor was used in the Delphi experiment at LEP in the `BSAURUS` package [14]. Nowadays NeuroBayes is further developed by the company <phi-t>, which is a spin-off of the University of Karlsruhe.

One of the key ingredients of NeuroBayes is the versatile preprocessing, which is presented in the following.

NeuroBayes preprocessing

The activation function in each NeuroBayes neuron is a sigmoid (see equation (5.2) and figure 5.2). This function has its strongest response in the region around zero. Therefore it is favourable to transform all input values in a way, that the values entering the neuron are in this region and that target events have on average higher values than non-target events. In NeuroBayes this is most commonly achieved by using the following preprocessing steps, which can be applied to continuous variables;

Flatten the inclusive distribution by using a histogram with non-equidistant bins. Each bin has than the same number of events and only the signal purity (= signal / background ratio) is used in the following. This procedure automatically constrains insignificant outliers to the lowest/highest bin. On the other hand regions where many events with similar values reside, are distributed over a wider numerical space.

Smoothen the purity distribution via a spline fit. This takes into account the values of neighbouring bins and further eliminates fluctuations.

Transform the spline values into a distribution with mean zero and width one. This is done to meet the sigmoid at its optimal working point.

The transformed input values are already well separated, outlier corrected and shifted into an interval around zero, which is convenient for the sigmoid function. The actual training process now can focus on learning inter-variable correlations and other effects to achieve a good signal and background separation.

5.2. Parameter estimation

The calculation of the single top quark production cross section is based on fitting the neural network discriminator templates for all the relevant processes to the according data distribution. This is explained in detail in section 6.5. In the following the technical and mathematical point of view of such a fit process is discussed. The likelihood function as well as algorithms to find the optimal parameter set by maximising a likelihood function, are introduced.

5.2.1. The likelihood function

Likelihood functions are in this analysis used to estimate the optimal set of parameters for a model, that describe the measured data. A "model" in this sense is a set of probability density functions (pdf) which describe the distribution of a measurable variable in a certain process k. These pdfs could either be theoretical functions or, as in our case, binned templates generated from simulated events. "Data", on the other hand, is a sample of events on which the variable was measured. The pdfs are used as templates and we are interested in that linear combination of these templates, that approximates the data distribution best. Hence this is a counting experiment in all of the template bins separately and all event sums in the bins have to be optimised simultaneously.

The sum of events over n_p processes in bin i, μ_i, is given by

$$\mu_i(\beta_1, \ldots, \beta_{n_p}) = \sum_{k=1}^{n_p} \beta_k \nu_k^{\text{exp}} \cdot \alpha_{k,i} \ , \tag{5.7}$$

where the template scale parameters β_k are normalised to the expected number of events for each process k.

$$\beta_k = \frac{\nu_k}{\nu_k^{\text{exp}}} \tag{5.8}$$

The fraction of events of process k that is expected to be found in bin i of the templates is denoted by $\alpha_{k,i}$. For each process the α's satisfy the normalisation condition

$$\sum_{i=1}^{n_p} \alpha_{k,i} = 1 \ , \tag{5.9}$$

where n_b is the number of bins in the templates.

In a counting experiment like this, Poisson statistic can be assumed in each of the bins. The general probability to observe N events with a given expectation value μ is given by

$$P_\mu(N) = \frac{\mu^N}{N!} e^{-\mu}. \tag{5.10}$$

In this case, with N_i selected data events per bin i and μ_i expected events as given by equation (5.7), equation (5.10) becomes

$$P_{\mu_i}(N_i) = \frac{\mu_i(\beta_1, \ldots, \beta_{n_p})^{N_i}}{N_i!} e^{-\mu_i(\beta_1, \ldots, \beta_{n_p})} \tag{5.11}$$

as the Poisson probability to observe this.

The full likelihood function is given by the product of all likelihoods for each bin.

$$L(\beta_1, \ldots, \beta_N) := \prod_{i=1}^{n_b} \frac{(\mu_i(\beta_1, \ldots, \beta_N))^{n_i}\, e^{-\mu_i}(\beta_1, \ldots, \beta_N)}{n_i!} \cdot \prod_{j=2}^{n_p} G(\beta_j|\sigma_j). \tag{5.12}$$

The extra term multiplying over $G(\beta_j|\sigma_j)$ adds prior knowledge about the scale parameters into the likelihood function. There is a certain degree of believe that the initial normalisation of the pdfs, i.e. the expected production cross section for the processes, is correct to some extend. Therefore this extra term is introduced, which is a Gauss function with mean 1 and width σ_j.

$$G(\beta_j|\sigma_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \cdot \exp\left(-\frac{(\beta_j - 1)^2}{2\sigma_j^2}\right) \tag{5.13}$$

The assumed widths are different for each process and are listed in section 6.4. This term makes it less likely for a scale parameter to deviate significantly from 1.0 during the optimisation.

The so defined likelihood function (5.12) can now be either used to estimate the parameter set β_k that maximises this function, or to calculate the marginalised posterior distribution that only depends on the signal scale parameter β_s anymore. Both are explained after a short remark on the technical implementation of likelihood functions.

Remark on numerical precision of likelihood functions

The likelihood function as given in equation (5.12) is neither well suited to be analytically maximised, nor to be implemented as software in todays computers. The reason for the first is just, that during the analytical maximisation the calculation of a product with many terms tends to get complicated.

On the other hand all factors in the likelihood function are probabilities and are therefore per definition in the range [0,1] and their product is in the same value range. In todays computers, with their representation of floating point numbers as exponent and mantissa, this is problematic for the following reason:

The smallest representable positive number with double precision [2] is $\approx 10^{-308}$. With many small probabilities in the likelihood function one might easily get below this threshold and the product will become zero.

Both issues, the analytical and the numerical, can be solved by taking the logarithm of the likelihood function. This has the effect that the value range is transformed from [0,1] to [-∞,0] and the product turns into a sum. It is numerically calculable at lower computing costs than a product and also finding the derivative means less effort.

Since historically optimisation algorithms were designed to minimise the cost function for a given parameter set, additionally the sign of the likelihood value is flipped. Thus in practise the likelihood function is not really maximised, but the negative log likelihood function is minimised. Still the result is the same in the end.

[2] Derived on a current 64bit PC. Other machines might yield different results.

5.2.2. Parameter estimation using likelihood functions

The previously discussed likelihood function is a measure for how well a given parameter set reflects data, the higher the likelihood the better is data described by the templates k scaled with the parameters β_k. Some algorithm is needed which scans the parameter space and finds the best fitting point in this space. This is in principle the same problem as training a neural network and finding the best set of weights there (see section 5.1.3). The main difference is, that here also the uncertainty/confidence interval for the estimated value of each parameter β_k has to be taken into account. This can be achieved by using the maximum likelihood estimation, which is done in this analysis for estimating the QCD contribution and to estimate the impact of some systematic uncertainties on the final result. This will yield the best fitting value for each parameter and also a confidence interval can be derived for each parameter individually.

For the actual cross section measurement one is not interested in the values of all nuisance parameters β_k, i.e. all model parameters except the signal strength parameter β_s. Also its confidence interval should include the uncertainties of the nuisance parameters. For this another approach, the full Bayesian marginalisation, is better suited and is explained after the maximum likelihood estimation.

Maximum likelihood estimation

Estimating the optimal set of parameters with respect to a given data sample can be done by maximising the likelihood function. In some cases this can be done analytically by calculating the first and second derivative and searching for roots. But in most cases numerical algorithms are needed to find a maximum. As stated in section 5.1.3,there are elaborated algorithms available which will numerically minimise any given loss function with respect to some parameters. Several of these algorithms are available in the Minuit [80] package, which is shipped as part of ROOT [32].

Such a minimizer is well suited for finding the optimal parameter set for fitting the QCD contribution in this analysis, as is described in section 6.2.7.

For the cross section measurement one is interested in only one parameter but would like to have all uncertainties included into its probability density function. This could be done with the maximum likelihood function, but there are other algorithms which have shown to converge faster and be more reliable.

5.2.3. Calculating Bayesian quantile with full marginalisation

The likelihood function of a fit model depends on many parameters. Usually only one of them is of interest in the end, the scale factor for signal templates β_s. All the other parameters are called nuisance parameters and are not directly needed for the final result, but their prior uncertainty has to be taken into account when calculating the confidence interval for β_s. Such an uncertainty could be for instance the cross section of a background process and its estimated precision, taken from earlier measurements. In this analysis this is achieved by integrating (marginalising) over all nuisance parameters. The so retrieved posterior distribution can then be used to determine the best fitting β_s as well as the confidence interval.

The definition of the posterior distribution is given in the next section.

Bayesian posterior of the likelihood function

With given data d, signal scale factor β_s and background parameters β_b, the general Bayesian posterior is given by

$$p(\beta_s, \beta_b|d) = p(d|\beta_s, \beta_b) \cdot \frac{p(\beta_s, \beta_b)}{p(d)}. \tag{5.14}$$

The different terms can be interpreted as follows:

$p(\beta_s, \beta_b|d)$ is the probability density for our model parameters β_s and β_b *after* considering the data d, therefore called the posterior.

$p(d|\beta_s, \beta_b)$ is the likelihood. It is the probability to observe the data d given the model parameters β.

$p(\beta_s, \beta_b)$ is the prior. Before looking at the dataset d there is some prior knowledge for all model parameters, e.g. derived from earlier measurements. Thus it is known to some extend, what the most probable value and the variation width for all nuisance parameters are. For example some background process has a certain cross section and therefore its template is normalised to the expected number of events. Hence its most probable scale parameter is one, but some variations depending on the precision with which the cross section is known are still allowed.

$p(d)$ is the probability for the observed data. This is usually not explicitly calculable a priori which makes the posterior calculation difficult. But it is independent from any model parameter β and cancels out when calculating the ratio for two different models, which are used in the Markov Chain Monte Carlo integration.

The full Bayesian posterior in equation (5.14) still depends on the model parameters β_b for the background processes, in which we are not interested for our final result. Instead we want to have the marginal posterior which only depends on β_s. This is achieved by integrating over all nuisance parameters β_b.

$$p(\beta_s|d) = \int p(\beta_s, \beta_b|d)\ \mathrm{d}\beta_b \tag{5.15}$$

The actual extraction of the marginal posterior equation (5.15) can be technically very difficult, if many nuisance parameters with several prior pdfs have to be integrated over. It can be done numerically using *Markov Chain Monte Carlo* which is explained next.

Markov Chain Monte Carlo for integration in multiple dimensions

Computing high dimensional integrals can be very difficult and often has to be done numerically. This can be done by using Monte Carlo integration. We will focus on Marcov Chain Monte Carlo (MCMC) in the following, which is one of several MC integration methods. A nice introduction to MCMC integration for Bayesian posteriors and the Metropolis Hastings algorithm can be found at [105] and is briefly summarised in the following.

If one wants to compute for example an integral of a function $h(x)$ and can decompose this function into the product of a function $f(x)$ and a probability density function $q(x)$ the integral can be expressed as the expectation value of $f(x)$ over $q(x)$

$$\int_a^b h(x)dx = \int_a^b f(x)q(x)dx = E_{q(x)}\left[f(x)\right]. \tag{5.16}$$

Thus by drawing a large sample of random numbers from density $q(x)$ the integral in the interval $[a, b]$ over $h(x)$ can be approximated by

$$\int_a^b h(x)dx \simeq \frac{1}{n}\sum_{i=1}^{n} f(x_i). \tag{5.17}$$

The difficult part is shifted from calculating the integral to drawing n random numbers from a potentially complicated pdf $q(x)$.

Applying this to equation (5.15) we get

$$\int p(\beta_s, \beta_b|d)\ \mathrm{d}\beta_b = E_{p(\beta_s,\beta_b|d)}\left[1\right] \tag{5.18}$$

Splitting up $p(\beta_s, \beta_b|d)$ into a pdf and another function is not possible in this case, because we do not know the actual value of $p(d)$ which would be needed to normalise the pdf. Hence the full vector of β values is drawn according to the posterior distribution and no further function has to be evaluated. The distribution of β_s then corresponds to the marginal distribution of the posterior.

Also drawing β_b's which are distributed according to $p(\beta_s, \beta_b|d)$ is complicated due to the lack of knowledge about $p(d)$, but it is possible using the Metropolis (Hastings) algorithm, which is a Markov Chain process.

A Markov Chain is a random walk process in which the next state only depends on the current state but not on previous ones and is therefore a memoryless process. Suppose we want to draw random numbers x distributed according to a probability density function $p(x) = f(x)/c$ but we do not know the normalisation constant c. $p(x)$ in this example corresponds to $p(\beta_s, \beta_b|d)$ while $f(x)$ is obviously $p(d|\beta_s, \beta_b) \cdot p(\beta_s, \beta_b)$ while c is $p(d)$. The Metropolis algorithm then uses the following steps to wander around in parameter space.

1. Select a starting point x_0 at which $f(x_0) > 0$.
2. Select another point x' and also calculate $f(x')$.
3. Calculate the ratio between the function values at x and x', which is also the pdf's ratio, since the unknown probability $p(d)$ in equation (5.14) cancels out.

$$\alpha = \frac{f(x')}{f(x)} = \frac{q(x')}{q(x)} \tag{5.19}$$

4. The actual next point of the chain,x', is not selected in all cases. Only if $\alpha > 1$ the new point is selected. If the probability decreases, i.e. $p(x') < p(x)$, the new point is selected with probability α and discarded in the other cases. Therefore a point x with a high $p(x)$ can be returned several times in a row until a new x is selected.

By repeating steps 2 – 4 several times, the selected x's are distributed according to the pdf $p(x)$. Though a certain amount of iterations have to be made until the drawn x's do not depend on the starting point x_0 anymore. After this "burn-in" period the algorithm reaches a stationary distribution. One is interested in keeping the burn-in period as short as possible, since it is just wasted computing-time. Choosing a smart starting point and also an elaborate proposal function for new points in step 2 is crucial for a fast convergence to the stationary distribution.

The above described procedure can be used to derive the marginal posterior distribution as defined in equation (5.15). With this the parameters which have the most probable posterior value can be identified and also the central interval which covers a certain credibility can be found.

The algorithms described above are implemented in the `theta` framework, which is briefly explained next.

Accounting for systematic shape uncertainties - template morphing

Some of the systematic uncertainties in the model are given by alternative templates, that have a different shape and potentially also a different rate as the nominal sample. These systematic samples were derived from events generated with different MC generator settings or by modifying or reweighting the events in nominal samples. In order to estimate the impact of these uncertainties the fitting or marginalisation procedure has to be done several times with different sets of inter- and extrapolated templates.

The process of generating such a systematically modified template is called "morphing" and can be done as follows:

1. Pick a set of templates consisting of the nominal template and one for the 1σ downwards variation of one systematic as well as one for the 1σ upwards variation.

2. Compute the expected number of events in each bin of these three templates, μ_0, μ_{down} and μ_{up}.

3. Find an inter-/extrapolation function through these three values. `theta` uses cubic polynomials to interpolate between the three points and a linear function to extrapolate outside. It is also required that the function shall be continuously differentiable at μ_{down} and μ_{up}.

4. Dice a μ' from a normal distribution with width 1 and mean 0.

5. Calculate the expected bin content for μ' with the inter-/extrapolation function and compute the difference to the nominal bin content Δ_i for each bin i.

The above steps are done for all processes k and each systematic uncertainty, each with an individual μ'. The derived Δ's are added up to form the morphed template, which can than be used in the statistic inference.

By only varying one systematic uncertainty at a time, one can study their individual impact on the final result, as is explained in section 6.5.1.

Credible interval

While the full Bayesian posterior, as given in equation (5.14), after marginalisation (see equation (5.15)) is the final result of our statistical inference, it is convenient to define certain values, derived from the posterior, to sum up the result. This is the central value of the distribution, which is usually the probable value, i.e. the one with the highest likelihood, but could also be the mean or the median of the Bayesian posterior. Also the credible level central interval shall always be given with a measured value since it expresses the precision with which the central value is derived. While any interval around the central value could be used, in the most cases the 68 %, 95 % or 99 % central intervals are quoted, since they correspond to the 1σ, 2σ and 3σ intervals if the posterior is a Gaussian distribution.

A credible interval of, e.g. 68 % can directly be interpreted as the value range in which the true value for β_s lies with 68 % probability. This is in contrast to the related confidence interval, which is used in frequentists statistics. There the interpretation is that in 68 % of a large enough sample of measurements of β_s, the outcome is within the confidence interval in 68% of the cases. But the probability for the true value of β_s to be inside the interval is either 1 or 0.

In this analysis, the central value is given as the left sided 50 % Bayesian quantile of the marginalised posterior, with the 16 % and 84 % credible interval borders as asymmetric lower and upper boundaries on the central 68 % credible interval.

5.2.4. `theta` - template fitting framework and statistical closure

According to [86] is `theta` "a framework for template-based statistical modelling and inference, focussing on problems in high-energy physics." An introductory manual can be found in [87]. It is a versatile statistics tool which is used for retrieving event yields at several points in this analysis.

1. Estimate the amount of QCD events by fitting the missing transverse energy (MET) in the electron channel and the transverse W mass (MTW) in the muon channel.

2. Retrieve the fraction of signal events and performing a full Bayesian marginalisation, taking all systematic uncertainties into account.

`theta` uses templates for building statistical models. These templates are expected to sum up to the data distribution. Thus for each physics process involved in data, an according (MC) template is needed. Systematic uncertainties are taken into account either as rate uncertainties, in which the templates normalisation is varied within a certain range for β_k as explained above. Or they are given as shape uncertainties, which require two additional templates for each process and each uncertainty. One describing the upwards fluctuation of this uncertainty and its impact on the template and one the downwards fluctuation. Intermediate distributions are then diced within the given uncertainty ranges, which are used to model the impact of a certain uncertainty on the fit result.

6. Analysis - measuring the single top t-channel cross section

So far, all the necessary parts have been presented which are used to measure the single top t-channel cross section. In the following they will be assembled to the full analysis work flow.

The first step in this work flow is to select the needed data events from CMS data, as well as the necessary generated Monte Carlo samples. In a second step selection cuts are applied on all events, to yield samples which are enriched in the desired single top t-channel events. The generated samples need additional corrections and reweighting to be in accordance with the selected data. These two steps are presented in sections 6.1 and 6.2.

The generated nominal MC samples are used to train NeuroBayes to distinguish between the single top t-channel events and the other processes, as is described in section 6.3. NeuroBayes is than used to produce discriminator templates for measured data, as well as for the generated MC.

There are several sorts of systematic uncertainties which need to be taken into account. They are discussed in section 6.4. In general they lead to alternative or reweighted samples that have to be used in the statistical inference. An individual pair of NeuroBayes discriminator templates is generated for most of the systematic uncertainties, modelling the up- and the downwards shift of the uncertainty.

All of these templates are used to build a fit model which is used in `theta` to perform a Bayesian marginalisation. This marginalisation includes most of the systematic uncertainties and yields the most probable signal yield as 50 % quantile of the Bayesian posterior distribution. This signal yield is then used to derive the single top t-channel cross section in section 6.5.

6.1. Datasets

This section covers the utilised samples. These are on the one hand the generated Monte Carlo samples, presented in section 6.1.1. On the other hand there is the list of primary data sets that are selected from measured CMS data, which are listed in section 6.1.2. Also the merging of the exclusive W+jets samples is presented here.

6.1.1. Generated Monte Carlo samples

For each relevant process in the selected data region, a particular Monte Carlo sample is used. All of them were officially produced for the CMS group during the `Summer11` production period. The data sets used in this analysis are given in table 6.1.

Table 6.1.: List of all used generated Monte Carlo datasets. All cross sections, taken from [100], are given for next to leading order calculations, except for the exclusive W+jets samples, were only leading order precision (LO) is available. The first column lists the abbreviated name for the data set which is used in the following. If not specified differently, the t and $\bar{t}$ are used together, as one sample.

dataset name	official dataset name	# events	cross section[pb]
t-channel(_t)	/T_TuneZ2_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	3900171	41.92
t-channel(_tbar)	/Tbar_TuneZ2_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1944826	22.65
s-channel(_t)	/T_TuneZ2_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	259971	3.19
s-channel(_tbar)	/Tbar_TuneZ2_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	137980	3.19
tW-channel_dr(_t)	/T_TuneZ2_tW-channel-DR_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	814390	7.87
tW-channel_dr(_tbar)	/Tbar_TuneZ2_tW-channel-DR_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	809984	7.87
$t\bar{t}$	/TTJets_TuneZ2_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v2/AODSIM	3701947	157.5
W+jets	/WJetsToLNu_TuneZ2_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	81352581	31314.0
W+2jets	/W2Jets_TuneZ2_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	25374787	1435.0 (LO)
W+3jets	/W3Jets_TuneZ2_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	7685944	304.2 (LO)
W+4jets	/W4Jets_TuneZ2_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	13133738	172.6 (LO)
Z+jets	/DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	36277961	3048.0
diboson (ww)	/WW_TuneZ2_7TeV_pythia6_tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	4225916	43.0
diboson (wz)	/WZ_TuneZ2_7TeV_pythia6_tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	4265243	18.2
diboson (zz)	/ZZ_TuneZ2_7TeV_pythia6_tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	4187885	5.9

Some of the systematic uncertainties, that are discussed in section 6.4, are given by alternative samples, which are modified on generator level according to the systematic uncertainty they model. The list of samples, representing a systematic uncertainty, is given in table 6.2.

W+jets sample blending

The inclusive W+jets sample WJetsToLNu_TuneZ2_7TeV-madgraph-tauola has relatively few events, considering that most of these events have only one or two additional jets. This makes it necessary to use the exclusive W+2,3,4 jets samples instead. Since there is no matching W+0 jets or W+1 jets samples available, this type of events have to be extracted from the inclusive sample by cutting on the number of final state partons on generator level.

A proper weighting is needed to merge these samples into one. Since for the exclusive samples only the leading order (LO) cross section is available, while all events should be weighted to the next to leading (NLO) cross section, the blending has to be made on LO cross sections and afterwards applying a k factor of $\sigma_{NLO}/\sigma_{LO} = 31314.0/27770.0$ on all events.

The retrieved sample has a sufficient number of events, even for categories were three or four jets are required, as is listed in tables 6.8 and 6.9.

6.1.2. CMS data events

The datasets for this analysis were collected during the so called `2011A` run, which was recorded between spring and late august 2011. The average instantaneous luminosity, and consequently also the event pile-up, was lower during this run, than in the `2011B` run which was recorded until end of november. The official JSON files (i.e. list of runs which are approved to be used for analysis) were used. Those runs, in which the used triggers were prescaled were removed from the JSON files. This was done to not have to take measures for the prescale uncertainty. However the excluded runs correspond only to a few pb^{-1}. The corrected JSON files were used to select the events and to calculate the integrated luminosity.

Muon channel

In the muon channel two different primary datasets are used. They are listed in table 6.3 together with their run ranges.

For the whole run range the `HLT_IsoMu17` trigger (see section 6.2.1) could be used and the quoted integrated luminosities were calculated using the `pixelLumiCalc` tool.

Electron channel

For the electron channel, three different datasets are used, which are listed in table 6.4

This was necessary since the SingleElectron trigger was prescaled in runs later than 163869, which is explained in more detail in section 6.2.1.

Table 6.2.: List of all Monte Carlo datasets, generated with systematic uncertainties. The cross sections[100] are the same as in the corresponding nominal dataset in table 6.1

dataset name	official dataset name	# events
t-channel(_t)_scaledown	/T_TuneZ2_scaledown_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1041924
t-channel(_t)_scaleup	/T_TuneZ2_scaleup_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1032197
t-channel(_tbar)_scaleup	/Tbar_TuneZ2_scaleup_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	565520
t-channel(_tbar)_scaledown	/Tbar_TuneZ2_scaledown_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	565454
s-channel(_t)_scaleup	/T_TuneZ2_scaleup_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	285972
s-channel(_t)_scaledown	/T_TuneZ2_scaledown_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	285602
s-channel(_tbar)_scaleup	/Tbar_TuneZ2_scaleup_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	153981
s-channel(_tbar)_scaledown	/Tbar_TuneZ2_scaledown_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	153971
tW-channel(_dr_t)_scaleup	/T_TuneZ2_scaleup_tW-channel-DR_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	437736
tW-channel(_dr_t)_scaledown	/T_TuneZ2_scaledown_tW-channel-DR_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	437819
tW-channel(_dr_tbar)_scaleup	/Tbar_TuneZ2_scaleup_tW-channel-DR_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	437798
tW-channel(_dr_tbar)_scaledown	/Tbar_TuneZ2_scaledown_tW-channel-DR_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	437863
$t\bar{t}$_scaleup	/TTjets_TuneZ2_scaleup_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	930483
$t\bar{t}$_scaledown	/TTjets_TuneZ2_scaledown_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	967055
$t\bar{t}$_matchingup	/TTjets_TuneZ2_matchingup_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1065323
$t\bar{t}$_matchingdown	/TTjets_TuneZ2_matchingdown_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1062792
W+jets_scaleup	/WJetsToLNu_TuneZ2_scaleup_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	9784907
W+jets_scaledown	/WJetsToLNu_TuneZ2_scaledown_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	10022324
W+jets_matchingup	/WJetsToLNu_TuneZ2_matchingup_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	10461655
W+jets_matchingdown	/WJetsToLNu_TuneZ2_matchingdown_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	9956679
Z+jets_scaleup	/ZJetsToLL_TuneZ2_scaleup_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1593052
Z+jets_scaledown	/ZJetsToLL_TuneZ2_scaledown_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1658995
Z+jets_matchingup	/ZJetsToLL_TuneZ2_matchingup_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1641367
Z+jets_matchingdown	/ZJetsToLL_TuneZ2_matchingdown_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1615032
t-channel(_t)_tMdown	/T_TuneZ2_mass166_5_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1046360
t-channel(_t)_tMup	/T_TuneZ2_mass178_5_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	1033425
t-channel(_tbar)_tMdown	/Tbar_TuneZ2_mass166_5_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	555893
t-channel(_tbar)_tMup	/Tbar_TuneZ2_mass178_5_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	540278
s-channel(_t)_tMdown	/T_TuneZ2_mass166_5_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	268357
s-channel(_t)_tMup	/T_TuneZ2_mass178_5_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	283965
s-channel(_tbar)_tMdown	/Tbar_TuneZ2_mass166_5_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	151177
s-channel(_tbar)_tMup	/Tbar_TuneZ2_mass178_5_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	146382
tW-channel(_t)_tMdown	/T_TuneZ2_mass166_5_tW-channel_DR_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	424239
tW-channel(_t)_tMup	/T_TuneZ2_mass178_5_tW-channel_DR_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	432690
tW-channel(_tbar)_tMdown	/Tbar_TuneZ2_mass166_5_tW-channel_DR_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	428943
tW-channel(_tbar)_tMup	/Tbar_TuneZ2_mass178_5_tW-channel_DR_7TeV-powheg-tauola/Summer11-PU_S4_START42_V11-v1/AODSIM	434305
$t\bar{t}$_tMdown	/TTJets_TuneZ2_mass166_5_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v3/AODSIM	1669034
$t\bar{t}$_tMup	/TTJets_TuneZ2_mass178_5_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v3/AODSIM	1648519
Comphep_ST_tchan_ele	/TToBENu_TuneZ2_t-channel_7TeV-comphep/Summer11-PU_S4_START42_V11-v1/AODSIM	1047087
Comphep_ST_tchan_mu	/TToBMuNu_TuneZ2_t-channel_7TeV-comphep/Summer11-PU_S4_START42_V11-v1/AODSIM	1087279

Table 6.3.: Datasets used in the muon channel

Dataset	Run ranges
/SingleMu/Run2011A-May10ReReco-v1	160404-163869
/SingleMu/Run2011A-PromptReco-v4	165098-167913

Table 6.4.: Datasets used in the electron channel.

Dataset	Run ranges
/SingleElectron/Run2011A-May10ReReco-v1	160404-163869
/ElectronHad/Run2011A-PromptReco-v4	165098-167151
/ElectronHad/Run2011A-Aug05ReReco-v1	170826-172619

6.2. Event selection and weighting

Events are selected by a list of selection criteria, which all events need to fulfil that are used for the statistical inference. The cuts have been chosen to either remove background events, or to remove events in which not all necessary information are available and hence are not well re-constructable, e.g. no good primary vertex is found.

6.2.1. Trigger

Several triggers are used to select datasets with a high fraction of single top t-channel events from the early 2011 CMS data.

muon channel: In the muon decay channel of the single top events, an isolated, high energetic muon is present in the final state (see section 2.4.3). Hence the HLT_IsoMu17 trigger was used to select events from the datasets listed in table 6.3. It requires an isolated muon with a transverse momentum $p_T > 17\,\mathrm{GeV/c}$ to be present in the event. Luminosity sections in which this trigger was prescaled were excluded from the event list. In the end a dataset with an integrated luminosity of $1170.21\,\mathrm{pb}^{-1}$ was selected and used for this analysis.

electron channel: In the electron channel an isolated electron is expected in the final event state. Since the the single electron trigger was prescaled after the first 2011 runs, also different triggers have to be used. Electron-hadron cross triggers are used, which also require a b tagged jet on trigger level. This is still in accordance with the expected event signature. The single electron trigger is usable for the runs 160404 to 163869, which have an effective integrated luminosity of $215.63\,\mathrm{pb}^{-1}$ for this trigger. The electron-hadron cross triggers are used for runs 163870 to 172619. These trigger require an isolated trigger electron, at least one trigger jet with $p_T > 30\,\mathrm{GeV/c}$, and at least one jet with TCHE b tag discriminator value above the medium working

point threshold. The second electron-hadron cross trigger used in run 165970 and later has tighter requirements for the electron identification and the isolation. The used triggers and their run range are listed in table 6.5.

Table 6.5.: High Level Triggers for the electron channel

High Level Trigger	first run	last run	$\mathcal{L}\,\mathrm{pb}^{-1}$
HLT_Ele27_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT	160404	163869	215.63
HLT_Ele25_CaloIdVT_TrkIdT_CentralJet30_BTagIP	163870	165969	136.38
HLT_Ele25_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_ Central-Jet30_BTagIP	165970	172619	1208.71

Here also individual luminosity sections in which the according trigger was temporarily prescaled were excluded. The cross triggers yield $1345.088\,\mathrm{pb}^{-1}$ in the runs 163870 to 172619. Altogether there are $1560.718\,\mathrm{pb}^{-1}$ available for analysis in the electron channel. Although the integrated luminosity is higher in the electron channel, than in the muon channel, the event yield is smaller due to a worse selection efficiency.

6.2.2. Primary Vertex Cut

Each event has to have a reconstructed primary vertex. The quality criteria for such a vertex are:

- The `IsFake` flag has to be set to false. In events where no vertex can be fitted the beam spot is taken as primary vertex and this flag is set to true. These events are excluded from this analysis.
- ndof >4, where number degrees of freedom is the weighted sum of the tracks used to reconstruct the primary vertex.
- Has to be within the central detector region, i.e. $|z| < 24\,\mathrm{cm}$ and $r < 2\,\mathrm{cm}$ with respect to the nominal primary interaction point.

6.2.3. Lepton cuts

In accordance to the trigger cuts, in the offline selection also an isolated lepton required. This is done to benefit from the more precise lepton identification in the offline event reconstruction and to remove events with multiple leptons. Hence the selection is refined by requiring exactly one reconstructed lepton.

A lepton has to pass certain criteria to be accounted as such, one of which is the relative isolation. The relative isolation is defined by Particle Flow (see section 4.2.6) quantities as the ratio between the sum of the charged, neutral and photon isolation over the lepton p_T.

$$I^{\ell}_{\mathrm{Rel}} = \frac{I^{\ell}_{CH} + I^{\ell}_{NH} + I^{\ell}_{Ph}}{p_{\mathrm{T},\ell}} \ , \tag{6.1}$$

I^{ℓ}_{CH} is the energy deposited by stable charged hadrons in a cone of $\Delta R = 0.4$ around the lepton's track. I^{ℓ}_{NH} and I^{ℓ}_{Ph} are the respective energies of the neutral hadrons and photons.

For electrons and muons it is required, that their track's transverse impact parameter is less than 0.02 cm with respect to the primary beam spot and the z position of the lepton vertex has to be within 1 cm around the primary vertex.

Additionally there are lepton specific requirements.

Muon requirements

Muons have to be reconstructed as global muons in the muon chambers, as well as tracker muons. The reduced χ^2 of the global fit has to be smaller than 10 while there must be more than 10 tracker hits associated to the candidate.

Electron requirements

The relative isolation, as well as the transverse momentum p_T, transverse energy E_T and pseudorapidity η are used to define selection criteria on the presence of leptons in the events. In general a electron candidate has to be within $|\eta| < 2.5$ excluding the ECAL-endcap-barrel transition region at $1.4442 < |\eta_{\text{sc}}| < 1.5660$, where η_{sc} is the pseudorapidity of the electron candidates supercluster. It has to pass the `VBTF70` electron-ID, which follows a simple cut based approach to identify electrons [59].

Conversion rejection

Within material, a photon can split into a pair of electron and positron. This happens a lot in the detector, especially in the ECAL, but it also occurs in tracker and hence this splitting will produce additional lepton tracks, which need to be rejected.

An electron is rejected if it did not produce hits in the inner tracker layer, but only in higher layers, which is typical for electrons from photon conversions. Additionally the partner track conversion veto [19] is used to remove electrons which are flagged as stemming from conversions. This algorithm works for each GSF-track and uses the fact, that electron positron pairs from photon conversion are parallel to each other at the decay point and even afterwards in the $r - z$ plane.

The observables of interest to detect these partner tracks are:

1. $\Delta \cot(\Theta) = \cot(\Theta_{\text{CTF-track}}) - \cot(\Theta_{\text{GSF-track}})$
   Difference of $\cot(\Theta)$ for the electron (GSF) track and neighbouring tracks with opposite charge.

2. `dist` = Distance between the parallel segments of the tracks.

If the absolute value of both variables is smaller than 0.02, the electron is considered to stem from a photon conversion and is therefore rejected.
All these criteria of course also apply to positrons in the according way.

Lepton requirement combinations

All lepton candidates passing the previously defined requirements are than accounted for in each event. An event has to pass either of the following criteria to be part of the electron or the muon channel data samples.

muon channel:

- One muon with a $p_T > 20\,\mathrm{GeV/c^2}$ and $\eta < 2.1$. Its relative isolation has to be between 0.15 and 0.2.
- No additional muon with $p_T > 10\,\mathrm{GeV/c^2}$, $\eta < 2.5$ and relative isolation < 0.2.
- No electron with transverse energy $E_T > 15\,\mathrm{GeV}$, $\eta < 2.5$ and relative isolation > 0.2.

electron channel:

- One electron with transverse energy $E_T > 30\,\mathrm{GeV}$ and $\eta < 2.5$ and a relative isolation of less than 0.125.
- The electron must not be a converted photon.
- No muon with $p_T > 10\,\mathrm{GeV/c^2}$, $\eta < 2.5$ and relative isolation < 0.2.
- No additional electron with $E_T > 15\,\mathrm{GeV}$, $\eta < 2.5$ and relative isolation < 0.2.

Using these criteria, a lepton that will lead to a selection of the event has to pass harder criteria than an additional lepton, that will lead to a rejection of the event. This guarantees that the primary lepton is unambiguously defined and that no dilepton events are selected due to reconstruction fluctuations.

Event selection efficiencies "cutflow"

The cutflow for all involved datasets until before the splitting by the number of jets are shown in table 6.6.

6.2.4. Number of jets and number of b tagged jets

From the signal event topology, as it is presented in 2.4.2 and 2.4.4, we expect the most single top t-channel events to have two jets, of which one is b tagged and the other is not. The b-jet is the one from the top quark decay, while the light jet is a remnant from the t-channel top quark production process. Additional jets can be found in signal events, if either the b from the initial state forms a jet which is found in the detector, or through initial or final state radiation, which produces extra gluons. Hence also in events with three or more jets, there will be signal events.

In this analysis the data sets are split into six different categories, based on the number of jets and the number of tagged jets. Groups are formed with events of two, three or four jets of which either one or at least two jets have a b tag. A jet has to fulfil the following requirements to be accounted for:

- number of constituents > 1

Table 6.6.: Cutflow for the event selection in the both lepton channels. The upper one is for the muon channel and the lower one for the electron channel. After the reweighting step, the event numbers are normalised to the expected number of events in $1\,\mathrm{pb}^{-1}$.

						W+jets					
muon	t-channel	s-channel	tW-channel	$t\bar{t}$	inclusive	0,1-jets	2-jets	3-jets	4-jets	Z+jets	diboson
preselection	956968	69037	523449	1288081	5074813	3215010	7106774	2601092	4990579	5852414	1914364
HLT	378308	24578	219929	504857	2285603	1442164	3273993	1171676	2126204	2117867	848802
reweighting	4.247	0.290	2.147	21.722	885.120	491.083	188.877	46.892	28.250	177.471	5.362
Primary Vertex	4.247	0.290	2.147	21.722	885.120	491.083	188.877	46.892	28.250	177.471	5.362
Moun Count	3.950	0.265	1.891	18.645	741.732	386.068	174.885	43.271	25.980	109.312	4.647
Loose Muon Veto	3.949	0.265	1.852	18.214	741.719	386.065	174.881	43.270	25.979	86.295	4.504
Electron Veto	3.944	0.264	1.678	16.384	741.425	385.997	174.770	43.227	25.945	85.750	4.363
IsoMu17 Eff.	3.897	0.262	1.656	16.169	736.402	383.372	173.575	42.933	25.788	85.336	4.327

						W+jets					
electron	t-channel	s-channel	tW-channel	$t\bar{t}$	inclusive	0,1-jets	2-jets	3-jets	4-jets	Z+jets	diboson
preselection	956968	69037	523449	1288081	5074813	3215010	7106774	2601092	4990579	5852414	1914364
reweighting	10.709	0.814	5.099	55.289	1956.779	1089.115	409.103	103.899	66.159	492.184	11.884
HLT	3.476	0.240	1.996	20.238	725.719	386.067	162.628	41.204	26.686	274.352	4.850
Primary Vertex	1.686	0.139	1.000	12.858	116.267	59.020	26.725	7.340	5.340	42.961	0.882
Electron Count	1.336	0.107	0.771	9.561	84.794	40.729	20.992	5.718	4.114	29.618	0.666
Muon Veto	1.336	0.107	0.694	8.579	84.788	40.727	20.991	5.717	4.113	29.567	0.646
Electron Veto	1.335	0.107	0.667	8.229	84.757	40.723	20.979	5.712	4.109	23.798	0.607
Conv. Rej.	1.332	0.107	0.665	8.213	84.562	40.623	20.930	5.700	4.100	23.709	0.605
Conv. Rej. 2	1.300	0.104	0.648	8.009	82.233	39.544	20.346	5.533	3.972	23.023	0.588

- charged electromagnetic energy fraction < 0.99
- neutral hadron energy fraction < 0.99
- neutral electromagnetic fraction < 0.99
- charged hadron energy fraction > 0
- charged hadron multiplicity > 0

It also has to have a transverse momentum $p_t > 30\,\mathrm{GeV/c}$ and be within $|\eta| < 4.5$, while a b tag requires a TCHP discriminator value above 3.41, the tight working point.

In the following these categories will be abbreviated by njmt, where n is the number of jets and m is the number of tagged jets. Hence the selected categories are 2j1t, 2j2t, 3j1t, 3j2t, 4j1t and 4j2t.

The corresponding event yield lists are given in the respective sections in tables 6.8 for the muon channel and in 6.9 for the electron channel.

According to the tables, the highest fraction of single top t-channel events is expected in the 2j1t category. The largest background in all categories is posed by $t\bar{t}$ events. That is the reason why the 4j categories are used, because they are well suited to constrain the $t\bar{t}$ contribution.

6.2.5. Trigger efficiencies

In this analysis there are basically three trigger cuts. The lepton requirement is present in all channels. Additionally in the electron channel cross triggers as listed in table 6.5 are used, which require a jet and and a b tag besides the isolated electron. These cross triggers are not simulated on MC events and hence they need to be approximated by using the trigger primitives and appropriate turn-on functions for the trigger cuts. But cutting on trigger level and offline on closely related, but not identical, quantities like jet p_T and b tag discriminator can lead to a turn-on behaviour in the selection efficiency, if the online and offline cut thresholds are close by.

Since it is assumed that the leptonic part, and the hadronic part, and the BTagIP factorize, in the following the three different trigger components and their efficiency turn-on functions will be discussed separately,

Lepton efficiency

The electron on trigger level is required to have at least $p_T > 27\,\mathrm{GeV/c}$ while the according offline requirement is already $E_T > 30\,\mathrm{GeV}$.

Similar things hold also for the muon trigger, where online a transverse momentum for the muon of $p_T > 17\,\mathrm{GeV/c}$ is required, while the offline cut demands $p_T > 20\,\mathrm{GeV/c}$.

These differences between the online and offline cut make it save to assume, that the lepton efficiency is close to 100 %. Hence it is not necessary to account for a turn-on function of the lepton trigger components.

Jet trigger efficiency

The offline jet p_T cut of 30 GeV/c is the same as the online trigger requirement for the jet part of the cross triggers. However in the trigger calorimeter jets are used, while in the offline selection the cut is done on Particle Flow jets. Therefore it can not be known which of the Particle Flow jets fired the trigger and hence the full combinatorics for each jet to be triggered or not has to be calculated.

The trigger probability in turn is not flat, but strongly p_T dependent around the threshold value. This is due to the fact, that the p_T of an online HLT calorimeter jet is not necessarily the same as that of an offline reconstructed Particle Flow jet. It is not even guaranteed, that all selected jet have a corresponding jet on trigger level. For the measurement of the PF jets efficiency in the HLT selection, an independent data sample based on the Mu24 trigger is used. The trigger jet objects are matched to the PF jets if they are within a cone with $\Delta R < 0.5$ next to each other. It is also required, that the triggering muon shall have a distance of $\Delta R > 0.5$ to all the considered jets. Then the trigger efficiency for each jet can be calculated in different p_T and η regions, where the efficiency is defined as the ratio between the number of PF jets having a trigger jet partner to the total number of jets in a certain region. In order to parametrize this efficiency in dependence of p_T a Gompertz-function is used, which is defined as

$$\varepsilon(p_T) = a \cdot e^{b \cdot e^{c \cdot p_T}} \tag{6.2}$$

where the free parameter a, b and c are retrieved with a maximum likelihood fit to the turn-on curves shown in figure 6.1

The fitted parameters suggest no strong dependence on the sign of η, while there is some dependence on the absolute value. It is also visible, that even jets with a p_T of only 20 GeV/c have a non-negligible HLT_CentralJet30 efficiency, while on the other hand jets with a p_T between 30 and 60 GeV/c do no yield 100% efficiency.

The trigger turnon for the HLT_CentralJet30 efficiency is only needed no b tag is required also in the trigger. For the triggers where a jet and a b tag are requested the following approach is used instead.

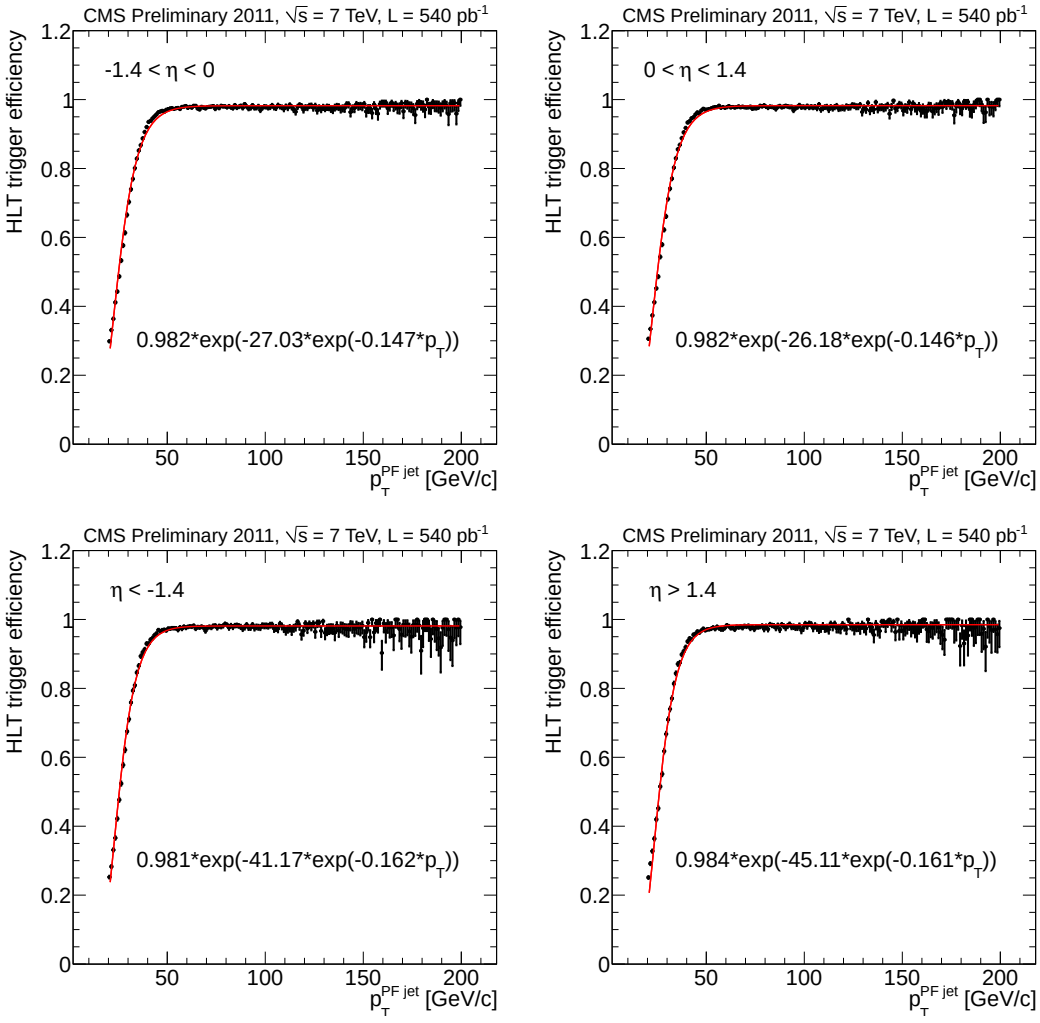


Figure 6.1.: Fitted Gompertz-functions to parametrise the jet trigger turn-on [78]. The trigger efficiency rises steeply around the cut threshold of $p_T > 30\,\mathrm{GeV/c}$ and reaches 100% efficiency only after $\approx 60\,\mathrm{GeV/c}$.

Combined jet p_T and b tag trigger efficiencies

Since the trigger objects are not used in the offline analysis themselves, but just some loosely related corresponding PF jets, it can not be said for each jet whether it fired a trigger or not. One can only say what the probability for each jet to be triggered is in dependence of its p_T and TCHP discriminator. Since the cross triggers are not simulated on MC, these probabilities have to be combined to retrieve an event weight. The full combinatorial probability for each individual jet being triggered or not has to be exploited and the probabilities for those events which yield a positive trigger outcome (in this case: at least one jet above the p_T threshold and being b tagged) have to be added up.

For the single top cross section measurements a turn-on parametrisation for the CentralJet30_BTagIP trigger component was developed. The two dimensional turn-on function [78] is given in figure 6.2.

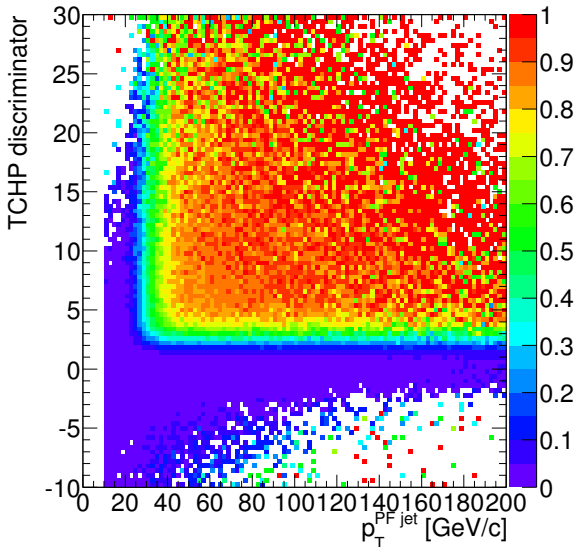


Figure 6.2.: The turn-on behaviour of the CentralJet30_BTagIP trigger is given in dependence of jet p_T and TCHP discriminator. [78]

The two dimensional distribution is projected in 15 TCHP discriminates bins and Gompertz functions in dependence of the jet p_T are used to parametrise the turn-on curves. Three of them are shown exemplary in figure 6.3

They are used to deduce per jet its probability to fire the cross trigger. The different probabilities are then combined by exploiting the full combinatorics for each jet to be triggered or not and adding up all the probabilities.

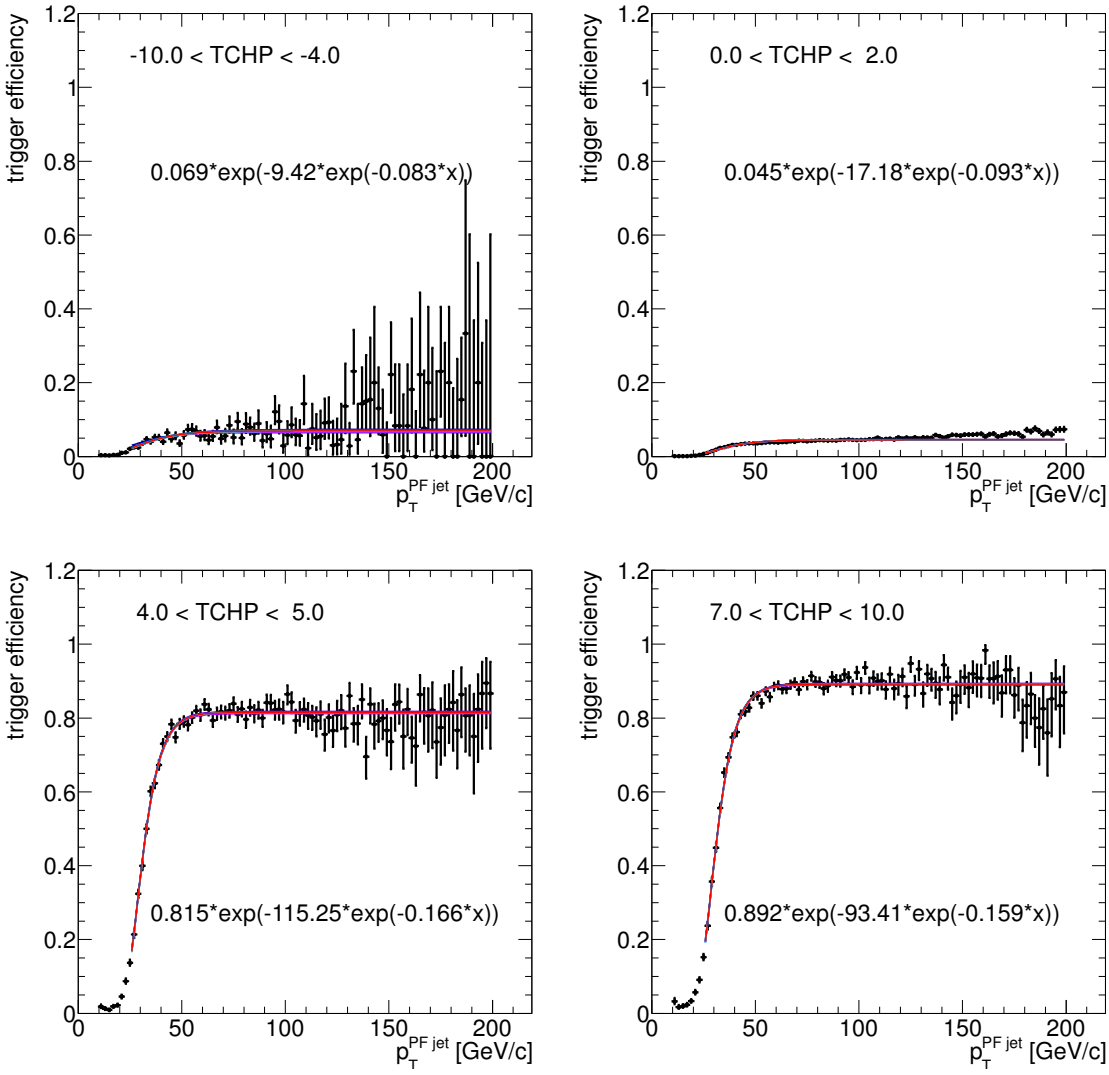


Figure 6.3.: Projections of the CentralJet30_BTagIP trigger turn-on functions in selected TCHP bins. [78]

6.2.6. Event weights

In order to achieve a better agreement between generated MC events and measured data, different types of weights need to be applied on the MC events. Some of them have already been mentioned in 5.1.3. In the following the different event weights that have to be used in this analysis will be presented.

3D pileup reweighting

As already stated in 4.3.2, a pileup of events occurs if multiple proton interactions happen during one bunch crossing, which is the usual case. The additional produced particle enter the detector and produce signals, that are not related to the one single interaction one wants to study. To account for this pileup noise, it has to be simulated also in generated events. In CMS, the generated MC events are produced with a flat pileup distribution and need to be weighted according to the measured pileup distribution in the used data sample.

For Summer11 MonteCarlo, as it is used here, it is additionally necessary to account for the out of time pileup on analysis level, since this was not simulated in these samples. By "out of time pileup" the additional detector signals from collisions in the previous and the following bunch crossing is meant. To account for this, three dimensional weight distributions depending on the before, in-time, and after pileup need to be used. These distributions are diced using the simulated pileup distribution and the data pileup distribution, by picking randomly distributed pileup scenarios from either distribution and using them to generate three Poisson distributed numbers in each case. The three dimensional distributions of data and MC are then divided by each other to generate the weight distribution. For each MC event the three pileup values are used to retrieve the according weight from the distribution.

The pileup scenario depends on the total inelastic cross section. The recommended value for this analysis is $71.0\,\text{mb} \pm 5.7\,\text{mb}$ for 2011's LHC collisions [100]. The uncertainties are taken into account by dicing also the distributions for the varied inelastic cross section and using the systematic shifts which are introduced due to this uncertainty in the inference later on.

The three dimensional distribution of weights is depicted in figure 6.4 for all considered total inelastic cross sections.

On each axis the possible pileup from 0 - 35 events is shown for the before, intime and after component. A "cigar" like shape is clearly visible. The area of highest density depends on the assumed inelastic cross section. The effect is also visible in figure 6.5, where the projections in the x-y-plane are given.

Cross section

The number of events in a generated sample is not related to the event rate of a certain process in measured data. Therefore each MC sample event is weighted to the expected rate in the integrate luminosity of the data set used. The weight is calculated as

$$w = \frac{\text{Cross section}}{\#\text{Events generated}} \times \text{integrated luminosity} \tag{6.3}$$

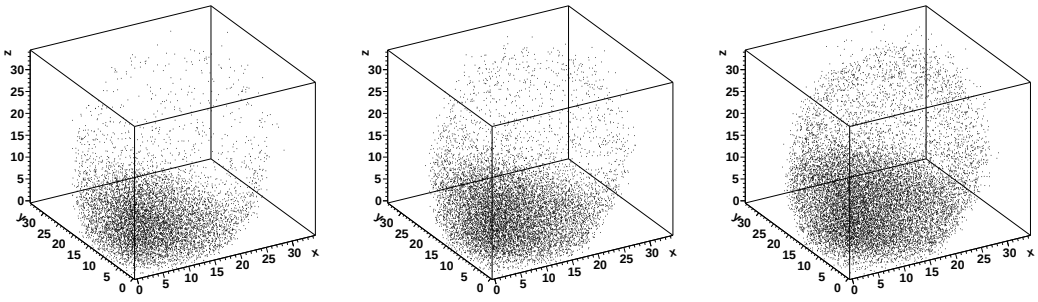


Figure 6.4.: Pileup weight distributions. left: 65.3 mb, middle: 71.0 mb, right: 76.7 mb. This distribution is just used to assign a probability to three successive pileup values. Thus there is no connection between the axis x,y,z and the before, intime and after pileup, but they are interchangeable.

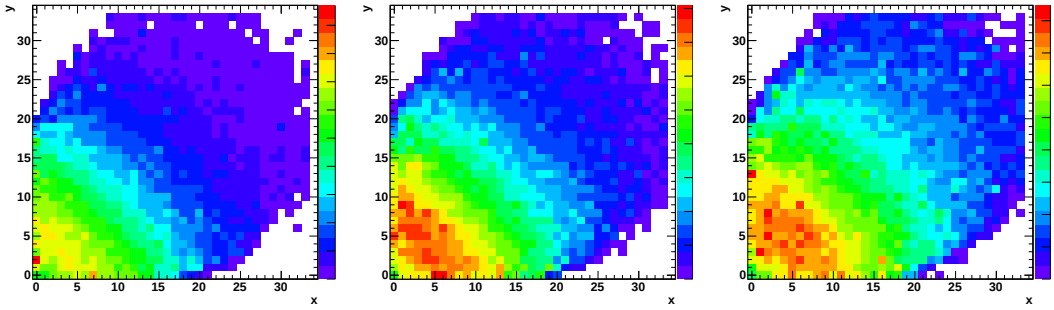


Figure 6.5.: Pileup weight distributions projected in the x-y-plane. left: 65.3 mb, middle: 71.0 mb, right: 76.7 mb. The colour code for the z direction represents the the probability [0,1] in dependence of two pileup vales.

The numbers for the used cross sections and the number of events available in each MC sample are listed in table 6.1 and 6.2. The integrated luminosities were given in section 6.1.2 and are $1170.21\,\mathrm{pb}^{-1}$ in the muon channel and $1560.718\,\mathrm{pb}^{-1}$ in the electron channel.

b-tag scale factor and efficiency

The b-tag scale factors SF_b and SF_{light} [50], which account for different tag efficiencies in MC and data, as well as the b-tag efficiency and the mistag rate, need to be accounted for in form of event weights.

Before applying the scale factors one needs the tag efficiencies for for b-, c- and light-jets. These are calculated for all processes separately in different p_T, η bins, also distinguishing between events with two or three jets to account for any possible dependence. Two of these distributions are shown exemplary in figure 6.6

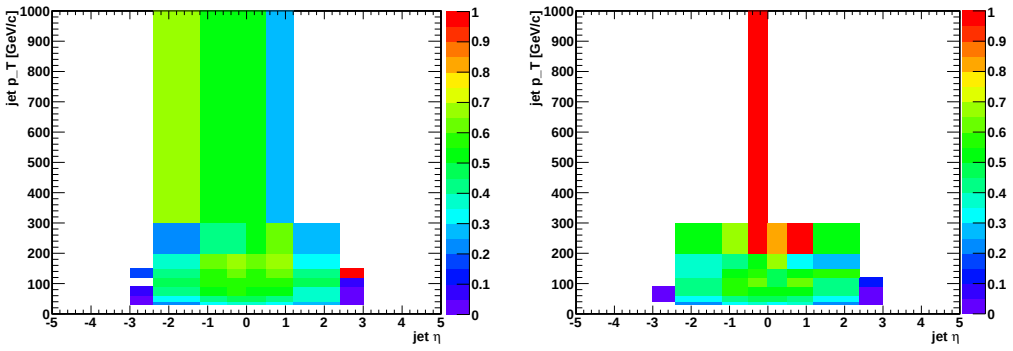


Figure 6.6.: b-tag efficiency for single top t-channel events with two jets. On the left in the electron channel and on the right in the muon channel

The scale factors are provided by the b-tag Physics Objects Group as functions which depend on the jet p_T [50]. For b-jets the scale factor on the track counting high purity tagger at the tight working point is

$$SF_b(x) = 0.895596 \cdot \frac{1.0 + 9.43219 \cdot 10^{-5} \cdot x}{1.0 - 4.63927 \cdot 10^{-5} \cdot x} \tag{6.4}$$

while the provided central function for light jets is

$$SF_{light}(x) = 1.20711 + 6.81067 \cdot 10^{-4} \cdot x - 1.57062 \cdot 10^{-6} \cdot x^2 + 2.83138^{-10} \cdot x^3 \tag{6.5}$$

where x is the jet transverse momentum p_T.

For both functions uncertainties are given. For the SF_b these are discrete values in several p_T bins between 30 and 670 GeV/c. The uncertainties range from 2.6 % to 8.6 %. For the light jets scale factor the uncertainties are given as alternative funtions. The b jet and light jet scale factors, together with their uncertainties, are shown in figure 6.7.

In order to calculate the event weight, the following steps need to be done.

- Find all possible combinations of assigning a tag to a jet or not, which results in the correct sum of tagged jets, required by the event selection.
- Calculate for each of these cases p_{MC} as the product of tag efficiencies ϵ for jets in the ensemble $\{t\}$, which are assumed to be b-tagged, and $1 - \epsilon$ for the other jets in the ensemble $\{!t\}$ which are assumed to be untagged. The actual b-tag discriminator value is not considered here directly.

$$p_{MC} = \prod_{i \in \{t\}} \epsilon_i \cdot \prod_{j \in \{!t\}} (1 - \epsilon_j) \tag{6.6}$$

- p_{Data} is calculated just like p_{MC}, but here the efficiency is multiplied by the scale factor.

$$p_{Data} = \prod_{i \in \{t\}} \epsilon_i \cdot SF_i \cdot \prod_{j \in \{!t\}} (1 - \epsilon_j \cdot SF_j) \tag{6.7}$$

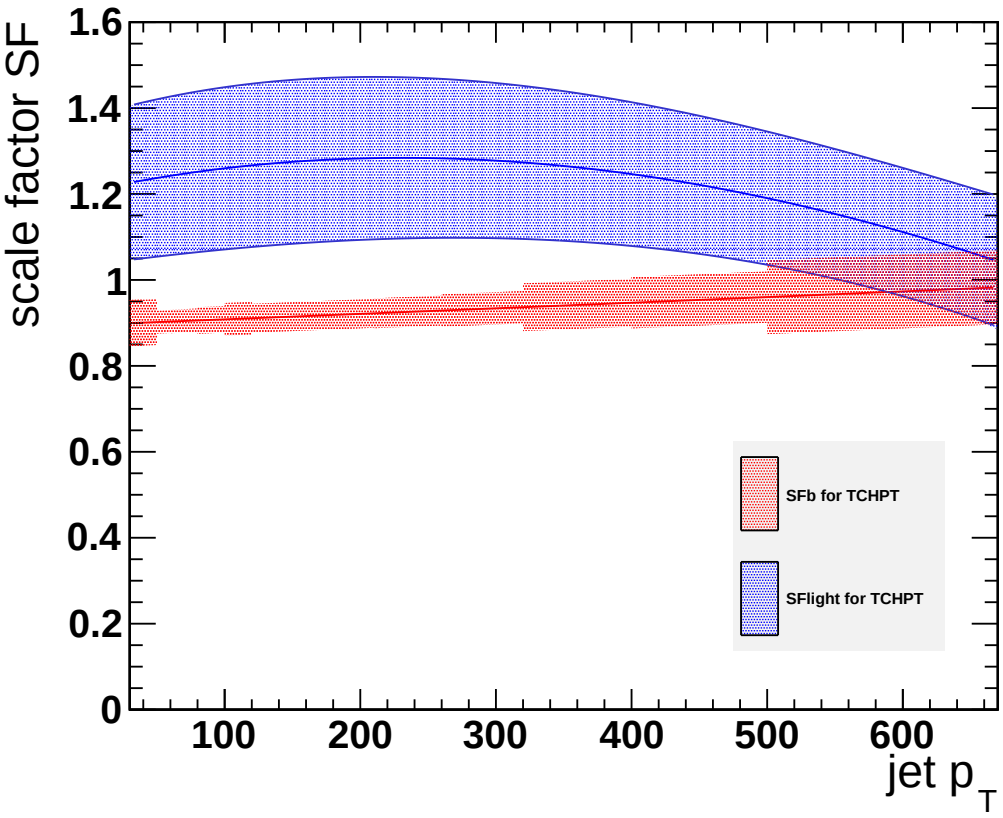


Figure 6.7.: b-tag scale factors for the track counting high purity tagger at the tight working point with uncertainty bands. The scale factors are shifted to either end of the uncertainty band in the statistical inference, to account for systematic uncertainties.

- p_{MC} and p_{Data} are added up separately for each possible combinations. For example if events with two jets of which one is b tagged are selected, than the two cases in which jet1 is b tagged and jet2 is not, and vice versa are evaluated in equations (6.6) and (6.7) and both probabilities are added.

The event weight is then given by

$$w = \frac{\sum p_{MC}}{\sum p_{Data}} \tag{6.8}$$

6.2.7. QCD modelling and suppression

For this analysis, no generated QCD Monte Carlo samples is available, which yields sufficient statistics after the selection cuts presented above. It would also be questionable if simulated QCD are well enough modelled in the selected phase space to be used in this analysis. Therefore the QCD templates are modelled from data events with altered

lepton selection criteria, and an additional cut is applied in each lepton decay channel to suppress the remaining QCD even further.

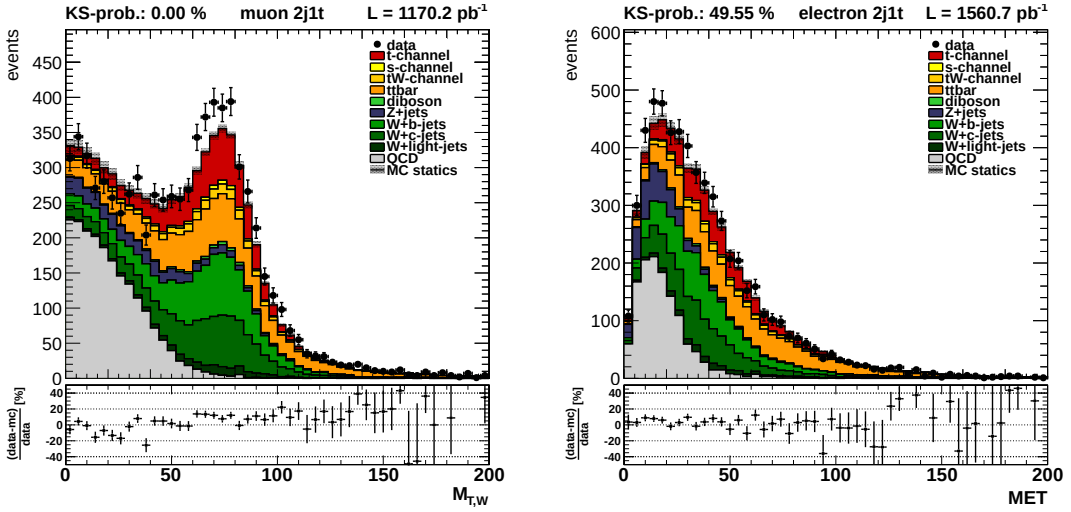
To select QCD events, the lepton criteria were altered in the following way:

muon: The relative isolation for the muon has to be between 0.3 and 1.0. This makes it necessary to use a different trigger, the `Mu15`, which requires a muon with transverse momentum $p_T > 15\,\mathrm{GeV/c}$, but no isolation.

electron: In the electron channel an event is selected as QCD event, if two of the following three criteria are fulfilled.

- $I_{rel} > 0.125$
- Impact parameter $d > 0.02\,\mathrm{cm}$
- fail the `VBTF70` electron ID

The so selected data events are reconstructed in the usual way. To estimate the actual amount of QCD events that pass the normal selection, template fits need to be made. In the muon channel the transverse W mass (MTW) is used, since it discriminates well between QCD events and others, as can be seen in figure 6.8(a). For the electron channel it was found, that MTW does not separate QCD and the other events as good as in the muon channel. Instead the missing transverse energy (MET) is fitted here, which shows a clear separation as can be seen in figure 6.8(b). Both distributions depicted in figure 6.8 are for events with two jets of which one is b tagged.



(a) Transverse W mass distribution in the 2j1t category for the muon channel.

(b) Missing transverse energy distribution in the 2j1t category for the electron channel.

Figure 6.8.: MTW and MET distribution in the muon and electron channel after fitting the QCD template from side-band with generated MC events to data. Only the scale factor for QCD events (gray) was applied, while all MC events are scaled to prediction.

For comparison the MTW distribution in the 2j1t electron channel is given in figure 6.9. In the electron channel there is a long tail of QCD events to higher MTW values. A

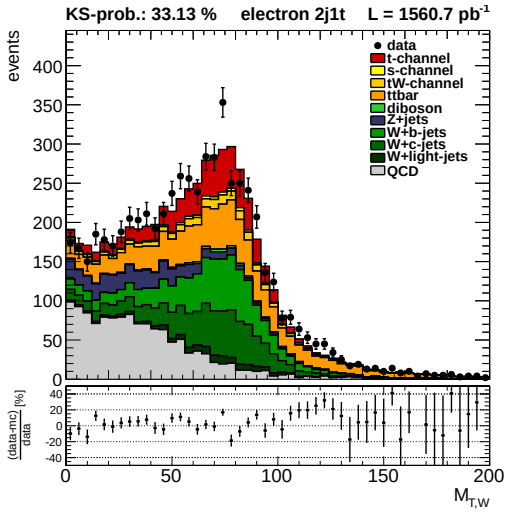


Figure 6.9.: Transverse W mass in the electron channel without any QCD suppression cut on MET. Since the QCD events have a long tail towards higher MTW values is distribution not as well suited to define a QCD suppression cut as MET.

clean cut to suppress QCD while not losing to much signal events is not possible here and hence the cut is applied on MET.

In order to reduce the amount of QCD events in the final data set and not use data events twice for statistical inference, in either channel a cut on MTW and MET respectively is applied. Only two free parameters are used in the fit, which are one scale factor for the QCD events β_{QCD}, and one common scale factor β_{MC} for the templates from the generated MC samples listed in table 6.1. The fit is done in each jet/tag category separately. For the muon channel, the QCD fit is done on events with MTW $< 40\,\mathrm{GeV}/c^2$ and for the electron channel the threshold is MET $< 35\,\mathrm{GeV}/c^2$. The events above these thresholds are then used for the rest of the analysis. The fit results are given in table 6.7.

As can be seen, the results for β_{MC} are in the order of 1.0, which is already a good result considering that the fit model has only two degrees of freedom. There is no prior expectation for the β_{QCD} values. Fitting values smaller than 1.0 in all cases but one is good, because it means that the retrieved templates contain more events from sideband data, than are acutally expected in the signal region.

QCD purity estimation

In order to estimate the purity of the selected QCD events, the previously described selection cuts were applied to the generated MC samples. The number of events passing these cuts are used to estimate the contamination of the QCD sample. It was found, that the fraction of none QCD events after the selection is smaller than 0.6 % in both lepton channels.

Table 6.7.: Fit values from the QCD template normalisation in the electron and muon channel

	muon channel		electron channel	
category	β_{MC}	β_{QCD}	β_{MC}	β_{QCD}
2j1t	0,832 ± 0,097	0,331 ± 0,025	1,084 ± 0,053	0,497 ± 0,038
2j2t	1,028 ± 0,249	0,273 ± 0,209	1,258 ± 0,184	0,375 ± 0,124
3j1t	1,317 ± 0,099	0,094 ± 0,064	1,263 ± 0,068	0,525 ± 0,096
3j2t	1,050 ± 0,072	0,000 ± 0,297	1,341 ± 0,191	0,260 ± 0,558
4j1t	1,228 ± 0,042	0,000 ± 0,086	1,408 ± 0,096	0,575 ± 0,279
4j2t	0,877 ± 0,104	1,786 ± 1,452	1,106 ± 0,059	0,000 ± 0,449

6.2.8. Event yield

After applying all selection cuts and weights, the samples are split by the number of jets to be 2,3 or 4 in an event and requiring either one or at least two of them to be b-tagged. Hence there are 6 subsamples for which the event yields in the different jet-tag categories are listed in tables 6.8 and 6.9 for the samples from table 6.1 in the muon and the electron channel.

Table 6.8.: Amount of selected MC events and the predicted number of events for the different processes in the muon channel.

sample	2j1t		2j2t		3j1t		3j2t		4j1t		4j2t	
	nExp	nMC	nExp	nMC	nExp	nMC	nExp	nMC	nExp	nMC	nExp	nMC
t-channel:	741.37	59306	21.67	2047	266.70	21153	63.65	5307	63.98	5242	22.22	1710
s-channel:	53.97	4049	12.37	1114	18.39	1365	6.37	474	4.84	358	1.68	130
tW-channel:	177.23	15776	4.28	496	197.14	17854	18.37	1637	91.90	9050	18.52	1617
$t\bar{t}$:	1048.43	21177	147.73	3704	2182.60	44508	584.55	12124	1798.49	38964	667.56	13288
W+b-jets:	935.973	17642	48.145	928	343.746	13362	33.314	1192	108.462	6765	19.035	1025
W+c-jets:	813.874	12636	2.366	41	207.649	7055	1.701	45	57.071	3289	.901	35
W+light-wjets:	139.770	2539	1.284	4	40.425	1865	.146	4	10.740	857	.090	4
Z+jets:	161.22	1599	4.24	57	68.23	724	3.61	38	20.47	232	2.75	25
diboson:	46.22	7205	3.61	966	17.20	2607	1.94	404	4.08	682	0.61	129
QCD:	255.95	773	3.27	12	26.67	284	0.00	9	0.00	78	7.14	4
total background:	3632.637	-	227.295	-	3102.050	-	650.001	-	2096.053	-	718.286	-
total expected:	4374.007	-	248.965	-	3368.750	-	713.651	-	2160.033	-	740.506	-
data:	4771	-	285	-	3828	-	690	-	2385	-	685	-

Table 6.9.: Amount of the predicted number of events and selected MC events for the different processes in the electron channel.

sample	2j1t		2j2t		3j1t		3j2t		4j1t		4j2t	
	nExp	nMC	nExp	nMC	nExp	nMC	nExp	nMC	nExp	nMC	nExp	nMC
t-channel:	429.89	29999	15.41	1091	171.59	11846	45.34	2844	45.12	3162	16.39	972
s-channel:	33.91	2185	8.53	593	12.60	782	4.39	247	2.82	206	1.50	87
tW-channel:	130.24	9943	4.55	358	146.00	11632	17.20	1133	72.17	6249	17.54	1171
$t\bar{t}$:	860.50	14919	135.94	2494	1726.30	30313	516.93	8133	1431.36	26759	597.78	8920
W+b-jets:	490.166	8874	31.319	493	209.915	7703	25.661	708	74.145	4221	15.067	635
W+c-jets:	412.527	5939	.969	15	122.549	4049	1.048	33	37.337	2046	.586	20
W+light-jets:	83.071	1489	.890	4	26.924	1221	.214	5	8.238	638	.033	1
Z-jets:	36.33	368	1.04	10	23.35	233	1.74	15	8.23	98	1.00	13
diboson:	26.41	3464	2.10	421	10.92	1504	1.40	201	3.14	460	0.50	68
QCD:	108.45	218	4.13	11	70.87	135	5.20	20	56.92	99	0.00	19
total background:	2181.604	-	189.468	-	2349.428	-	573.783	-	1694.360	-	634.006	-
total expected:	2611.494	-	204.878	-	2521.018	-	619.123	-	1739.480	-	650.396	-
data:	2652	-	224	-	2755	-	535	-	1990	-	597	-

6.3. NeuroBayes training

This section is devoted to the details on the NeuroBayes training in both lepton channels. If not specified otherwise, the presented settings are used for both trainings.

6.3.1. Training parameters

A NeuroBayes training was performed on all selected events with at least two jets of which at least one is b tagged. Thus there is only one training per lepton channel and no splitting into jet or b tag categories is done, to have a network that is trained with the maximal possible number of events.

Besides the default settings, NeuroBayes was configured to perform an iterative training, using the BFGS method (see section 5.1.3). In the hidden layer one node more than in the input layer is used, which is 38 for the muon and 40 in the electron channel. The ratio between signal and background events in the trainings sample is reweighted to a prior signal probability of 50 % with formula (5.6). This is done to hide the actually lower prior probability for signal events from NeuroBayes since otherwise the retrieved discriminator values would not cover the full range [-1,1] and the signal and background shapes were not as distinct.

To be sure that no overtraining occurs, i.e. the network learns insignificant features from the training sample by heart, the `RTRAIN` option is invoked. NeuroBayes is setup to use only 80 % of the available dataset for the network training and use the residual 20 % to calculate the loss function on an independent sample. Overtraining would be indicated if the error on the test sample rises while it still decreases on the trainings sample. This would lead to an interruption of the training process. Due to the extensive variable preprocessing and the thorough variable selection, this was not observed in any training for this analysis.

The NeuroBayes input variables are chosen from a large set of available variables, which exploit the signature of single top t-channel events. Only those variables, which yield at least three standard deviations of additional significance to the whole network performance, are selected in the training.

The variables presented in the following pass this 3σ criterion in at least one of the network trainings.

6.3.2. NeuroBayes input variables

A brief explanation for each variable used in NeuroBayes together with the stacked distribution for the 2j1t category in both lepton channels is listed here. Variables which require the presence of three jets, are taken from the 3j1t category samples. These categories are chosen because they contain the highest signal contribution. The according distributions for the other jet/tag categories can be found in appendix A.

There are four categories of variables that are used in NeuroBayes:

1. Lepton properties, since the isolated lepton from the W boson decay is a very distinct feature for signal events in this analysis.

2. Jet properties for the first, second and third jet. These are the hadronic counter part of the event signature. Variables which exploit the spatial distance between two jets or a jet and the lepton are used as well.

3. Reconstructed candidate properties from the top quark, W boson and also the reconstructed neutrino. As explained in section 4.5 these candidates and their properties like the transverse W boson mass MTW are reconstructed precisely, to use them as NeuroBayes inputs.

4. Variables describing the over all event typology. These are the missing transverse energy MET and the absolute sum of all transverse momenta in the event H_T. MET gives a well separation between single top events and W+jets or QCD. In contrast H_T has a long tail to higher values for $t\bar{t}$ events and is rather narrow for signal events, which can be expected, since the two decaying top quark in $t\bar{t}$ events yield multiple high energetic jets or leptons.

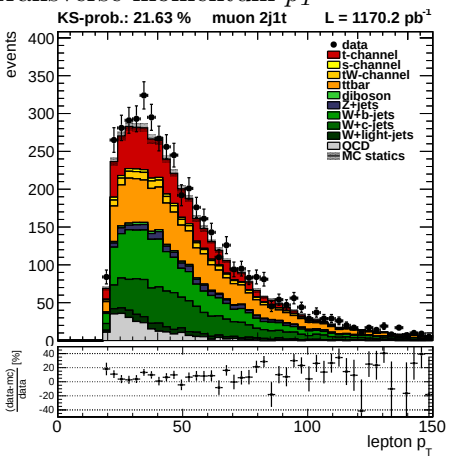
All contributions from simulated samples are scaled to their expected rate, as given by cross section and integrated luminosity. QCD is scaled by the parameter retrieved from the MTW or MET fit (see section 6.2.7). In each of the figures, the distribution of the stacked model templates is shown together with the data distribution for this variable. In the top row, the Kolomogorov-Smirnov test value is printed, which gives an estimate for the compatibility of both (data and stack) distributions.

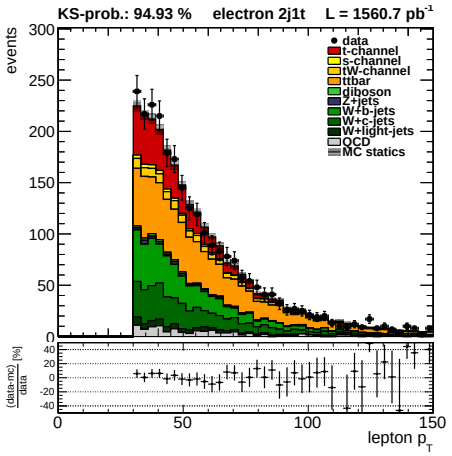
Below the plots the relative residuals for each bin are given.

Lepton properties

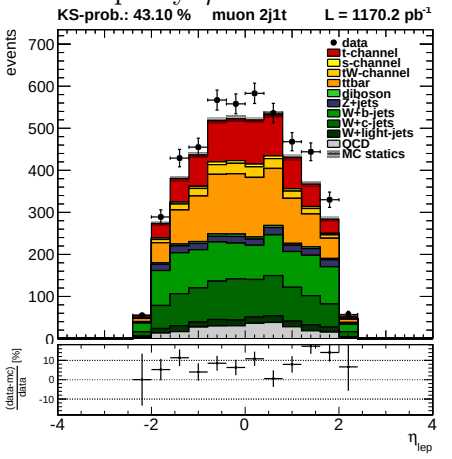
Three basic information about the selected, isolated lepton are collected. As is explained in 2.4.3 the lepton is believed to be produced by the leptonic W boson decay.

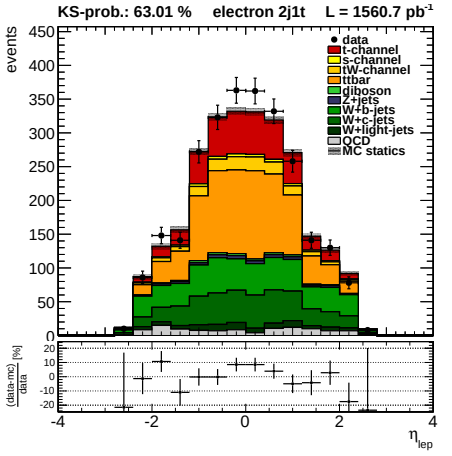
- Transverse momentum p_T



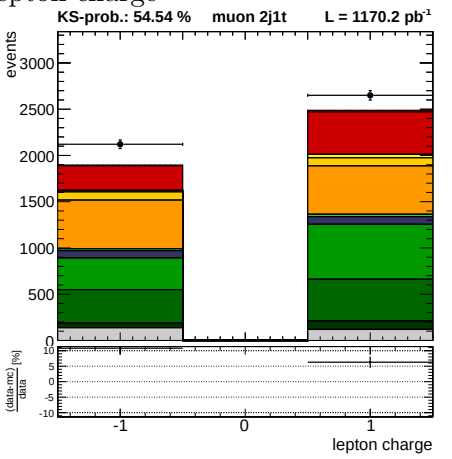


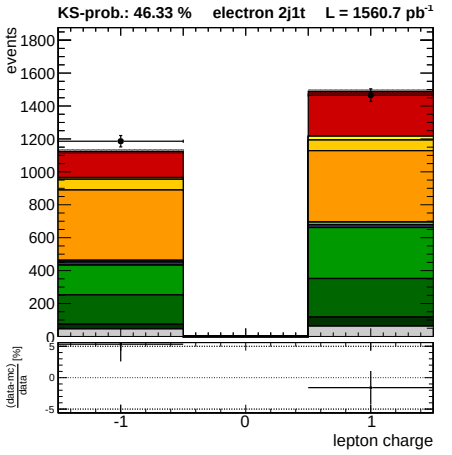
- Pseudo rapidity η





- lepton charge

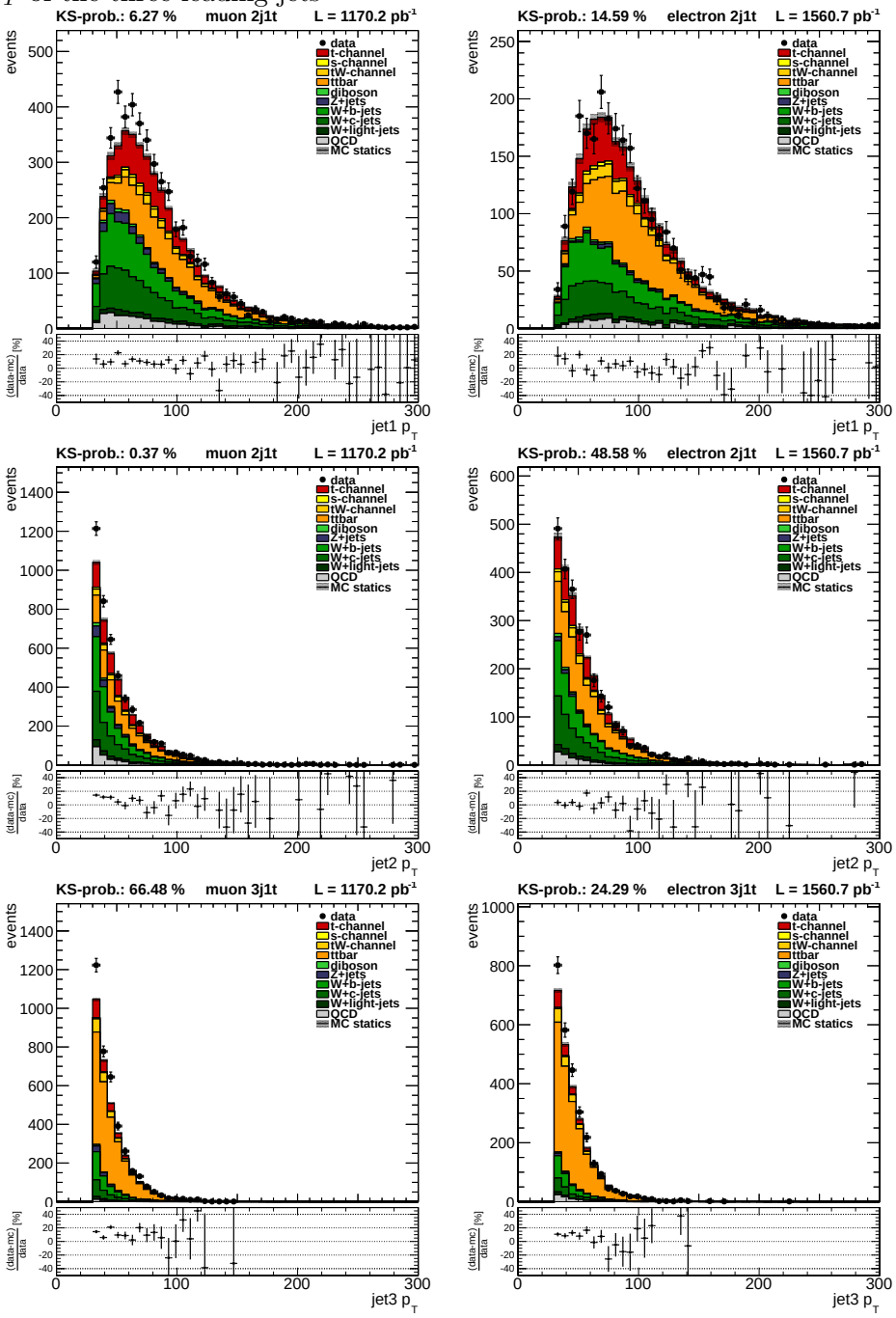




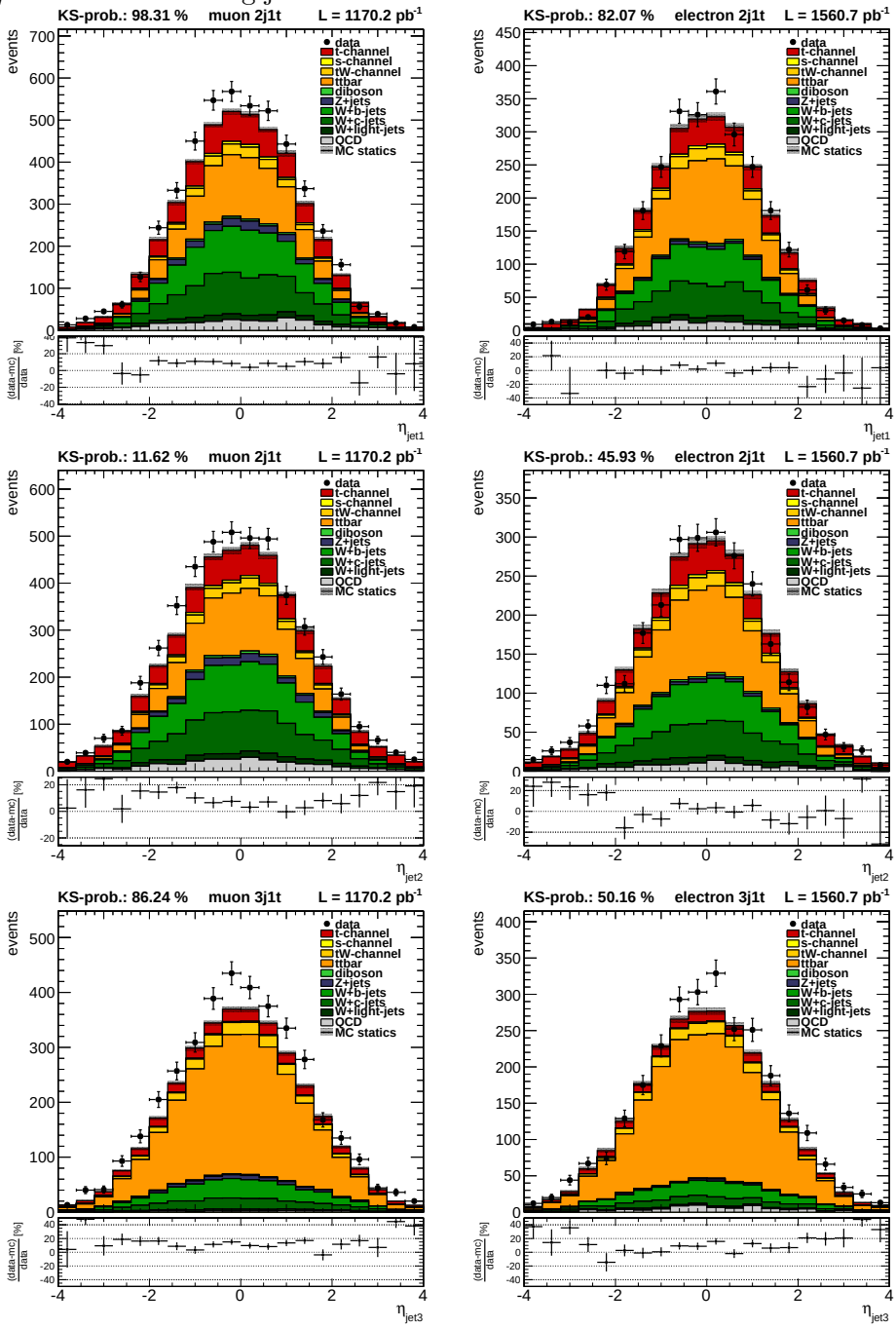
Jet properties

Events are selected with at least two and up to four jets. For the first three jets, ordered by p_T, basic quantities like p_T and η are used as training input. Also for each possible jet pair the invariant mass, their distance ΔR in the $\eta - \phi$ plane and the difference in pseudorapidity η is calculated. The difference in η between each jet and the lepton is also used as an input.

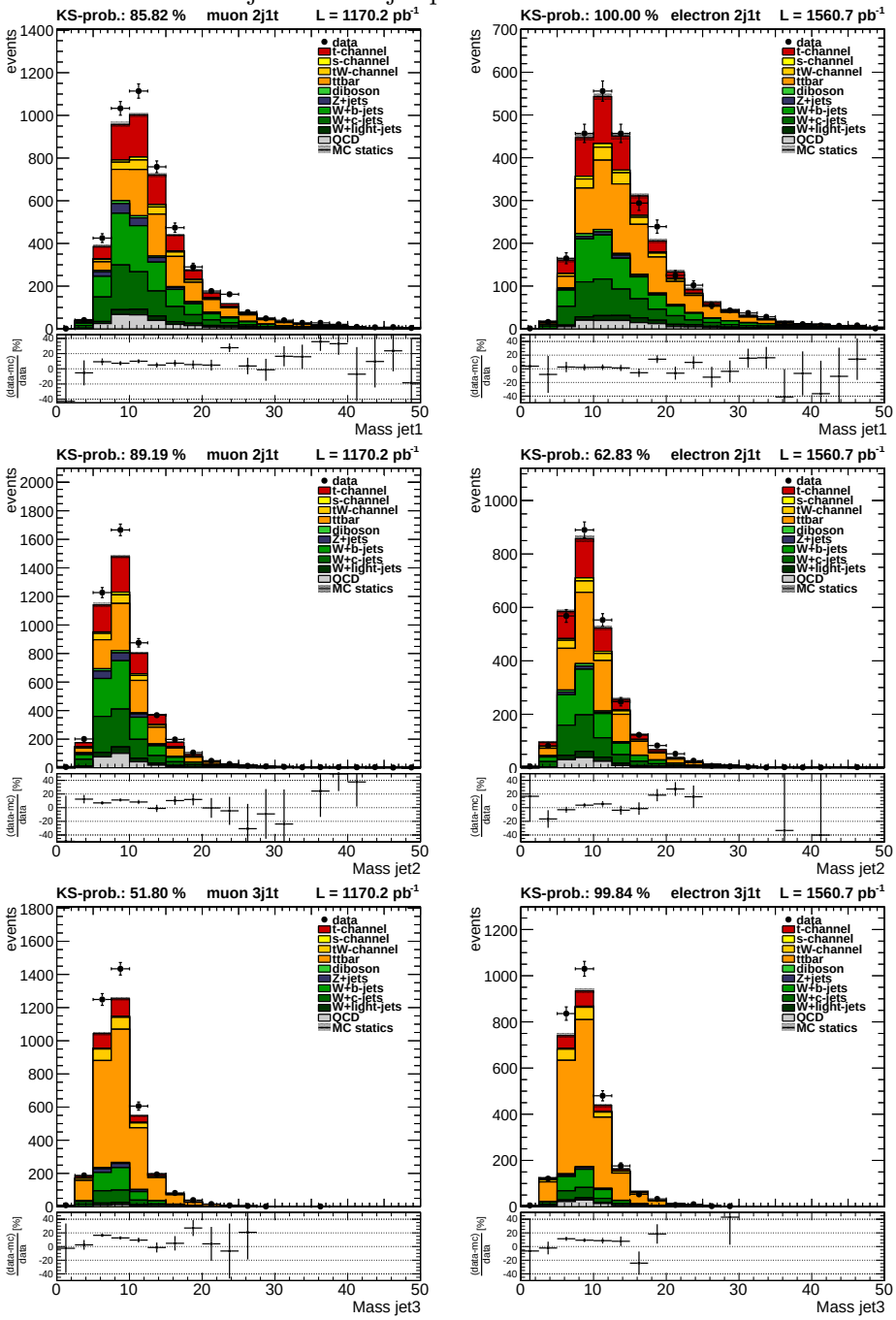
- p_T of the three leading jets

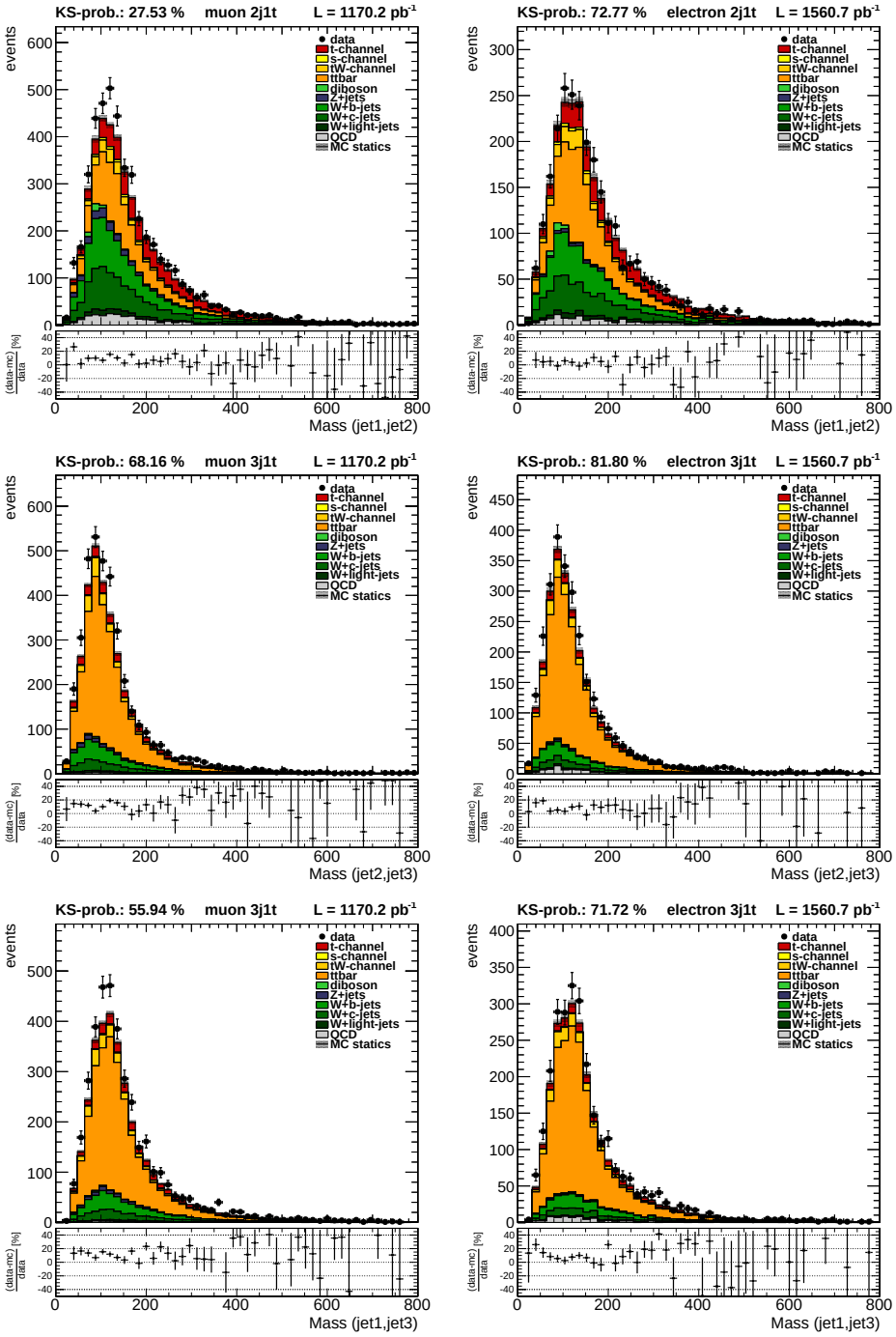


- η of the three leading jets

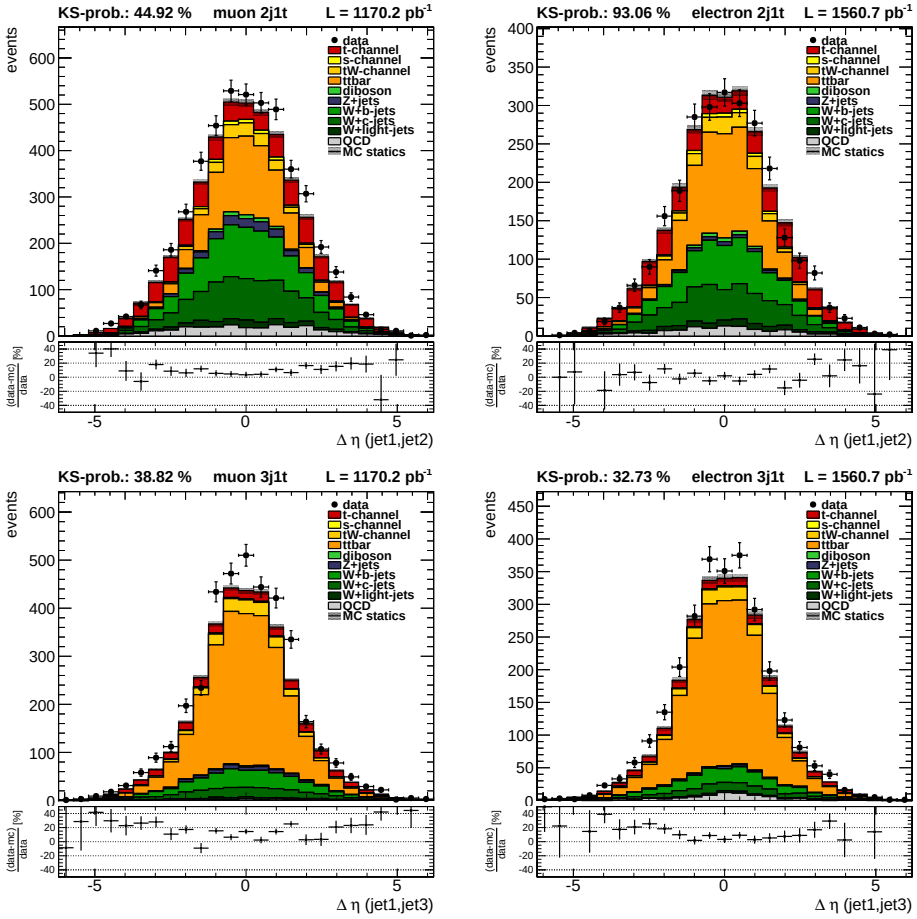


- Masses of individual jets and of jet pairs

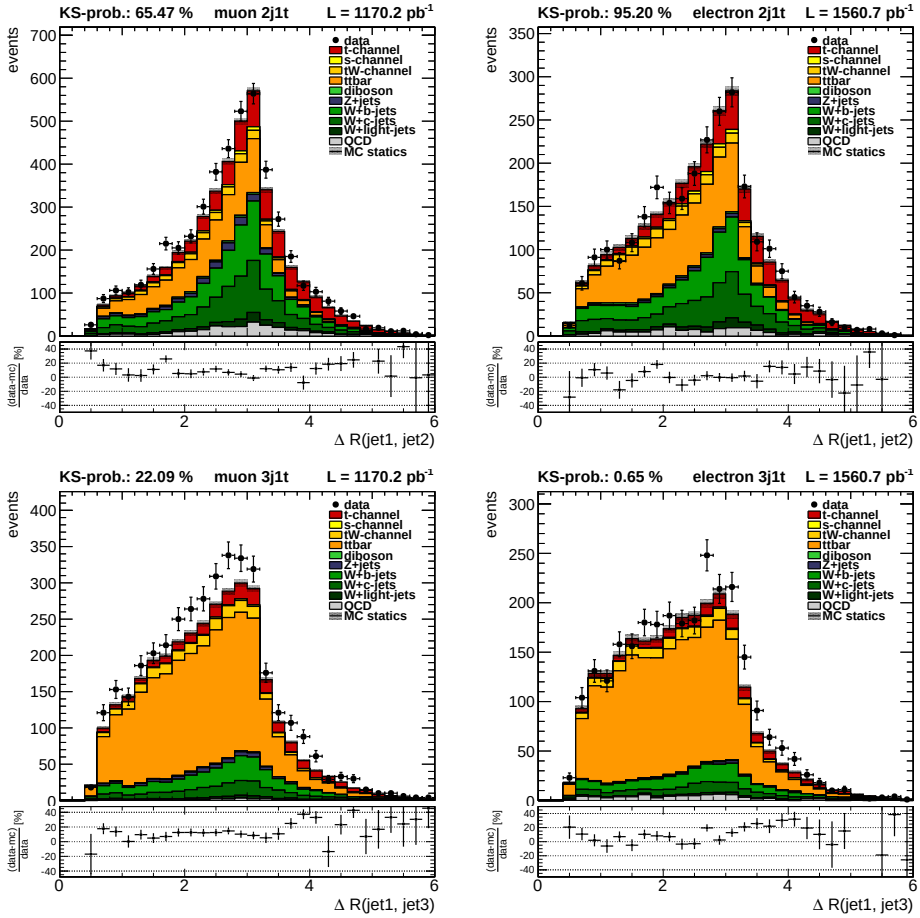



KS-prob.: 27.53 % muon 2j1t L = 1170.2 pb-1
KS-prob.: 72.77 % electron 2j1t L = 1560.7 pb-1
KS-prob.: 68.16 % muon 3j1t L = 1170.2 pb-1
KS-prob.: 81.80 % electron 3j1t L = 1560.7 pb-1
KS-prob.: 55.94 % muon 3j1t L = 1170.2 pb-1
KS-prob.: 71.72 % electron 3j1t L = 1560.7 pb-1
events
(data-mc)/data [%]
Mass (jet1,jet2)
Mass (jet2,jet3)
Mass (jet1,jet3)
data
t-channel
s-channel
tW-channel
ttbar
diboson
Z+jets
W+b-jets
W+c-jets
W+light-jets
QCD
MC statics

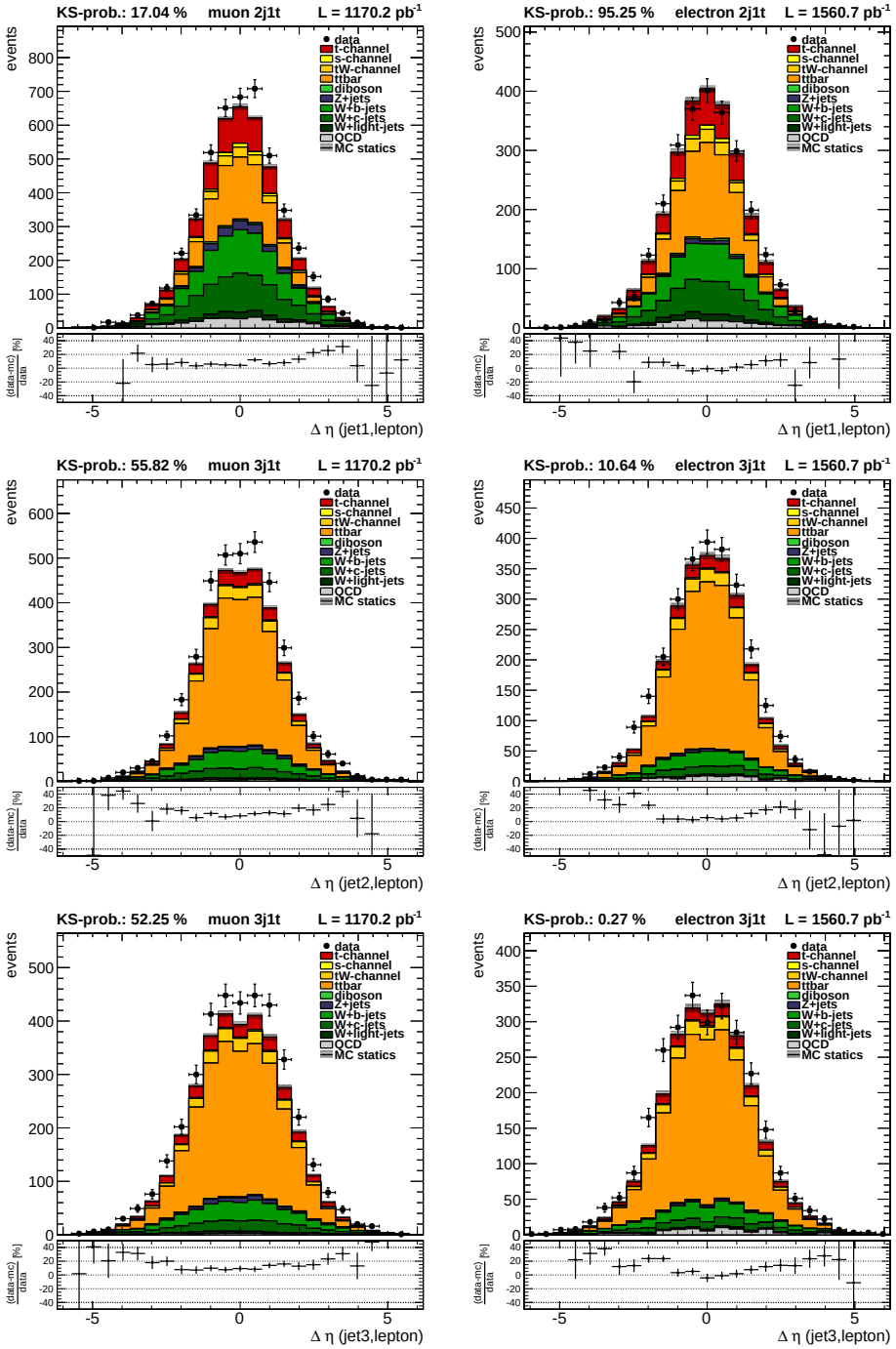
- Difference in pseudorapidity between two jets $\Delta\eta$



- Distance in $\eta - \phi$ plane between two jets ΔR



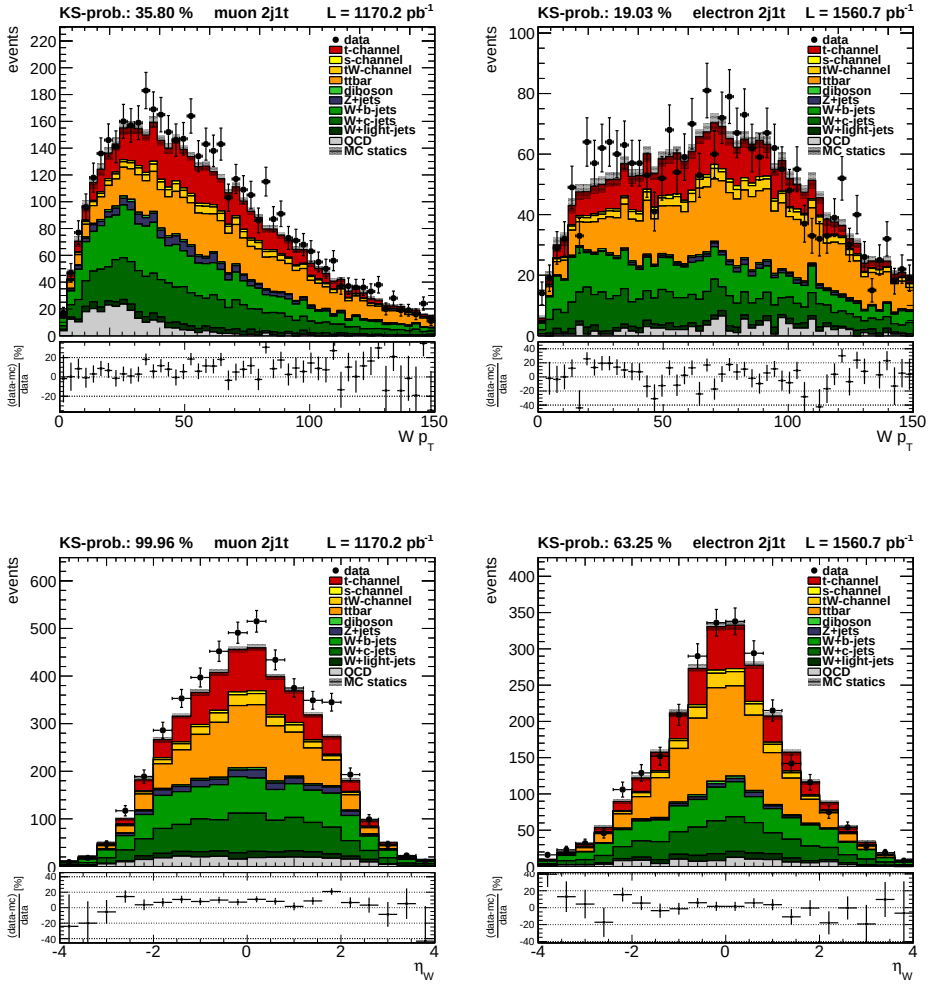
- Difference in pseudorapidity η between a jet and the lepton

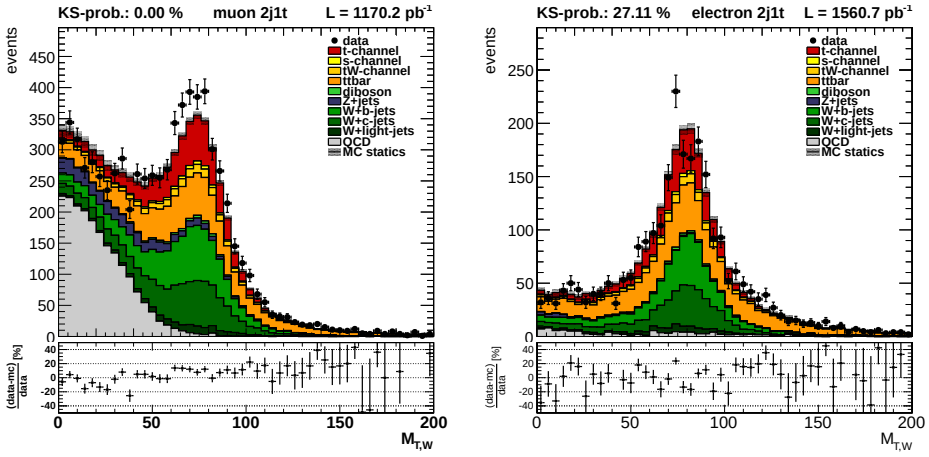


Reconstructed candidate properties

The jets in the previous list were descendingly ordered by the transverse momentum p_T. But there are also other information which can be used to select a jet, like the role it is assigned in the top candidate reconstruction explained in section 4.5. In fact it turned out, that the explicit grouping of information for the light quark candidate jet and the not tagged light-quark jet still yields additional significance. Also properties from the reconstructed W and top candidates and of the reconstructed neutrino are valuable information.

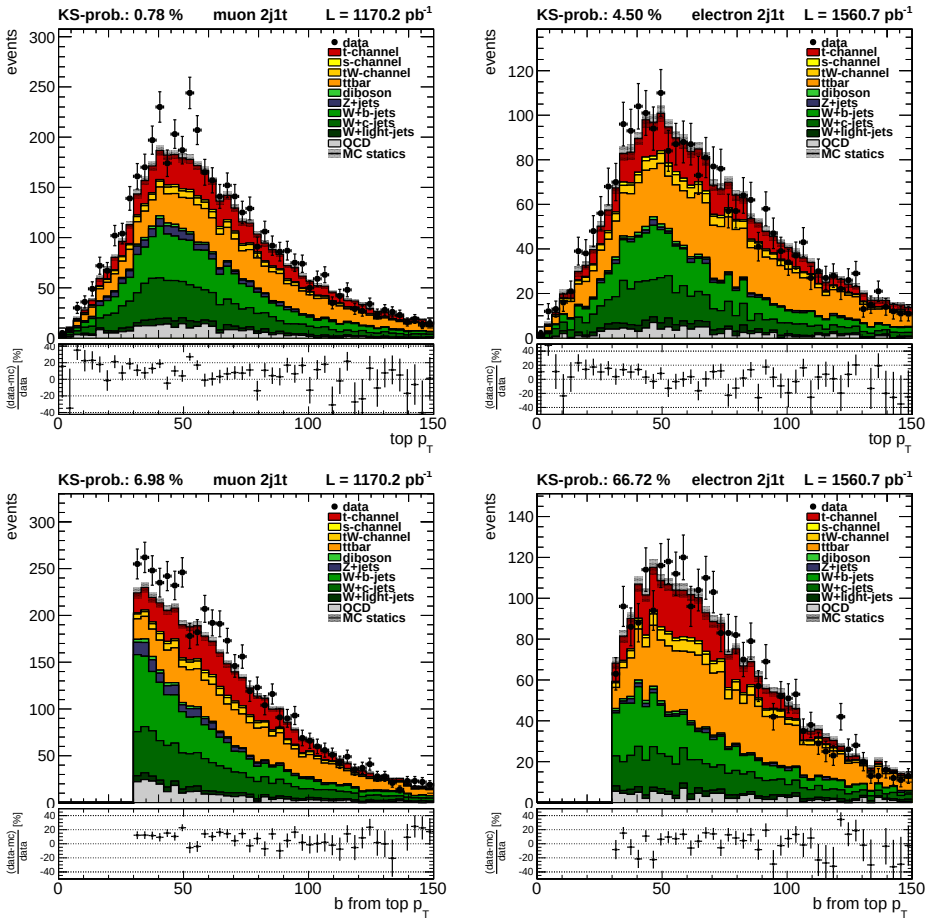
- Transverse momentum p_T, transverse mass and pseudo rapidity η of the reconstructed W boson candidate.



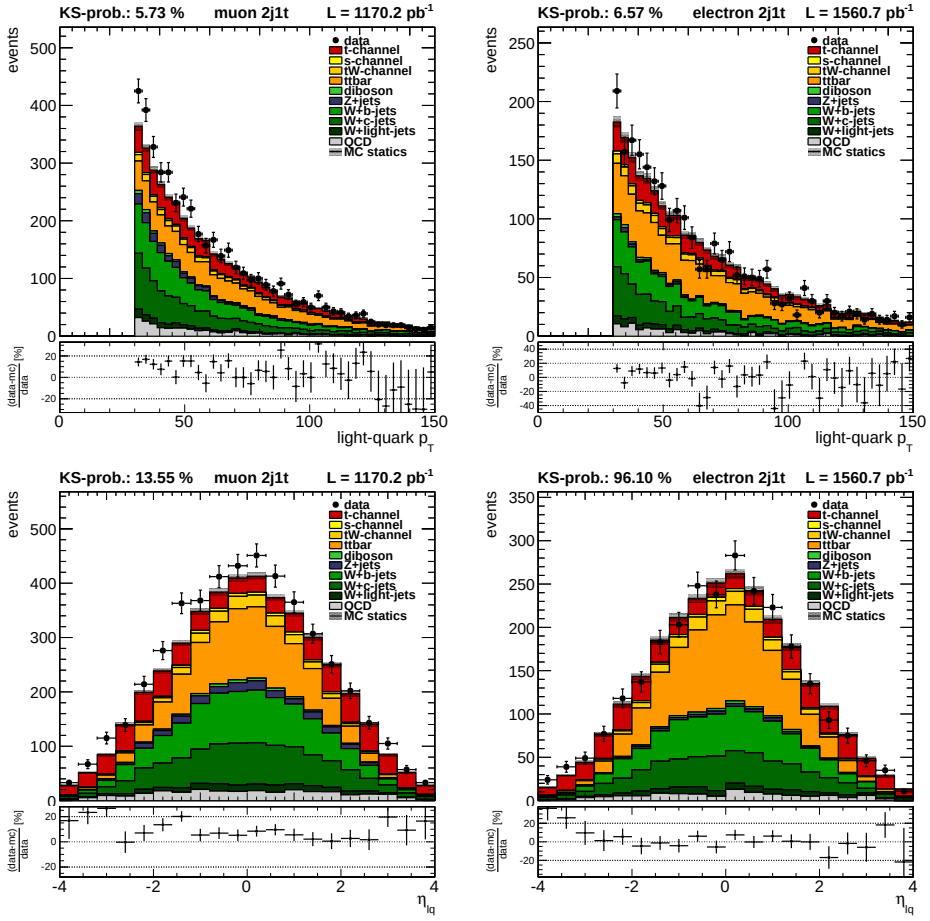


The MTW distribution in the muon channel is shown before the QCD suppression cut of MTW $> 40\,\mathrm{GeV}/c^2$

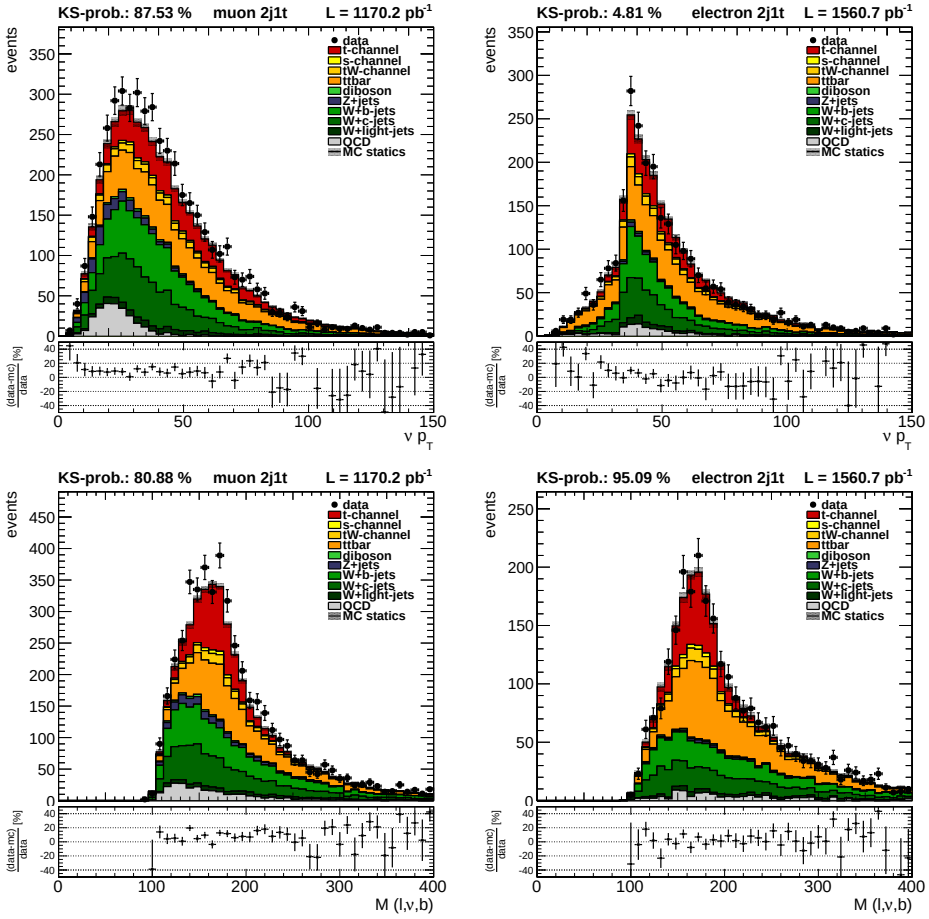
- Transverse momentum p_T of the top candidate and the b jet which is assumed to be produced by the top decay



- Transverse momentum p_T and pseudo rapidity η for the light quark jet in the single top reconstruction hypothesis.

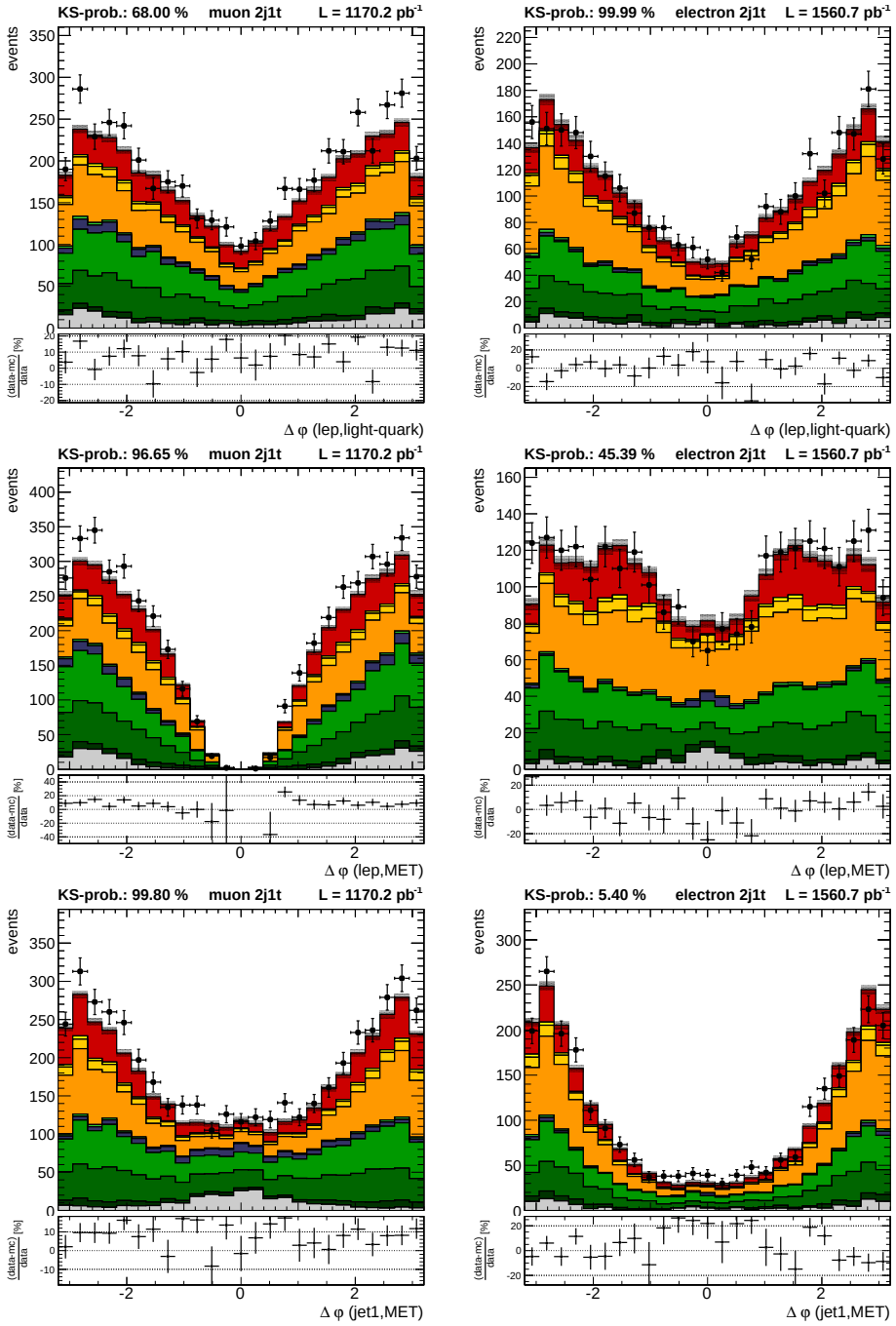


- Transverse momentum p_T for the reconstructed ν candidate and the invariant mass of lepton, ν and and b jet.

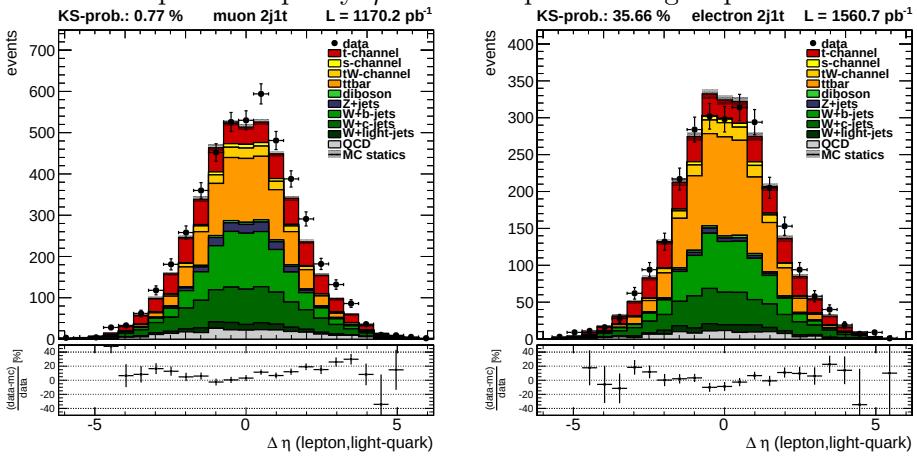


- Difference in φ between
  1. lepton and light quark jet.
  2. lepton and the direction of the missing transverse energy MET.
  3. jet with highest p_T and the direction of the missing transverse energy MET.

  These variables are shown with the signed values, but in NeuroBayes only the absolute values are used, since the sign of the difference turned out to be not discriminating.

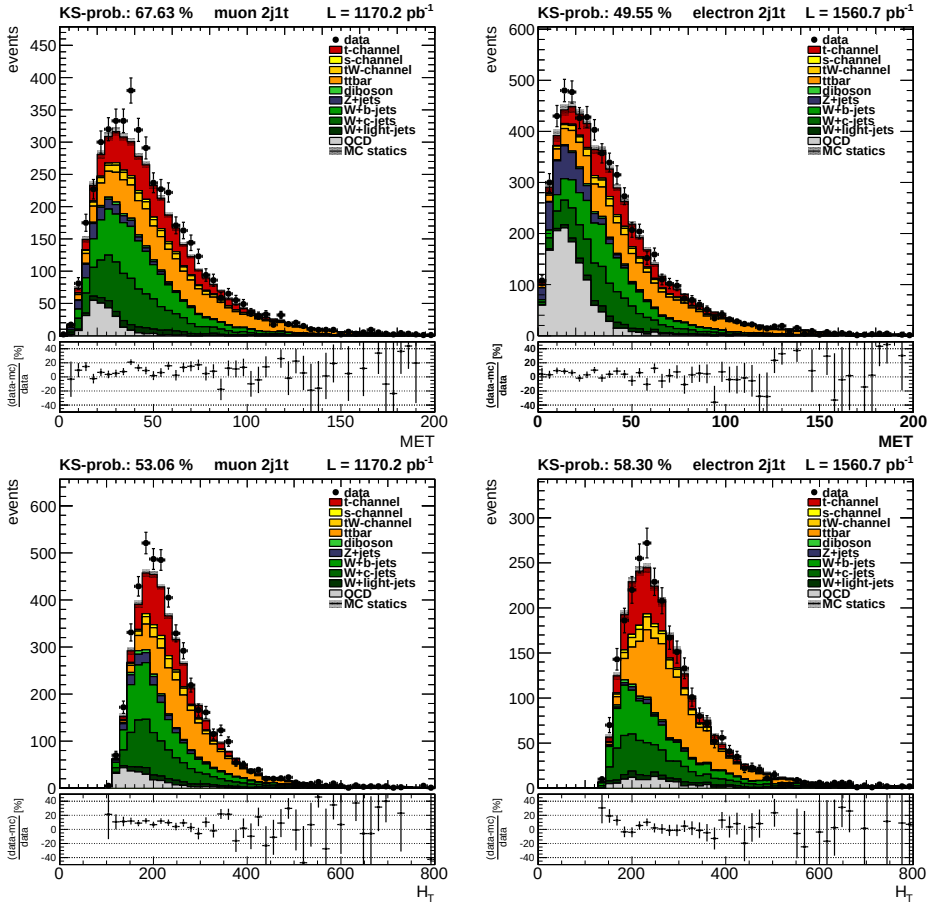


- Difference in pseudo rapidity η between lepton and light quark



Event topology

Two variables which describe the overall event topology are used as NeuroBayes inputs. One is the missing transverse energy MET, the other is the absolute sum of all transverse momenta for all jets, leptons and MET, H_T.

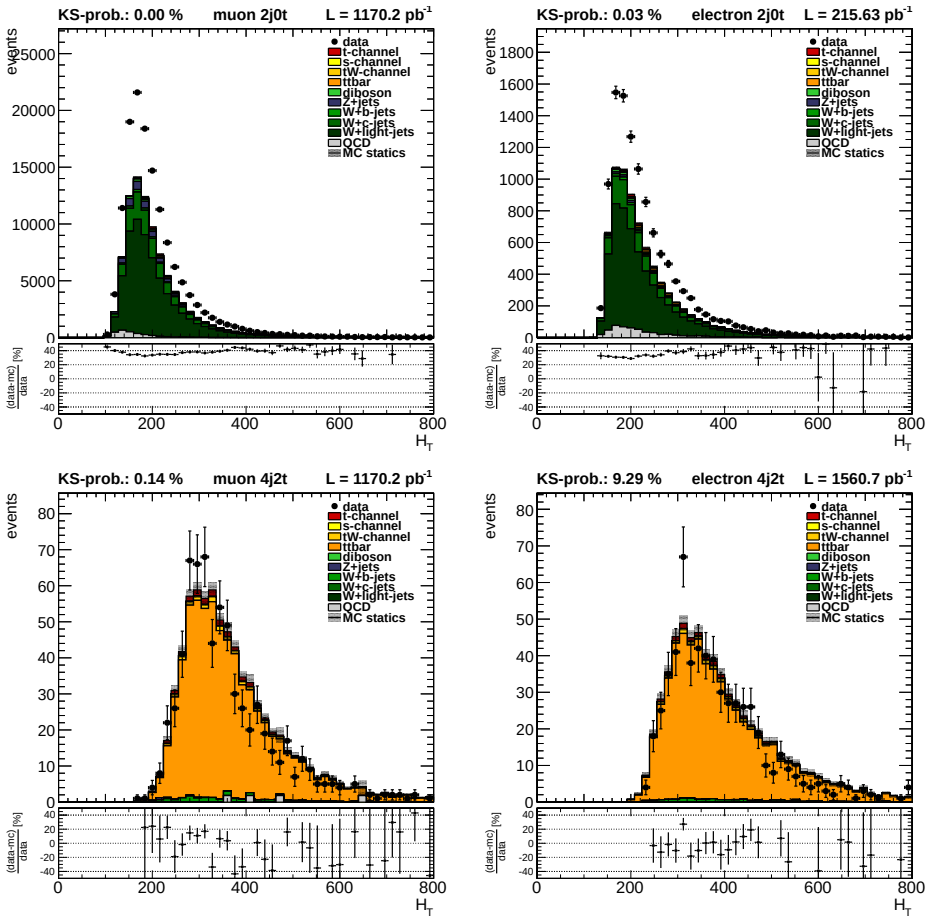


The missing transverse energy distribution in the electron channel is drawn without the QCD suppression cut of MET $> 35\,\mathrm{GeV}/c^2$

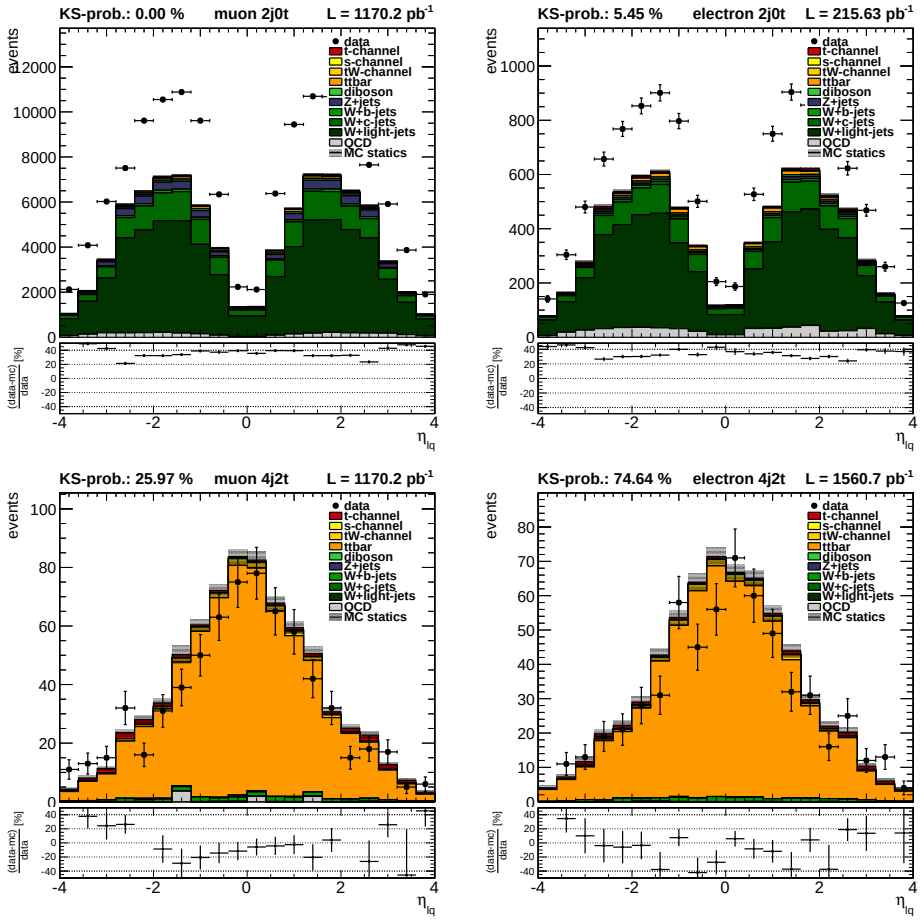
Variable distributions in control regions

The figure given above show the distribution of input variables in the signal enriched region. To gain confidence in the modelling, the distribution of the two most import variables (see section 6.3.3) in two additional categories will be shown. This is on the one hand the 4j2t category, which mainly consists of $t\bar{t}$ background events, and the 2j0t region which is dominated by W+jets events.

- Distribution of H_T in the sideband regions



- Distribution of η_{lq}



The given distribution reveil a good agreement in the 4j2t category between data and MC, while there are rate and shape differences in the W+jets dominated 2j0t region. These differences in the W+jets contribution is one of the reasons for the disentangeling of the cross section nuissance parameter for the W+jets events by the different flavour of the leading jet and assinging large prior uncertainties on these cross sections, as is listed in table 6.12. It is also the reason why the categories with no b tagged jet are neither used in the NeuroBayes training nor in the statistical inference.

6.3.3. Variable ranking

From the variables presented in section 6.3.2 and in appendix A, the ones in table 6.10 are used in the NeuroBayes discriminator for the muon channel and those in table 6.11 for the electron channel. The variables are ranked by the significance of their correlation to the target. This ranking is done by iterating over all variables and removing in each step the variable which has smallest target correlation significance. The added significance of a variable is then calculated as the difference between the total target correlation of the variable set with and without this variable.

The most discriminating variable in both trainings is the pseudorapidity of the light quark jet η_{lq}. This variable was also used as the only discriminating variable in another CMS single top t-channel cross section measurement [71].

Other variables have a high rank in both trainings as well. There is for example H_T which is second best in both trainings. The invariant mass of lepton, neutrino and b jet, which is essentially the sum of all top quark decay products, is on rank 5 in the muon NeuroBayes training and on rank 6 in the electron channel training, respectively.

There are other variables instead, that were found to yield an additional significance of more than 3σ in only on lepton channel. These are for example the reconstructed top p_T and the p_T of the b jet that is assumed to stem from the top quark decay. Both are only present in the electron channel, but have been pruned from the list of input variables in the muon channel.

On the other hand are the absolute difference in φ between the second jet and the first jet or MET only used in the muon channel.

But overall most of the input variables yield comparable significance in both data sets.

Table 6.10.: Variable ranking by relevance for the NeuroBayes discriminator performance in the muon channel.

rank	additional significance	only this var	loss when removed	global corr. to others [%]	variable name
1	109.27	109.27	13.39	87.2	$\eta_{\text{light-quark}}$
2	61.22	77.47	38.81	85.5	H_T
3	62.17	64.02	15.42	94.3	Mass (jet1,jet2)
4	33.28	21.63	34.85	63.0	light-quark p_T
5	30.79	50.14	23.88	50.7	Mass (l,ν,b)
6	30.15	50.43	13.70	89.6	Mass (jet1,jet3)
7	22.09	37.53	12.05	73.9	Mass (jet2)
8	22.22	44.87	12.49	50.8	$M_{T,W}$
9	21.07	26.08	14.53	69.9	jet3 p_T
10	20.71	49.26	21.53	52.4	Mass (jet2,jet3)
11	17.69	31.64	16.97	11.2	lepton charge
12	16.38	10.25	5.75	65.1	η_W
13	14.58	106.21	16.17	90.2	$\Delta\eta$_jet1_jet2
14	10.85	43.67	18.58	77.2	$\Delta\eta$_jet2_lep
15	13.84	30.49	18.97	71.4	$\Delta\eta$_jet1_lep
16	13.36	51.55	14.25	79.3	lepton p_T
17	10.09	78.63	14.59	81.3	$\Delta\eta$ (lepton,light-quark)
18	9.10	71.93	13.30	82.9	η_{jet2}
19	12.90	66.27	12.73	78.5	η_{jet1}
20	11.65	46.29	14.47	70.7	Mass (jet1)
21	12.10	37.78	12.37	84.8	jet1 p_T
22	10.30	28.86	12.05	71.4	$\Delta\eta$ (jet3,lep)
23	8.38	21.78	10.00	34.3	$\lvert\Delta\varphi$(lep,lq)$\rvert$
24	7.22	47.30	7.07	88.5	W p_T
25	8.71	33.46	8.22	87.5	MET
26	7.35	16.88	6.18	61.1	Mass (jet3)
27	6.69	42.05	5.89	76.3	η_{jet3}
28	5.99	29.67	7.64	53.3	b-jet from top p_T
29	5.73	91.92	7.49	90.4	ΔR (jet1,jet2)
30	5.50	65.52	7.27	90.2	$\Delta\eta$ (jet1,jet3)
31	5.33	36.05	5.87	76.6	top quark p_T
32	5.39	15.43	5.95	64.9	η_{lepton}
33	5.18	57.44	5.00	87.9	ΔR (jet1,jet3)
34	4.40	31.26	4.28	86.6	jet2 p_T
35	3.78	30.61	3.66	70.6	$\lvert\Delta\varphi$ (lep,met)$\rvert$
36	3.08	11.78	3.85	43.0	$\lvert\Delta\varphi$ (jet1,met)$\rvert$
37	3.35	31.00	3.35	84.8	ν p_T

Table 6.11.: Variable ranking by relevance for the NeuroBayes discriminator performance in the electron channel.

rank	additional significance	only this var	loss when removed	global corr. to others [%]	variable name
1	87.78	87.78	11.00	86.9	$\eta_{\text{light-quark}}$
2	55.39	70.09	31.71	89.3	H_T
3	51.24	16.32	31.58	63.8	light quark p_T
4	38.57	90.54	17.32	90.2	$\Delta\eta$ (jet1,jet2)
5	26.47	43.12	9.76	87.5	Mass (jet1,jet3)
6	26.35	41.82	20.04	49.2	Mass (l,ν,b)
7	23.65	47.29	12.15	73.7	$M_{T,W}$
8	19.75	50.49	8.86	89.0	$\Delta\eta$ (jet2,jet3)
9	16.02	23.52	9.23	70.8	jet3 p_T
10	14.78	26.48	14.09	12.2	lepton charge
11	14.38	51.53	17.36	90.2	Mass (jet1,jet2)
12	10.90	34.58	14.12	78.6	$\Delta\eta$ (jet2,lep)
13	13.80	27.61	15.61	72.8	$\Delta\eta$ (jet1,lep)
14	11.73	37.68	7.61	74.1	Mass (jet2)
15	10.44	5.84	4.82	54.0	η_W
16	9.41	44.93	11.16	72.7	Mass (jet1)
17	10.33	37.10	9.76	85.3	jet1 p_T
18	7.49	64.79	10.12	82.7	$\Delta\eta$ (lepton,light-quark)
19	8.31	24.85	8.39	74.1	$\Delta\eta$ (jet3,lep)
20	7.37	56.08	8.65	88.6	$\Delta\eta$ (jet1,jet3)
21	5.81	57.37	8.91	78.0	η_{jet1}
22	6.14	55.54	7.74	82.4	η_{jet2}
23	6.14	76.52	8.95	92.2	ΔR (jet1,jet2)
24	5.89	12.00	6.54	37.6	$\lvert\Delta\varphi$ (lepton,light-quark)$\rvert$
25	5.66	39.04	6.94	84.6	Mass (jet2,jet3)
26	4.55	41.06	5.54	87.0	ΔR (jet2,jet3)
27	4.76	8.75	3.51	54.1	$\lvert\Delta\varphi$ (jet2,MET)$\rvert$
28	3.16	45.88	7.39	94.3	W p_T
29	4.96	47.56	8.93	86.7	lepton p_T
30	4.08	37.42	7.86	87.8	MET
31	4.84	37.25	5.29	74.6	ν p_T
32	4.54	31.38	4.54	80.3	$\lvert\Delta\varphi$ (lepton,MET)$\rvert$
33	4.39	12.51	4.54	56.9	η_{lepton}
34	4.31	35.46	4.31	78.0	η_{jet3}
35	4.39	48.13	4.30	87.4	Δ R (jet1,jet3)
36	3.55	16.19	3.56	64.3	Mass (jet3)
37	3.22	36.90	3.47	85.4	jet2 p_T
38	3.22	15.62	3.22	74.7	$\lvert\Delta\varphi$ (jet1,jet2)

6.3.4. Discriminator stacks in the signal region

For the 2j1t and 3j1t samples the stacked distribution for the templates, together with the data distribution is shown in figures 6.10 and 6.11. All templates are scaled to the predicted amount of events by cross section and integrated luminosity.

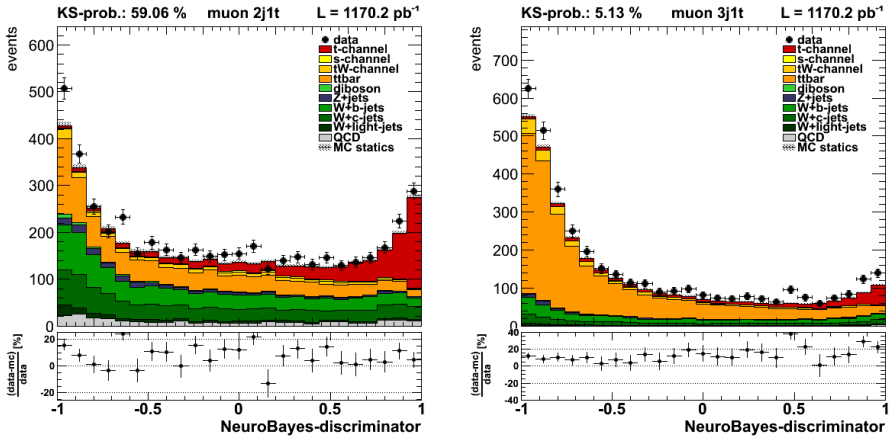


Figure 6.10.: The NeuroBayes discriminator distributions for the electron channel are given as stack for all generated samples and QCD together with the data distribution. On the left there is the 2j1t, and on the right the 3j1t sample.

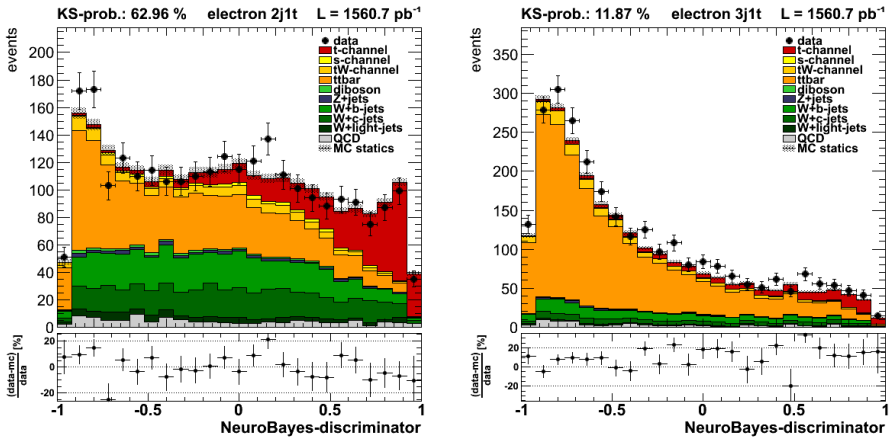


Figure 6.11.: The NeuroBayes discriminator distributions for the electron channel are given as stack for all generated samples and QCD together with the data distribution. On the left there is the 2j1t, and on the right the 3j1t sample.

6.4. Systematic uncertainties

There are various sources of systematic uncertainties than can influence the measured signal yield. Hence they need to be taken into account during the statistical inference as nuisance parameters. This could for example be rate uncertainties like the cross section for some of the background processes, or the uncertainties on the jet energy corrections. While the process cross sections only affect the total number of selected events, other uncertainties also alter the shape of the NeuroBayes discriminators, and are therefore called "shape uncertainty".

It is necessary to use separate pairs of discriminator templates for each process in which build by using either the systematic upwards or downwards shift.

All uncertainties for which it is possible are included in the Bayesian marginalisation. Their nuisance parameters are then integrated out by `theta`.

For some shape uncertainties it is not possible to include them in the marginalisation, because their modelling does probably not cover all possible variations. This can lead to an overconstrain on the observed residual uncertainty [21, 90]. Instead they are studied individually and their expected impact is quadratically added to the marginalised uncertainties.

The expected impact is derived by using a simplified model which only contains the rate uncertainties and the one shape uncertainty one wants to study. Toy-data are build using those templates which mimic either the up- or downwards variation due to this uncertainty. The nominal model is used to derive the signal scale factor β_s for a fit of the nominal model to the shifted toy-data. The relative variation for β_s to 1.0, which is the nominal result, is then quoted as residual systematic impact of this uncertainty on the measured signal strength.

In the following the different sources of systematic uncertainties are discussed.

6.4.1. Process cross sections

Each of the cross sections listed in table 6.1 is assigned a systematic uncertainty. These are:

The signal process t-channel is not constrained. For the QCD contribution a large uncertainty is assumed, since it is not based on theoretical calculations, but it is just estimated from side-bands (sec. 6.2.7). The W+jets cross section was split into 9 independent nuisance parameters by the number of jets and the flavour of the leading jet and a large uncertainty of 200 % is assigned. This was done in order to account for the Q^2 scale and the matching threshold uncertainties in this process, as is explained in detail in sections 6.4.8 and 6.4.9.

6.4.2. Jet energy resolution

For a precise reconstruction of event properties, like the invariant mass of the decaying top quark, it is crucial to know the jet energy and the jet transverse momentum p_T as precisely as possible. This makes it necessary to calibrate the jet energy and momentum reconstruction on data. The CMS detector's jet momentum response and the energy

Table 6.12.: Systematic uncertainties on process cross sections.

process	uncertainty width	uncertainty type
t-channel	∞	flat
s-channel	15%	log-normal
tW-channel	13%	log-normal
$t\bar{t}$	15%	log-normal
W+b with 2jets	200%	log-normal
W+c with 2jets	200%	log-normal
W+light with 2jets	200%	log-normal
W+b with 3jets	200%	log-normal
W+c with 3jets	200%	log-normal
W+*light* with 3jets	200%	log-normal
W+b with 4jets	200%	log-normal
W+c with 4jets	200%	log-normal
W+*light* with 4jets	200%	log-normal
Z+jets	30%	log-normal
diboson	30%	log-normal
QCD	100%	log-normal

resolution where calibrated for example by using the p_T balance in dijet events and in γ/Z+jets events [49].

The updated, but not yet published results, suggest a 10 % worse p_T resolution in data than in the detector simulation. The uncertainty on this measurement is $|\eta|$ dependent. The uncertainties in the different $|\eta|$ regions are:

Table 6.13.: Jet energy resolution uncertainties in dependence of jet $|\eta|$.

$\lvert\eta\rvert$ range	≤ 1.7	$1.7 < \lvert\eta\rvert \leq 2.3$	$\lvert\eta\rvert > 2.3$
jet p_T resolution uncertainty	5 %	9 %	15 %

Each generated event is shifted according to the Jet Energy Corrections as explained in section 4.3.1. Two additional samples are generated for each of the MC datasets in table 6.1 were the jet energy resolution is varied by either the up- or downwards variation from table 6.13.

6.4.3. Jet energy scale

The jet energy scale (JES) uncertainties refers to a whole group of independent uncertainties which affect the measured energy for a given particle in different ways. As listed in [72], these are:

Absolute: absolute scale uncertainty. Mainly an uncertainty in combined photon (EM) and Z→ $\mu\mu$ (tracking) reference scale and correction for FSR+ISR.

HighPtExtra: high pT extrapolation. Based on Pythia6 Z2/Herwig++2.3 differences in fragmentation and underlying event (FullSim).

SinglePion: high p_T extrapolation. Based on propagation of ±3 % variation in single particle response to PF Jets (FastSim).

Flavor: jet flavour (quark/gluon/charm/b-jet). Based on Pythia6 Z2/Herwig++2.3 differences in quark and gluon responses relative to QCD mixture (charm and b-jets are in between uds and g).

Time: JEC time dependence. Observed instability in the endcap region, presumed to be due to the EM laser correction instability for prompt 42X data.

RelativeJER- EC1, EC2, HF: eta-dependence uncertainty from jet pT resolution (JER). The JER uncertainties are assumed fully correlated for endcap within tracking (EC1), endcap outside tracking (EC2) and hadronic forward (HF).

RelativeFSR: eta-dependence uncertainty due to correction for final state radiation. Uncertainty increases toward HF, but is correlated from one region to the other.

RelativeStat- EC2, HF: statistical uncertainty in determination of eta-dependence. Averaged out over wider detector regions, and only important in endcap outside tracking (EC2) and in HF.

PileUp- DataMC, OOT, Pt, Bias, JetRate: uncertainties for pileup corrections.

- **DataMC** parametrizes data/MC differences vs eta in Zero Bias data.
- **OOT** estimates residual out-of-time pileup for prescaled triggers, if reweighing MC to unprescaled data.
- **Pt** covers for the offset dependence on jet pT (due to e.g. zero suppression effects), when the correction is calibrated for jets in the p_T=20-30 GeV range.
- **Bias** covers for the differences in measured offset from Zero Bias (neutrino gun) MC and from MC truth in the QCD sample, which is not yet fully understood.
- **JetRate** covers for observed jet rate variation versus the number of primary vertices in 2011 single jet triggers, after applying L1 corrections.

All together there are 16 individual jet energy scale uncertainties. They are summarised in a combined uncertainty, JES_Total.

For each of these 17 possible variations, a separate sample including either the up- or downwards variation was produced from the nominal samples listed in table 6.1. During the modification of the jets in each event, the missing transverse energy was adjusted as well, to rebalance the transverse energy sum in the event.

Tests were performed to see, whether the usage of the individual uncertainties yields a different result than what is retrieved by using the total uncertainty. This test was done on two simplified models. They only included the rate uncertainties and one the

16 individual JES uncertainties and the other the JES_Total as only shape uncertainties. Both models were used in the marginalisation procedure on toy-data and the width of the Bayesian posterior was calculated as the difference of the 16 % and the 84 % quantile.

It turned out, that the residual width was the same in both cases for the electron and the muon channel. For this reason only the total uncertainty variations are used further on.

6.4.4. Missing transverse energy

The missing transverse energy (MET) is rather large in the events of interest, because of the undetectable neutrino which escapes the CMS detector. This variable is varied by $\pm$ 10 %. The event needs to be rebalanced after MET is altered, i.e. the jets transverse momentum components (p_x, p_y) need to be adjusted to account for the added/removed MET.

6.4.5. b tagging

As stated in section 6.2.6, the b tagging scale factors are given with an uncertainty of about 2 - 10 %, depending on the p_T of the jet. This is also depicted as error bands in figure 6.7.

Separate samples are generated varying the b tag scale factor to either the up- or downwards end of the given uncertainty. This is done for the b jet scale factor SFb and the light jet scale factor independently. Hence this yields four additional sets of templates for each process. The nuissance parameter for this uncertainty are denoted by btagcb for the SFb variations and btaglight for the light jet scale factor variations.

The current model for the b tagging scale factors has only a discrete p_T binning and is constant over η. In order to model a possible different η and p_T dependence of the scale factor uncertainties, they are multiplied by Chebychev polynomials $T_n(f(p_T))$ and $T_m(g(|\eta|))$. The Chebychev polynomials of the first kind are given by

$$T_n(x) = \cos\left(n \arccos x\right). \tag{6.9}$$

p_T and η are mapped to the interval $[-1, 1]$ by the functions

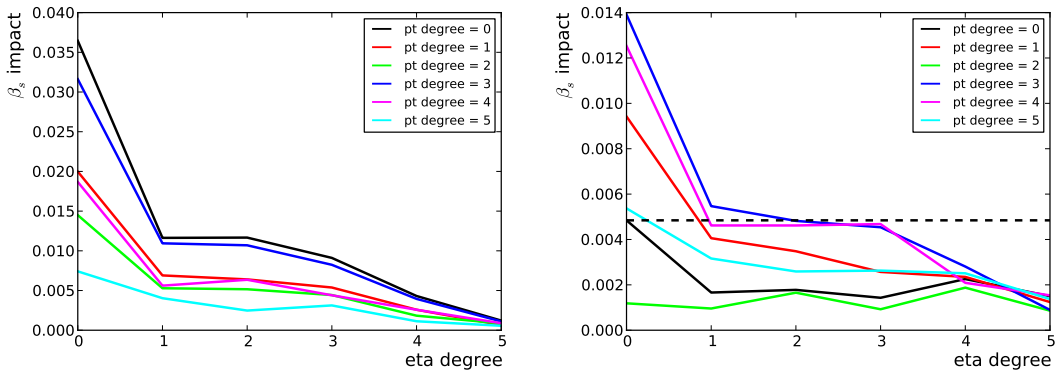
$$f(p_T) = -1 + 2 \cdot p_T / p_T^{max} \tag{6.10}$$

and

$$g(|\eta|) = -1 + 2 \cdot |\eta| / |\eta|^{max}. \tag{6.11}$$

The influence of polynomials up to $n, m = 0 \ldots 5$ was accounted for in form of varied NeuroBayes discriminator templates for each dataset in the statistical evaluation.

The original b tagging uncertainty is accounted for in the Bayesian marginalisation, while for the additional Chebychev variations the residual impact was estimated using maximum likelihood fits on the shifted toy-data, as presented above. It turned out, that the additional variations does not yield larger uncertainties than the normal uncertainty, as depicted in figure 6.12(b).



(a) Impact on β_s with current statistics.

(b) Impact on β_s with 10 times higher statistics than currently available.

Figure 6.12.: Study of the additional dependence on the b tagging scale factors on p_T and η. pT-degree and eta-degree are the used parameters n in the Chebychev polynomials (6.9). If there were an additional dependence, than the nominal uncertainty would not yield the largest variation. But as can be seen in figure 6.12(a), the highest observed shift is due to the nominal uncertainties with p_T and η Chebychev parameters 0. This changes if one assumes to have ten times more statistics in figure 6.12(b), than it is currently the case. [91]

If the here presented analysis were sensitive to additional dependence, than this would be visible in figure 6.12(a) as a different point that yields the highest observed shift than the one with p_T and η degree 0, which is the nominal case.

Therefore there is no need to quote an additional p_T or η dependent b tagging uncertainty for this analysis.

6.4.6. Trigger efficiencies

The leptonic part of the triggers is accounted for as a pure rate systematic. But hadronic parts of the cross triggers do introduce some shape uncertainty.

The parametrisation of the two dimensional CentralJet30_BTagIP uses Gompertz functions which have three free fit parameters. Each of these parameters is fitted with some uncertainty. The three parameters in the Gompertz function are decorrelated and each of the decorrelated parameters is shifted to either end of its systematic uncertainty. The resulting event weights are calculated as presented in section 6.2.5. This leads to additional six event samples that are used as systematic shape uncertainties in the Bayesian marginalisation.

6.4.7. Event pileup

Generated events need to be pileup reweighted, as is explained in sections 4.3.2 and 6.2.6. The inelastic cross section is measured to be 71.0 mb with an uncertainty of ± 5.7 mb, which is accounted for in form of alternative three dimensional pileup distributions, as given in figures 6.4 and 6.5. Samples for either pileup weight for an inelastic cross section 65.3 mb, 71.0 mb and 76.7 mb [100] are generated to account for this uncertainty.

6.4.8. Q^2 scale

The parton density functions depend on the energy scale $\mu = Q^2$ as explained in section 2.2.1. This energy scale is usually set to the mass of the top quark, or to the mass of the W or Z boson if they are produce. This is not with out some arbitrariness. Therefore the effect of different Q^2 scales is taken as a systematic uncertainty.

Dedicated samples with double and halve the nominal Q value are available for all single top samples (t-channel, s-channel and associated tW production), as well as for $t\bar{t}$. The parton shower simulated with Pythia was configured to use an $k_\perp^2$ evolution scale for α_s either 0.25 (4) times the default value for space-like parton showers for the scale down (scale up) sample.

While there are also samples available for W+jets and Z+jets, they can not be used directly, since they are only very few events in these samples, as is listed in table 6.2. Instead the scale uncertainty is neglected for the Z+jets sample, because the Z+jets contribution is expected to be small and systematic uncertainty on this is expected to be insignificant.

Q^2 scale reweighting for W+jets

For W+jets a reweighting method is applied on the blended W+$n$jets sample (sec. 6.1.1) to model this uncertainty. The required event weights are calculated by considering

event properties on generator level in a three step procedure, was presented in [93]. The CTEQ6ll pdf sets from LHAPDF [106, 107] have been used.

These steps are:

1. Hard Matrix element: Calculate $\alpha_s(Q^2)$ and $\alpha_s(\frac{1}{4}\cdot Q^2)$, $\alpha_s(4\cdot Q^2)$ for each event in the default sample. The first part of the event weight is then given by

$$w_{down} = \left(\frac{\alpha_s(\frac{1}{4}\cdot Q^2)}{\alpha_s(Q^2)}\right)^N$$

$$w_{up} = \left(\frac{\alpha_s(4\cdot Q^2)}{\alpha_s(Q^2)}\right)^N$$

   where N is the number of additional partons.

2. Parton density function variation: The PDFs for the two initial partons are evaluated at the shifted factorisation scales.

$$w_{[down,up]} = \frac{pdf(x_1, id_1, Q\cdot[\frac{1}{2}, 2])}{pdf(x_1, id_1, Q)} \cdot \frac{pdf(x_2, id_2, Q\cdot[\frac{1}{2}, 2])}{pdf(x_2, id_2, Q)}$$

3. Neural network based reweighting: Additional reweighting is necessary since the Pythia showering information are not accessible on the samples anymore. Thus a neural network is trained on those variables, which are not well described after the first two reweighting steps (rapidities, ΔR and invariant masses). The neural networks output is used as third weight to account for the last residual differences.

Since the rate variation in the W+jets Q^2 scale systematics are up to 200 % compared to the nominal templates, it was decided to disentangle the rate and the shape uncertainty in this case. The rate uncertainty is accounted for by the large prior width for the different W+jets cross sections, while the residual impact on β_s due to the shape uncertainty is evaluated by the separate maximum likelihood fit method described above.

Since there is no reason to assume a lepton flavour dependence in this uncertainty, the Q^2 scale uncertainty is derived in the combined fit only and used also for the measurements in the electron and the muon channel.

6.4.9. Matching threshold

The matching threshold for the transition from the hard matrix element simulation to the soft radiation and showering simulation is also an possible systematic bias. This is accounted for by using samples with either 0.5 and 2 times the nominal scale factor. They uncertainty is present in the MadGraph samples only, which are $t\bar{t}$, W+jets, and Z+jets. The latter two have again to few events. For Z-jets this uncertainty is hence thus ignored. Since in this case there is no reweighting method available for W+jets, the cross section for this process was disentangled for the different jet multiplicities. This together with the large prior uncertainties well covers the observed event rate variations due to this uncertainty.

Also in this uncertainty there is no lepton flavour dependence assumed and the result for the combined fit is applied also on the muon and electron channel result.

6.4.10. Parton density function

Another source of systematic uncertainty arises from the parton density functions described in 2.2.1. This uncertainty is studied by using the systematic variations for the CTEQ6ll pdf set from LHAPDF [106, 107], which consists of 22 independent eigenvectors that each yield different event weights. The expected effect on the templates was found to be very small. Due to this, and because 22 additional nuisance parameter could make the model unstable, this uncertainty is excluded from the marginalisation and instead the impact on β_s for each of the eigenvectors is evaluated by the maximum likelihood fit method on the shifted toy-data. The individual variations on β_s are then symmetrised and quadratically added.

6.4.11. Top quark mass

The influence of the top quark mass to the measured value has been studied by using samples which were generated with shifted top quark mass. Values of $m_t = 166.5\,\mathrm{GeV/c^2}$ and $m_t = 178.5\,\mathrm{GeV/c^2}$ are used for processes which include a top quark. These are listed in table 6.2. Also this uncertainty is not marginalised but the observed impact on the toy-data fit is quoted.

6.4.12. Single top generator

The default MC generator for single top events in this analysis is Powheg. Alternative samples generated with CompHEP [23, 94] (see also section 4.1.4) are used as alternative single top t-channel samples. Toy-data are build where the standard Powheg single top t-channel sample is replaced by the one generated with CompHEP and the β_s shift is derived using the default model, which includes the Powheg template.

6.4.13. Luminosity

The measurement of the instantaneous luminosity for each is crucial to deduce the integrated luminosity for the whole selected sample. For this analysis the so called `pixelLumiCalc` tool was used, which uses information from the track pixels to improve the precission on the measurement of the instantaneous luminosity. The systematic uncertainty for the measurement of the luminosity is taken to be 2.2 % [51]. It is not included in the statistical inference directly, but it is qudratically added to the final uncertainty.

6.5. Cross section measurement

In this section the results on the measurement of the single top t-channel cross section is given. First the unmarginalised uncertainties and their residual impact on the signal yield is presented in section 6.5.1. Then the central cross section value together with the residual uncertainty after the marginalisation is given in section 6.5.2.

Finally the complete result and a comparison with other measurements is presented in 6.5.5.

6.5.1. Unmarginalised uncertainties

Most of the uncertainties mentioned in section 6.4 are marginalised. Only four of them are treated in a different way to estimate their impact on the final result, as is explained in the beginning of this chapter.

The uncertainties and the reasons for them not being marginalised are:

Q^2 scale: The uncertainty on the Q^2 scale is modelled with large variations of factor two and one half on Q. However it is not clear, whether interpolating between the up, down and nominal templates does cover the full spectrum of possible variations, which is a prerequisite to apply the marginalisation. Hence this uncertainty is studied separately with independent nuisance parameters for each of the processes. The observed β_s shift for each of the processes are listed the following table. As already stated in section 6.4.8, for the W+jets component only the shape uncertainty is studied separately and the rate uncertainty is covered by large enough prior uncertainty widths for the different W+jets cross section nuisance parameters. The reason is, that the observed variations in the event rate are up to 240 %, as can be seen in table 6.18 and 6.19.

direction	t-channel	s-channel	tW associated	$t\bar{t}$	W+jets (shape only)
↑	0.34%	0.06%	0.28%	0.46%	+5.6%
↓	0.52%	-0.18%	-0.05%	2.00%	-0.0%

To be conservative with this uncertainty, the largest deviation for all processes is are added quadratically. The final result as residual uncertainty due to the Q^2 scale uncertainty is therefore 5.98 %.

Matching threshold: Also this uncertainty is varied arbitrarily by a factor of two and one half on the central value (top quark mass). For the same reasons given above, it is excluded from the marginalisation. The only remaining process for which decent alternative templates due to this uncertainty are available is $t\bar{t}$. The observed shifts are +0.21% and +1.49 % for the down- and the upwards variation. The larger one is quoted as systematic uncertainty on the final result.

Top quark mass: The available datasets vary the top quark mass to 166.5 GeV/c^2 and to 178.5 GeV/c^2. This is a shift of ≈ 6 GeV/c^2 in either direction around the current world average of 173.2 ± 0.9 GeV/c^2[103], which in turn is more than five standard

deviations of the current uncertainty. Since this would mean an overestimation of the expected impact, this uncertainty is also excluded from the marginalisation. In the three statistical inferences the following shifts are observed on β_s.

top mass	muon	electron	combined
178.5 GeV/c^2	-1.11%	3.71%	1.13%
166.5 GeV/c^2	0.89%	-0.78%	-0.31%
average shift	1.00%	2.25%	1.17%

Considering the small deviation that are observed in the different measurements for the large variations of the top mass, it can safely be assumed, that the sensitivity to the actual top mass uncertainty is negligible. This uncertainty is therefore not included in the final result.

Parton density function: The parton density function uncertainties are parametrised with 22 independent eigenvectors. Considering the small effect on the discriminator templates, adding these 22 nuisance parameters to the model could make the parameter estimation unstable or underestimate the effect of this uncertainty. Hence the individual impact for each eigenvector is studied in a maximum likelihood fit and the symmetrised uncertainties are quadratically added. In the muon channel the result is 1.041%, in the electron channel 1.162% and in the combined inference the square root of the quadratic sum of PDF uncertainties is 0.976%.

Single top MC generator The dependence on the Monte Carlo generator is studied by using alternative samples, that are generated with the CompHEP generator [23, 94]. The toy-data are build using the CompHEP generated single top t-channel events instead of the default Powheg sample. The latter is then still used in the model, which is fitted to the toy-data.

In the electron channel a shift of -8.58 % is observed, while in the muon channel it is -8.17 %. For the combined fit in both lepton channels a shift of -8.37 % is measured.

6.5.2. Marginalised uncertainties and central values

All rate uncertainties together with the rest of the shape uncertainties are used to calculate the marginal Bayesian posterior. Before being applied on the actual data, this method is studied on toy-data to test its robustness and estimate the impact of each individual uncertainty.

First the nominal fit value for β_s, as the 50 % quantile for the marginal Bayesian posterior, is determined by using a the stacked distribution of all nominal templates as toy-data distribution. This is a fit of the model to itself and the expected result is 1.0 with only minor statistical fluctuations. Also the 84 % and the 16 % quantile are derived in the same way and give the expected upper and lower uncertainty for the measurement. Based on 100 Bayesian inferences the nominal values for β_s in the different channels are measured as listed in table 6.14.

These results for the expected uncertainty already include two kinds of statistical uncertainties.

Table 6.14.: Central values for the nominal fit and expected uncertainties

channel	nominal value	expected uncertainty
muon	1.017133 +- 0.002032	$^{+10.00\,\%}_{-9.90\,\%}$
electron	1.029743 +- 0.011554	$^{+11.35\,\%}_{-11.19\,\%}$
combined	1.019887 +- 0.006444	$^{+7.99\,\%}_{-7.59\,\%}$

1. The statistical uncertainty which arises from the finite number of data events. This is included by assuming Poison statistics in each bin of the toy-data distribution.

2. The statistical uncertainty on the MC template distribution. As given in table 6.8, the ratio between the available amount of MC events and the expected number of events after weighting is about 10:1 for most of the processes in all jet/tag categories. Still the uncertainty on the template stack which arises from the finite number of MC events is taken into account using the Barlow-Beeston method[20], which is implemented in `theta`. The Barlow-Beeston method adds for the statistical uncertainty in each template bin another independent model parameter

$$\mu_i \to \mu_i + \gamma_i$$

where μ_i denotes the number of entries in bin i in the model and γ_i has in each bin i a Gaussian prior around zero and a width according to the (MC) statistical uncertainty. Hence there is only one additional nuisance parameter in each bin of the model for all processes combined and not for each template individually, since this would add to many nuisance parameters.

The dependence on the γ parameters is analytically maximized out and the usual statistical inference is performed on the modified likelihood function with the remaining parameters.

The individual impact for each uncertainty is studied in a similar way as it is done for the unmarginalised uncertainties with the maximum likelihood fit. Also here either uncertainty is studied individually and the distribution width for the nuisance parameter of all the others is set to zero, i.e. they can not be varied. Then the uncertainty under investigation is shifted to either side $(+1\sigma, -1\sigma)$ and two different fits are performed. One were also the width of the last uncertainty is fixed, and one were this width is kept free and this uncertainty can be marginalised. The results are listed in tables 6.15, 6.16 and 6.17, respectively.

Table 6.15.: List of expected uncertainties after the marginalisation in the muon channel. In the left two columns the residual uncertainty for the systematic is given when all uncertainties are marginalised. In the right two columns the uncertainty under investigation is excluded from the marginalisation. The difference between both methods gives an estimate on how much the uncertainty can be constrained.

systematic	up marg.	down marg.	up unmarg.	down unmarg.
btagcb	-4.41%	4.09%	-7.20%	6.61%
btaglight	1.02%	-1.23%	6.12%	-9.05%
diboson	0.01%	-0.30%	0.11%	-0.31%
JER	0.67%	-0.23%	1.50%	-0.83%
JES	-2.79%	0.74%	-11.90%	6.61%
MET	-0.12%	-0.18%	0.11%	-1.44%
muid	-2.87%	3.05%	-2.77%	2.99%
pileup	1.14%	-1.10%	1.65%	-1.26%
qcdmu	-0.97%	1.29%	-0.89%	3.49%
s-channel	-0.33%	0.49%	-0.16%	0.20%
$t\bar{t}$	0.52%	-0.83%	13.84%	-15.73%
tW-channel	-0.02%	-0.14%	-0.13%	0.16%
W+b-jets_2j	1.72%	0.46%	7.39%	-22.09%
W+b-jets_3j	-0.39%	0.01%	-3.30%	-0.21%
W+b-jets_4j	-0.33%	0.60%	-0.78%	1.08%
W+c-jets_2j	-2.63%	0.57%	-6.75%	17.22%
W+c-jets_3j	-0.49%	0.92%	-0.98%	1.59%
W+c-jets_4j	-0.15%	0.37%	-0.46%	0.18%
W+light-jets_2j	0.58%	-0.88%	0.26%	-1.39%
W+light-jets_3j	0.13%	-0.02%	0.09%	-0.05%
W+light-jets_4j	-0.18%	0.10%	0.09%	0.20%
Z+jets	0.42%	-0.31%	0.16%	-0.62%

Table 6.16.: List of expected uncertainties after the marginalisation in the electron channel. In the left two columns the residual uncertainty for the systematic is given when all uncertainties are marginalised. In the right two columns the uncertainty under investigation is excluded from the marginalisation. The difference between both methods gives an estimate on how much the uncertainty can be constrained.

systematic	up marg.	down marg.	up unmarg.	down unmarg.
btagcb	-4.02%	3.54%	-6.55%	6.41%
btaglight	0.45%	-0.59%	2.38%	-5.91%
diboson	0.27%	-0.22%	0.06%	-0.32%
eleid	-2.82%	3.13%	-3.08%	2.76%
JER	0.47%	-0.32%	1.73%	-0.06%
JES	-4.10%	-0.29%	-16.30%	8.27%
MET	-1.30%	0.15%	-4.56%	2.39%
pileup	1.31%	-1.19%	1.72%	-1.43%
qcdele	-1.44%	1.14%	-1.96%	4.25%
s-channel	-0.24%	0.17%	-0.17%	-0.02%
$t\bar{t}$	0.63%	-0.98%	19.81%	-19.15%
ttobtag1	0.33%	0.00%	0.34%	0.06%
ttobtag2	-0.22%	0.35%	-0.02%	0.45%
ttobtag3	-0.03%	0.21%	-0.01%	0.23%
tW-channel	0.02%	0.00%	0.10%	0.01%
W+b-jets_2j	1.49%	-0.43%	5.15%	-18.20%
W+b-jets_3j	-1.12%	0.05%	-3.81%	1.45%
W+b-jets_4j	-0.57%	0.54%	-0.93%	1.49%
W+c-jets_2j	-3.53%	0.46%	-8.17%	18.81%
W+c-jets_3j	-0.71%	0.10%	-1.06%	0.87%
W+c-jets_4j	-0.25%	0.43%	-0.30%	0.47%
W+light-jets_2j	-0.06%	0.22%	-0.06%	0.08%
W+light-jets_3j	0.03%	0.17%	0.11%	0.35%
W+light-jets_4j	0.18%	0.04%	0.05%	0.15%
Z+jets	-0.24%	0.11%	-0.14%	0.05%

Table 6.17.: List of expected uncertainties after the marginalisation in the combined model with both lepton channels. In the left two columns the residual uncertainty for the systematic is given when all uncertainties are marginalised. In the right two columns the uncertainty under investigation is excluded from the marginalisation. The difference between both methods gives an estimate on how much the uncertainty can be constrained.

systematic	up marg.	down marg.	up unmarg.	down unmarg.
btagcb	-3.35%	3.29%	-7.57%	6.40%
btaglight	0.83%	-0.81%	6.85%	-11.45%
diboson	0.04%	-0.25%	0.34%	-0.45%
eleid	-1.29%	1.45%	-2.55%	1.99%
JER	0.46%	-0.05%	1.50%	-0.37%
JES	-1.74%	-0.51%	-13.63%	9.11%
MET	-0.06%	0.27%	-1.18%	1.66%
muid	-1.77%	1.53%	-2.16%	2.98%
pileup	1.37%	-0.72%	1.97%	-1.32%
qcdele	-0.87%	0.72%	-1.56%	2.84%
qcdmu	-0.12%	0.32%	-0.59%	1.13%
s-channel	0.22%	0.18%	-0.14%	0.22%
$t\bar{t}$	0.73%	-0.40%	20.01%	-17.88%
ttobtag1	0.20%	-0.04%	0.38%	-0.03%
ttobtag2	0.09%	-0.03%	0.28%	0.14%
ttobtag3	-0.09%	-0.15%	0.14%	0.24%
tW-channel	0.08%	-0.00%	0.29%	0.21%
W+b-jets_2j	1.44%	0.77%	7.39%	-22.90%
W+b-jets_3j	-0.34%	-0.03%	-3.11%	-0.88%
W+b-jets_4j	-0.35%	-0.07%	-0.75%	1.22%
W+c-jets_2j	-2.86%	0.10%	-8.30%	20.92%
W+c-jets_3j	-0.53%	0.19%	-0.88%	1.64%
W+c-jets_4j	0.04%	0.08%	-0.14%	0.02%
W+light-jets_2j	0.18%	-0.17%	0.06%	0.17%
W+light-jets_3j	0.16%	0.04%	0.08%	0.12%
W+light-jets_4j	-0.07%	0.02%	-0.15%	0.25%
Z+jets	0.38%	-0.66%	0.25%	-0.27%

6.5.3. Event yield changes by systematic uncertainties

All systematic shape uncertainties also change the expected number of events in the selected sample. This impact on the event rates in the 2j1t category is listed in table 6.18 for the muon channel and table 6.19 for the electron channel. Missing values indicate that the uncertainty was not modelled for a certain process, like the matching uncertainty for the Powheg generated samples As can be seen, most of the uncertainties affect the expected rate by only a few percent, but the Q^2 scale uncertainty yields a rather large shift in the W+jets samples of up to 240 %. It was already stated above, that due to this fact, the scale uncertainty shape variation is disentangled from the induced rate uncertainty for the W+jets samples.

Table 6.18.: Rate impact of systematic shape uncertainties in the 2j1t category for the muon channel

process		btagcb	btaglight	MET	scale	JES	JER	matching	mass
t-channel	↑	1.63%	1.15%	3.55%	0.98%	-4.20%	-0.38%	—	2.40%
	↓	-1.65%	-1.16%	-4.16%	-1.34%	-0.81%	0.24%	—	-1.48%
s-channel	↑	1.09%	0.46%	3.60%	-1.50%	-3.21%	-0.64%	—	-2.69%
	↓	-1.25%	-0.47%	-5.10%	-3.62%	-1.05%	1.03%	—	-6.62%
associate tW	↑	0.58%	1.87%	4.45%	4.38%	-8.75%	-0.12%	—	-20.18%
	↓	-0.61%	-1.89%	-5.66%	-0.70%	4.10%	0.12%	—	-21.79%
$t\bar{t}$	↑	0.66%	0.93%	3.13%	10.10%	-11.62%	-0.07%	-0.60%	-2.87%
	↓	-0.75%	-0.94%	-3.81%	-6.92%	7.52%	0.08%	-3.96%	5.70%
W+b-jets	↑	0.96%	2.60%	4.10%	-24.81%	-2.09%	-2.09%	3.15%	—
	↓	-1.00%	-2.61%	-5.23%	175.68%	-3.50%	-3.50%	27.70%	—
W+c-jets	↑	1.40%	11.27%	5.06%	-13.24%	7.50%	7.50%	-3.20%	—
	↓	-1.40%	-11.32%	-6.52%	240.68%	-3.64%	-3.64%	32.71%	—
W+light-jets	↑	0.00%	-2.61%	4.40%	-13.46%	3.19%	3.19%	8.75%	—
	↓	0.00%	-1.17%	-4.86%	158.85%	-1.29%	-1.29%	22.15%	—
diboson	↑	1.44%	5.62%	4.93%	—	-0.31%	-0.40%	—	—
	↓	-1.51%	-5.73%	-6.34%	—	-1.98%	-0.01%	—	—

Table 6.19.: Rate impact of systematic shape uncertainties in the 2j1t category for the muon channel

					cross trigger turn-on parameter							
process		btagcb	btaglight	MET	1	2	3	scale	JER	JES	matching	mass
t-channel	↑	1.46%	1.15%	6.11%	0.10%	0.37%	0.22%	0.59%	0.23%	1.60%	—	8.10%
	↓	-1.48%	-1.15%	-3.67%	0.16%	-0.10%	0.06%	0.35%	-0.14%	1.01%	—	-7.57%
s-channel	↑	1.09%	0.40%	6.39%	0.31%	0.55%	0.39%	1.66%	-0.05%	-0.31%	—	-0.57%
	↓	-1.25%	-0.40%	-4.09%	0.33%	0.10%	0.27%	-2.52%	0.42%	-0.91%	—	-8.50%
associated tW	↑	0.42%	1.61%	4.42%	0.17%	0.42%	0.26%	0.66%	0.04%	-4.00%	—	-21.19%
	↓	-0.45%	-1.63%	-1.07%	0.20%	-0.04%	0.14%	-4.29%	-0.37%	4.54%	—	-26.46%
$t\bar{t}$	↑	0.66%	0.92%	1.91%	0.27%	0.50%	0.33%	11.03%	0.36%	-8.12%	-3.39%	-2.98%
	↓	-0.76%	-0.92%	-0.12%	0.26%	0.04%	0.23%	-6.10%	0.20%	7.75%	-3.98%	-0.59%
W+b-jets	↑	0.72%	1.98%	12.67%	0.17%	0.47%	0.35%	-35.63%	1.32%	5.07%	-8.75%	—
	↓	-0.76%	-1.98%	-7.52%	0.30%	-0.01%	0.11%	91.90%	0.26%	3.95%	41.68%	—
W+c-jets	↑	1.29%	11.21%	12.74%	0.22%	0.53%	0.32%	-37.78%	1.06%	17.34%	9.03%	—
	↓	-1.30%	-11.28%	-6.17%	0.29%	-0.08%	0.14%	96.21%	-1.73%	1.80%	6.79%	—
W+light-jets	↑	0.00%	-5.02%	8.19%	0.26%	0.53%	0.28%	-35.82%	0.51%	15.60%	15.95%	—
	↓	0.00%	0.34%	-1.09%	0.18%	-0.10%	0.16%	92.24%	-0.06%	1.34%	46.24%	—
diboson	↑	1.17%	5.26%	9.24%	0.22%	0.52%	0.36%	—	0.51%	6.69%	—	—
	↓	-1.23%	-5.39%	-4.66%	0.33%	-0.01%	0.15%	—	-0.67%	2.33%	—	—

6.5.4. Shape comparison for selected sytematic uncertainties

The b tag uncertainty btagcb and the jet energy scale JES systematic uncertainties yield the largest residual impact after the marginalisation as can be seen in tables 6.15, 6.16 and 6.17 in the left columns. In figure 6.13 the ratio between the b tag uncertainty btagcb and the nominal samples is given for the up and down variation in dependence of the NeuroBayes discriminator value. The up variation is indicated by triangle markers with the tip at the top and the down variation with the tip at the bottom accordingly. It can be seen, that for all processes there is a strong correlation between this uncertainty and the neural network discriminator value, which explains the large residual uncertainty after the marginalisation of at least 3.5 %.

For the JES uncertainty the according plots are given in figure 6.14. There is no strong correlation between the shape of the ratio and the discriminator value but the observed rate changes are larger than for the btagcb uncertainty, as can also be seen in tables 6.18 and 6.19, hence the large observed residual impact on β_s.

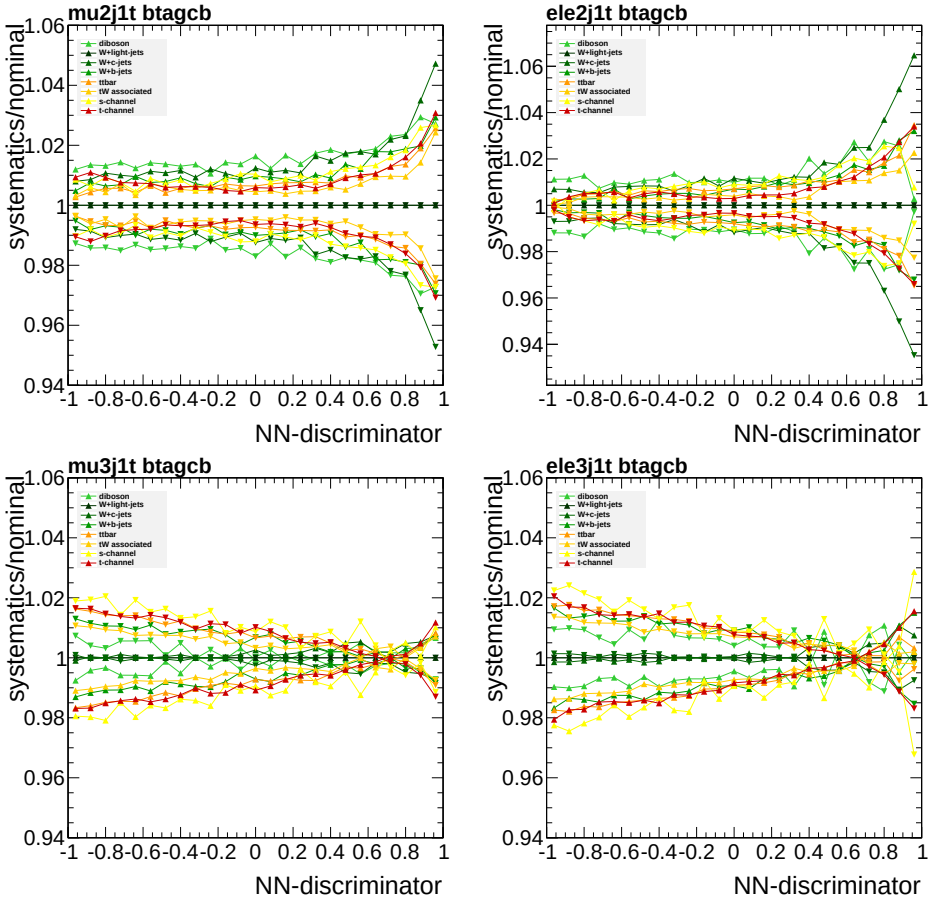


Figure 6.13.: Ratio between the b-tagging systematic up/down variation and the nominal sample in the signal enriched regions. The direction of the triangle indicates the up or down shift of the systematic.

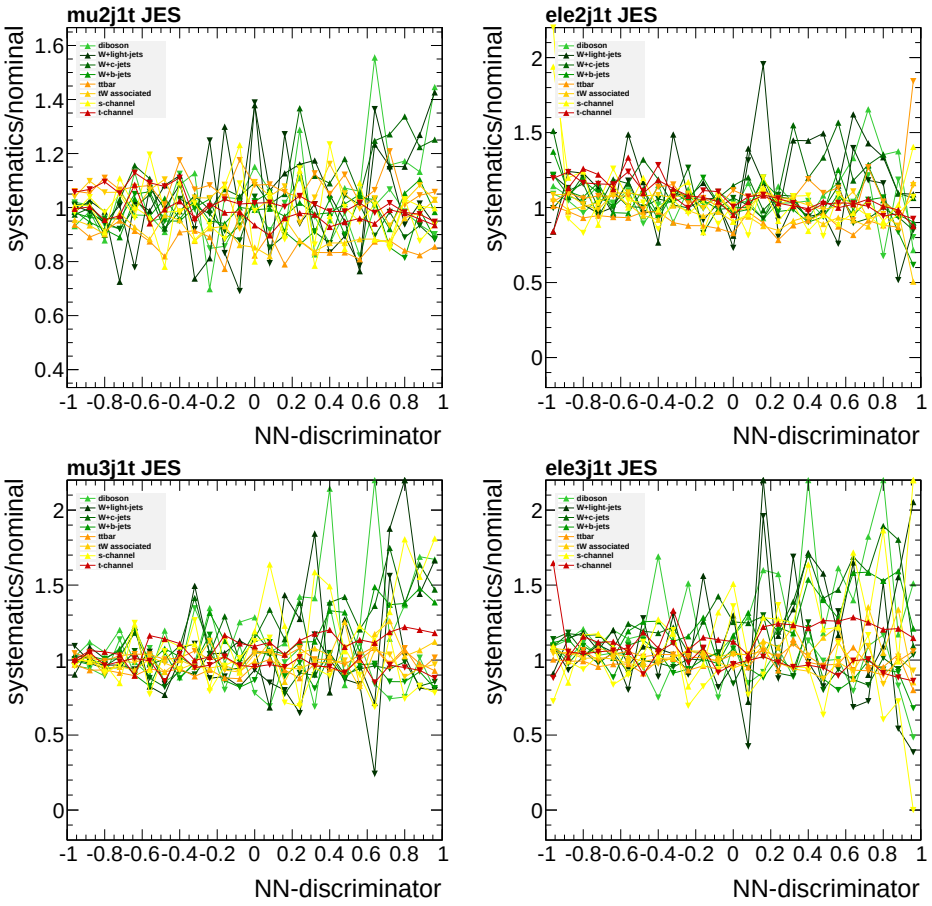


Figure 6.14.: Ratio between the jet energy scale (JES) systematic up/down variation and the nominal sample in the signal enriched regions. The direction of the triangle indicates the up or down shift of the systematic.

The Q^2 scale uncertainty yields a large impact on β_s of 5.98 %, where the largest contribution is due to the W+jets events, as explained in section 6.5.1. In figure 6.15 the ratio between the normalised Q^2 scale up or down discriminator distribution and the normalised nominal discriminator distribution for each process is given. It can be seen, that the ratio between the systematic and the nominal discriminator distributions is higher for high discriminator values in the W+jets samples. Unlike for the btagcb uncertainty presented in figure 6.13, this is not the case for the single top quark t-channel events. Hence the large observed shift in β_s for the W+jets Q^2 scale uncertainty can be explained by this difference between signal and the W+jets events with respect to this uncertainty.

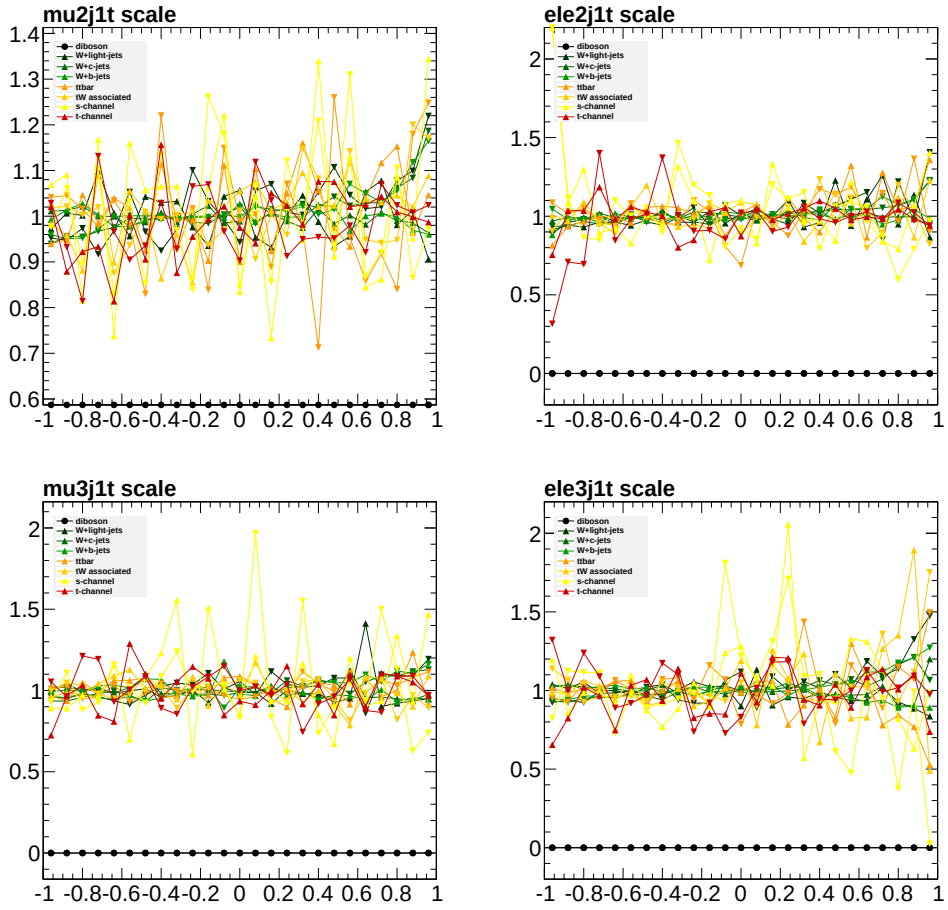


Figure 6.15.: Ratio between the normalised Q^2 scale matching uncertainty up/down variation and the normalised nominal sample in the signal enriched regions. The direction of the triangle indicates the up or down shift of the systematic. The normalisation was done to study the pure shape effect, independent of the large rate shifts that are observed in the W+jets samples. There are no samples modelling the Q^2 scale systematic for the diboson events, hence the ratio is flat.

6.5.5. Single top t-channel cross section results

The Bayesian inference is performed with the data samples and the 50 % quantile is calculated as best parameter estimate for β_s. The 84 % and 16 % quantile are quoted as upper and lower boundaries for the 1σ credible interval. The values for the so derived single top quark production cross section in the t-channel are given in table 6.20.

Table 6.20.: Measured single top t-channel cross sections for all three statistical inferences. The quoted errors are the marginalised only.

		uncertainty	
channel	measured cross section [pb]	absolute [pb]	relative [%]
muon	70.09	$^{+7.19}_{-6.88}$	$^{+10.25}_{-9.82}$
electron	67.40	$^{+8.08}_{-7.79}$	$^{+11.98}_{-11.55}$
combined	68.95	$^{+5.20}_{-5.74}$	$^{+7.55}_{-8.32}$

The given uncertainties do already include the before mentioned two sources of statistical uncertainty, the data statistics and the statistics of the model templates via the Barlow Beeston light method. The observed relative uncertainties are comparable to the expected ones, listed in table 6.14.

To the uncertainties quoted in table 6.20 the unmarginalised uncertainties listed in section 6.5.1 have to be added quadratically. An overview is given in table 6.21.

Table 6.21.: Overview of the residual systematic uncertainties in each channel and their total quadratic sum. All values are given in percent to the central value.

Marginalised down	-9.70	-11.47	7.80
Marginalised up	8.44	11.32	7.11
Q^2 scale	5.98	5.98	5.98
Matching thershold	1.49	1.49	1.49
PDF	1.04	1.16	0.98
Signal generator	-8.17	-8.58	-8.37
Luminosity	2.20	2.20	2.20
Total down	-14.31	-15.79	-13.22
Total up	13.49	15.68	12.82

This yields the final results which are listed in table 6.22.

Table 6.22.: Measured single top t-channel cross sections for all three statistical inferences. The quoted errors are the total errors including the marginalised and the unmarginalised.

		total uncertainty	
channel	measured cross section [pb]	absolute [pb]	relative [%]
muon	70.09	$^{+10.29}_{-10.09}$	$^{+14.69}_{-14.39}$
electron	67.40	$^{+10.90}_{-10.68}$	$^{+16.17}_{-15.85}$
combined	68.95	$^{+9.69}_{-9.33}$	$^{+14.06}_{-13.53}$

$|Vtb|^2$ extraction

The CKM-matrix element $|Vtb|^2$ can be extracted from the measured cross section in the combined channel as

$$|V_{tb}|^2_{\text{measured}} = \frac{\sigma_t^{\text{measured}}|V_{tb}^{\text{SM}}|^2}{\sigma_t^{\text{SM}}} = 1.068^{+0.107}_{-0.113} \tag{6.12}$$

Here the assumptions are made, that $|V_{tb}^{\text{SM}}|^2| \approx 1$ and $|V_{tb}^{\text{SM}}|^2 \gg |V_{td}^{\text{SM}}|^2 + |V_{ts}^{\text{SM}}|^2$. The derived value for $|V_{tb}|$ is then given by

$$|V_{tb}| = \sqrt{\frac{\sigma_t^{\text{measured}}|V_{tb}^{\text{SM}}|^2}{\sigma_t^{\text{SM}}}} = 1.033^{+0.052}_{-0.055} \tag{6.13}$$

Error propagation is used to take into account the uncertainty for the measured and the predicted cross sectiuons.

The derived value is in good agreement with the current world average value as given in the CKM-matrix (2.4).

Goodness of fit

In order to visualize how well the agreement between the templates and the data distribution is for the derived parameter set, also maximum likelihood fits were performed to data. These yield the same central values for β_s as the Bayesian marginalisation and also the best fitting value for all nuisance parameters can be derived. This also includes the template morphing due to shape uncertainties, if necessary. The best fitting distributions for the 2j1t samples in both lepton channels are shown exemplary in figure 6.16.

For the other categories the distributions can be found in appendix B.

Result comparison

The single top t-channel cross section is measured in this analysis as $68.95^{+9.69}_{-9.33}$ pb in the combined fit. This is in good agreement with the predicted value from [74] which is

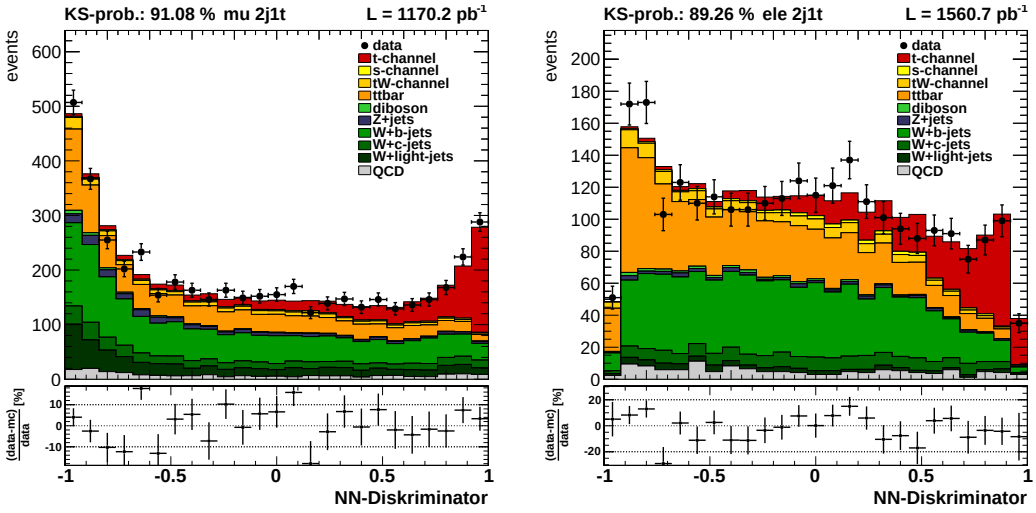


Figure 6.16.: Best fitting distributions of the NeuroBayes discriminator for the 2j1t category samples in the muon channel (left) and the electron channel (right).

$64.57^{+2.09\ +1.51}_{-0.71\ -1.74}$ pb for $\sqrt{s} = 7\,\text{TeV}$ LHC collisions. In figure 6.17 the retrieved value from the combined measurement is shown together with the predicted cross section shape in dependence of $\sqrt{s}$ as published in [27, 75].

The measured value for $|V_{tb}| = 1.033^{+0.052}_{-0.055}$ is also in good agreement with the current CKM matrix element value of $0.999133^{+0.000044}_{-0.000043}$ [69].

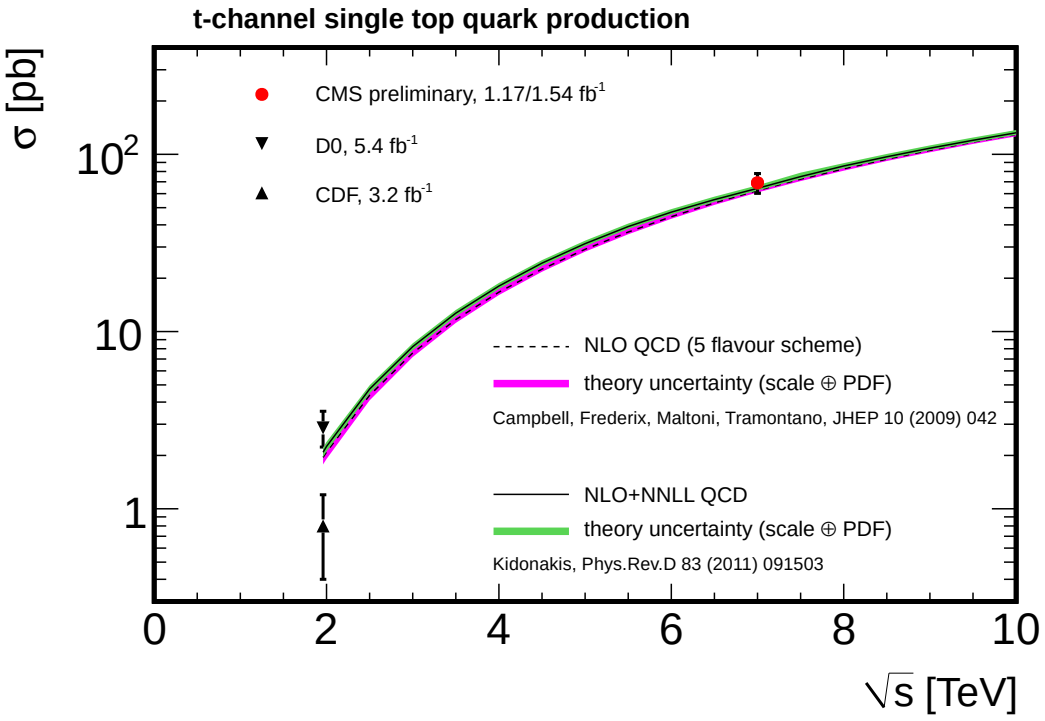


Figure 6.17.: Result for the combined single top t-channel cross section fit together with the theoretically predicted cross section shape and two Tevatron results.

7. Conclusion and outlook

The top quark was the last elementary particle to be newly discovered by physicists. It had been predicted to exist 22 years already [77], until it was first observed in $p\bar{p}$ collisions at the Tevatron by the CDF and DØ experiments in 1995 [10]. Another 14 years later the electroweak production of single top quark events has been observed by the same experiments. This was again a big success for the standard model of particle physicists, because also this process was predicted long before it was observed.

The standard model not only predicts the existence of certain processes, but it also gives quantitative statements about the production cross section. For the Tevatron the combined cross section for single top t-channel and s-channel events is calculated as $\sigma_{ts} = 3.12 \pm 0.09\,\text{pb}$ [74]. The latest measurement by the CDF collaboration for this quantity was published in [31] as $3.04^{+0.57}_{-0.53}\,\text{pb}$, while the DØ collaboration measured this quantity to be $3.43^{+0.73}_{-0.74}\,\text{pb}$ [7]. Both values agree well with the theoretical prediction. In [7] also the t-channel cross section at the Tevatron was published separately as $2.86^{+0.69}_{-0.63}\,\text{pb}$ which is also in good agreement with the predicted value of $1.04^{+0.00}_{-0.02} \pm 0.06\,\text{pb}$ for top and anti-top production in the t-channel.

In 2010, the same year as the first observation of single top events at the Tevatron, the Large Hadron Collider finally started running and within a short time it produced hadron collisions with before unprecedented center of mass energies of $\sqrt{s} = 7\,\text{TeV}$. Although it was mainly designed and planed to discover or rule out the Higgs boson and beyond the standard model theories like Super Symmetry, also all already known standard model particles and processes had to be rediscovered by the two large experiments at the LHC, ATLAS and CMS. This has already been done successfully for all particles, also the top quark [67]. Hence precision measurements on the standard model parameters have to be performed. One of these is the measurement of the single top production cross section in the t-channel.

The cross section for single top t-channel events is measured in this thesis using data from the CMS detector at the LHC. These data were taken in 2011 with a center of mass energy of $\sqrt{s} = 7\,\text{TeV}$. Events are selected to exploit the semileptonic decay channel of top quarks, in which the top decays into a W boson and a b quark, and the W then further decays into a muon or electron and its affiliated neutrino. The final state event signature is hence described by a high energetic lepton, a b jet and a large amount of missing transverse energy due to the undetected escaping neutrino.

In each lepton channel a neural network is trained to discriminate between the single top t-channel events and the dominating background, as there are $t\bar{t}$, W+jets, Z+jets, diboson, QCD and single top events from s-channel and associated tW production. The networks are then used to derive discriminator shape templates for each single process and systematic uncertainty.

Extensive studies were made to test the impact of the systematic uncertainties on β_s and to gain confidence that no systematic uncertainty is over- or underestimated (see

sections 6.4, 6.5.1 and 6.5.2) and it turned out, that most of the systematic uncertainties are well modelled.

A full Bayesian marginalisation is done to integrate out the systematic uncertainties that are well modelled. The Bayesian posterior for the signal strength parameter β_s is derived using Marcov Chain Monte Carlo techniques. Analogous to the 1σ interval of Gaussian errors, the upper and lower boundaries for the 1σ credible interval are taken to be the 84 % and the 16 % quantile, which covers the central 68 % for the posterior. The central value for the cross section measurement is taken as the 50 % quantile of the posterior. The uncertainties that did not qualify to be marginalised are studied individually using maximum likelihood fits and their conservatively estimated impact on β_s is added quadratically to the observed credible interval boundaries from the marginalisation.

This inference is performed three times, once in each lepton channel, and once using them simultaneously in a combined measurement. The derived results are listed in the following table. The quoted errors include the statistical uncertainty, as well as all systematic uncertainties.

		total uncertainty	
channel	measured cross section [pb]	absolute [pb]	relative [%]
muon	70.09	$^{+10.29}_{-10.09}$	$^{+14.69}_{-14.39}$
electron	67.40	$^{+10.90}_{-10.68}$	$^{+16.17}_{-15.85}$
combined	68.95	$^{+9.69}_{-9.33}$	$^{+14.06}_{-13.53}$

The next-to-leading order cross section calculations for these process in LHC collisions at a center of mass energy of $\sqrt{s} = 7\,\text{TeV}$ is $64.57^{+2.09}_{-0.71}\ ^{+1.51}_{-1.74}\,\text{pb}$ as published in [74]. The observed difference between the central values and the predicted cross section for the single top t-channel is well within the uncertainties and therefore both values are in good agreement.

Another CMS measurement for this quantity was performed on data with an integrated luminosity of $36\,\text{pb}^{-1}$ and yields a cross section of $83.6 \pm 29.8\,\text{pb}$ [71]. The ATLAS collaboration has published an according measurement on $1.04\,\text{fb}^{-1}$ of data in [2] and measured a cross section of $83 \pm 4\ (\text{stat.})\ ^{+20}_{-19}\ (\text{syst})\,\text{pb}$. These measurements also indicate an upward shift of the single top t-channel cross section at the LHC, but also here none of them is significant.

The CKM-matrix element $|V_{tb}|$ was derived from the measured cross section as $1.033^{+0.052}_{-0.055}$ which is again in good agreement with current world average of $0.999133^{+0.000044}_{-0.000043}$ [69].

In the following possible improvements for future measurements of this quantity are discussed.

Outlook

The analysis presented in this thesis indicates a small upward shift for the single top t-channel cross section in LHC data compared to the theoretically predicted value. This deviation is not significant since the residual uncertainties are of the order of 13 %. Therefore future measurements are needed with lower residual uncertainties.

Most of the uncertainties considered in this analysis could be constrained in situ using the Bayesian marginalisation, but there are some uncertainties that are not well understood yet, such as Q^2 scale and matching threshold uncertainty. The current approaches to model them probably do not reflect the true behaviour of these uncertainties, as is explained in section 6.5.1. For these uncertainties it is not yet known on how they can be constrained in data and large prior uncertainties need to be assumed. Hence for future measurements it is advisable to find a way to constrain them in situ, or in a dedicated measurement.

It is also worth investigating the relatively large impact of about 8.5 % that was observed in the comparison of Powheg and CompHEP as Monte Carlo generator for signal events. It might well be that with further understanding of LHC processes and more precise theory calculations, future releases of both programs will yield comparable results, if they model the single top t-channel process with more accuracy.

In 2012 the LHC will run with a center of mass energy of $\sqrt{s} = 8\,\text{TeV}$. This also means an increased cross section, which will need to be measured for the single top production processes. It will be interesting to see if the spread between predicted and observed cross section becomes more significant.

A. NeuroBayes input variables distributions

In this chapter, the NeuroBayes input variables are listed in the same order as in section 6.3.2. In each figure the distributions are given for one variable at a time in all six jet/tag categories, seperated by the muon and the electron channel.

In each figure the following jet/tag categorie pattern is used to align the plots.

2j1t	3j1t	4j1t
2j2t	3j2t	4j2t

On the upper half of the page the distributions for the muon channel are shown, while there are the distributions for the electron channel on the lower half. All distributions follow the same color code for the different processes. The legened is omitted in some plots, because it would have interfered with the histograms. Empty plots occur in the two jet regions, if information about a third jet are needed to fill the distribution. They are drawn to keep the pattern, but can be ignored. The MTW/MET distributions are drawn without the QCD suppresion cuts for the muon/electron channel.

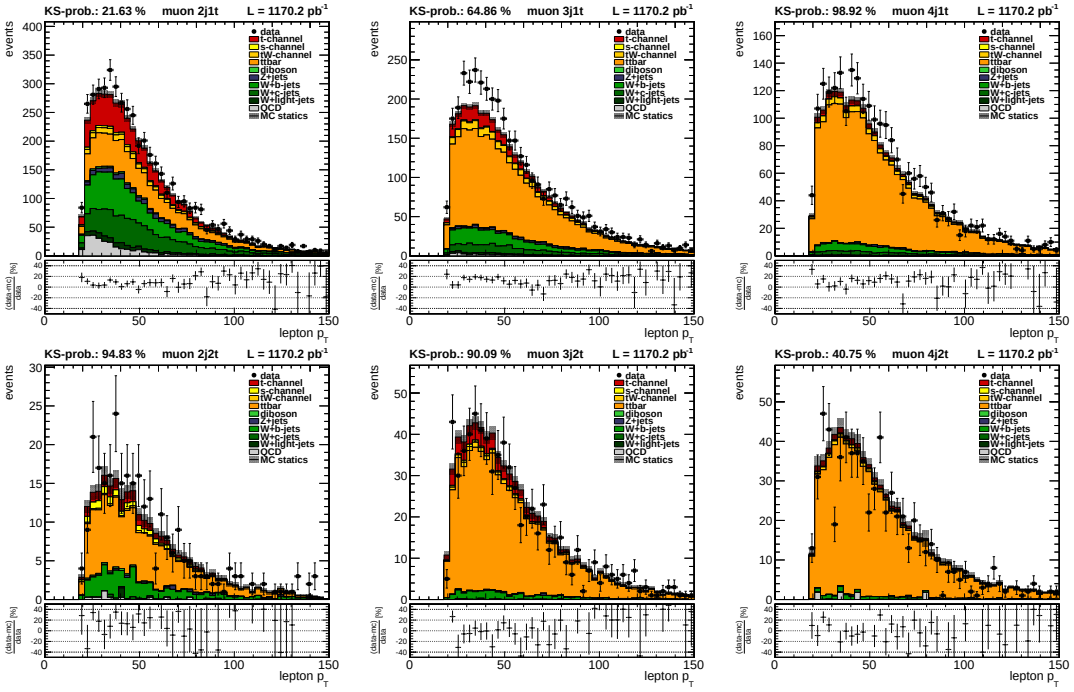


Figure A.1.: Muon transverse momentum p_T.

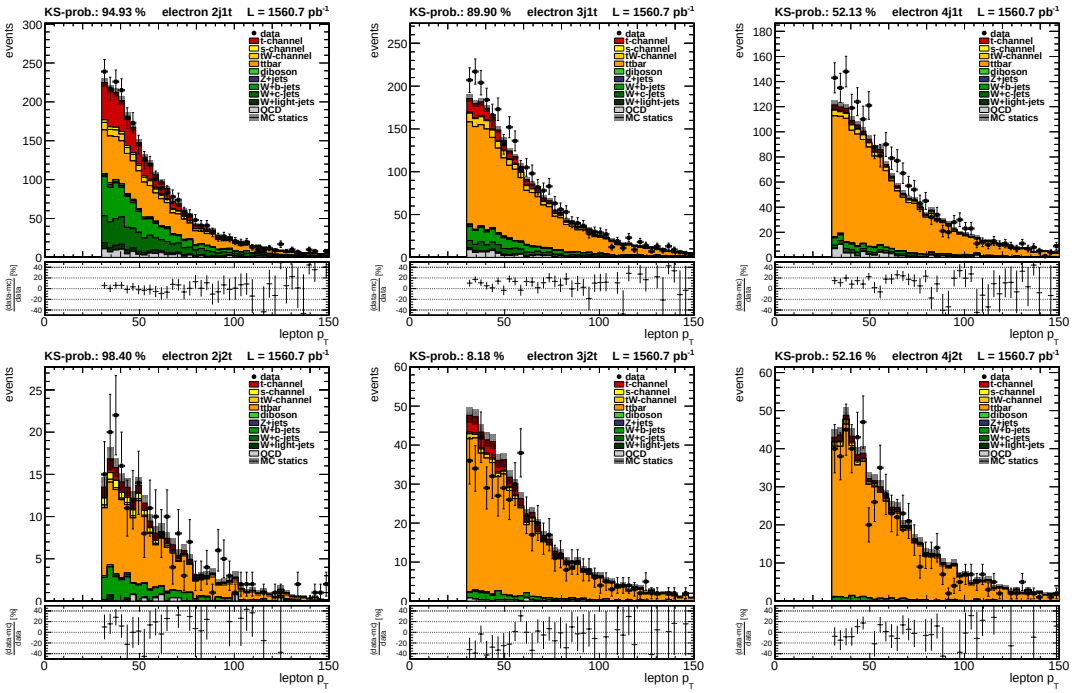


Figure A.2.: Electron transverse momentum p_T.

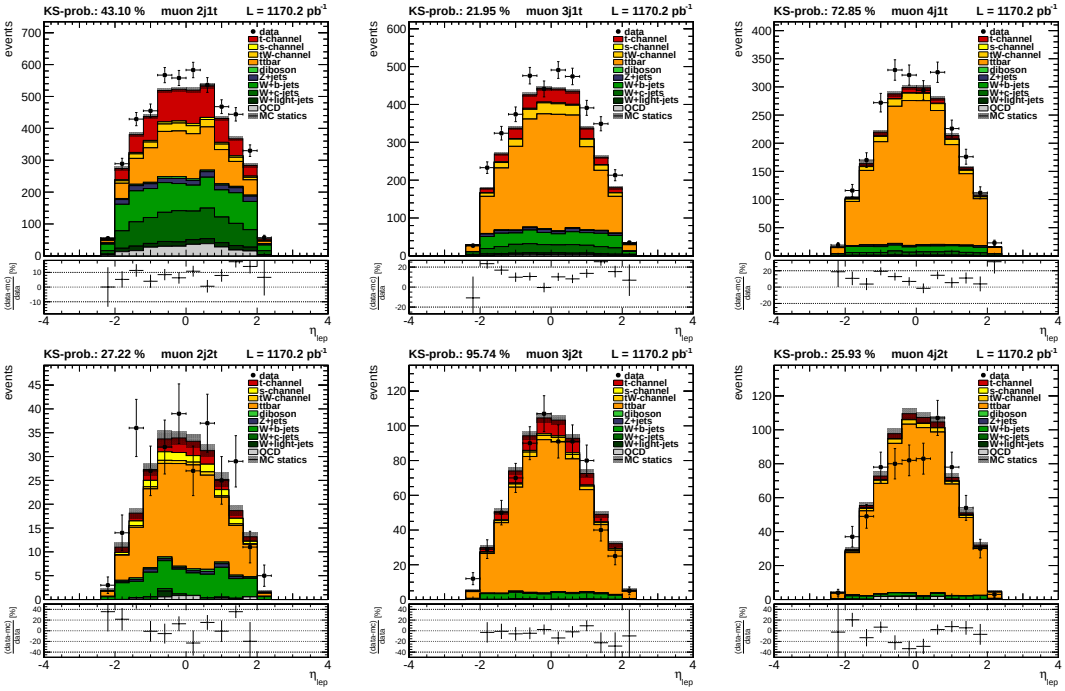


Figure A.3.: Muon pseudo rapidity η.

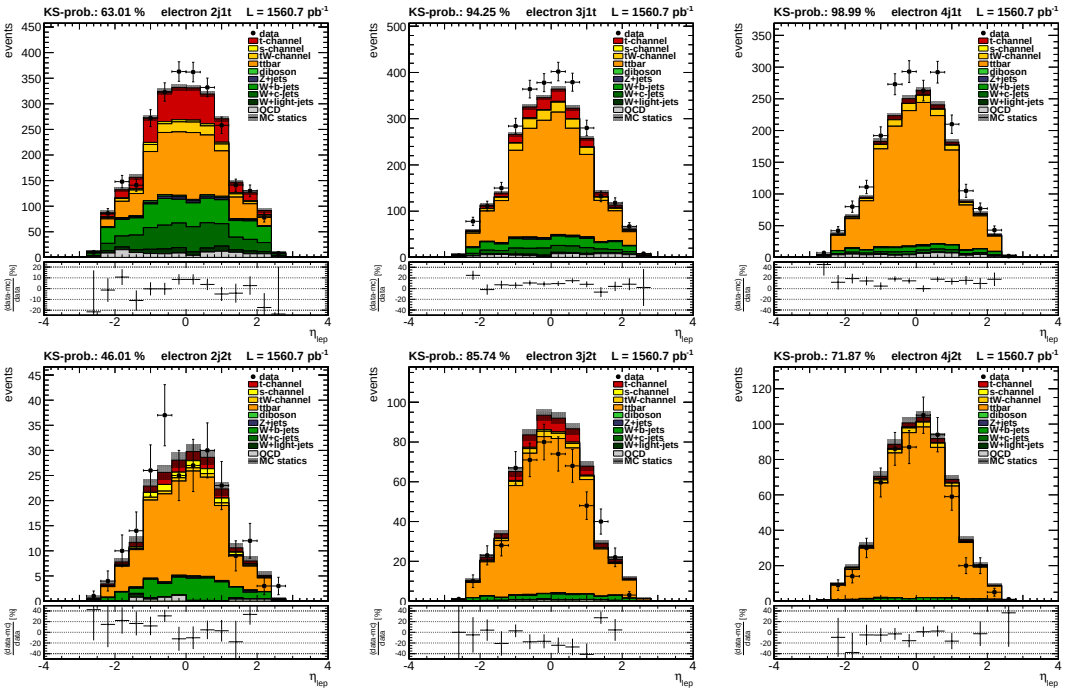


Figure A.4.: Electron pseudo rapidity η.

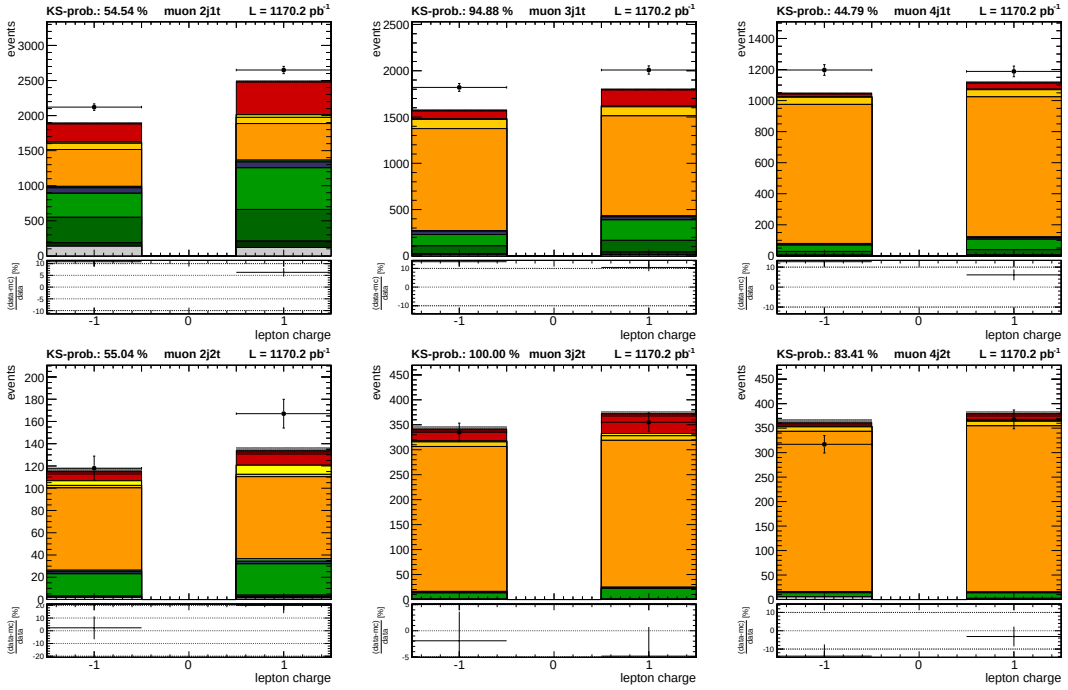


Figure A.5.: Muon charge

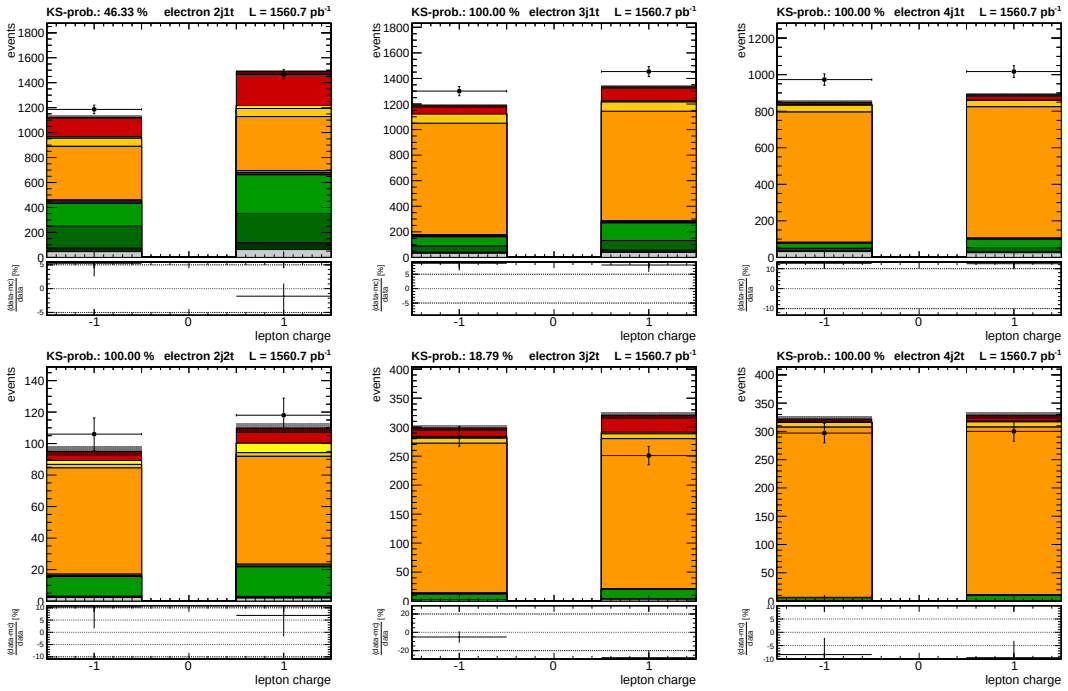


Figure A.6.: Electron charge.

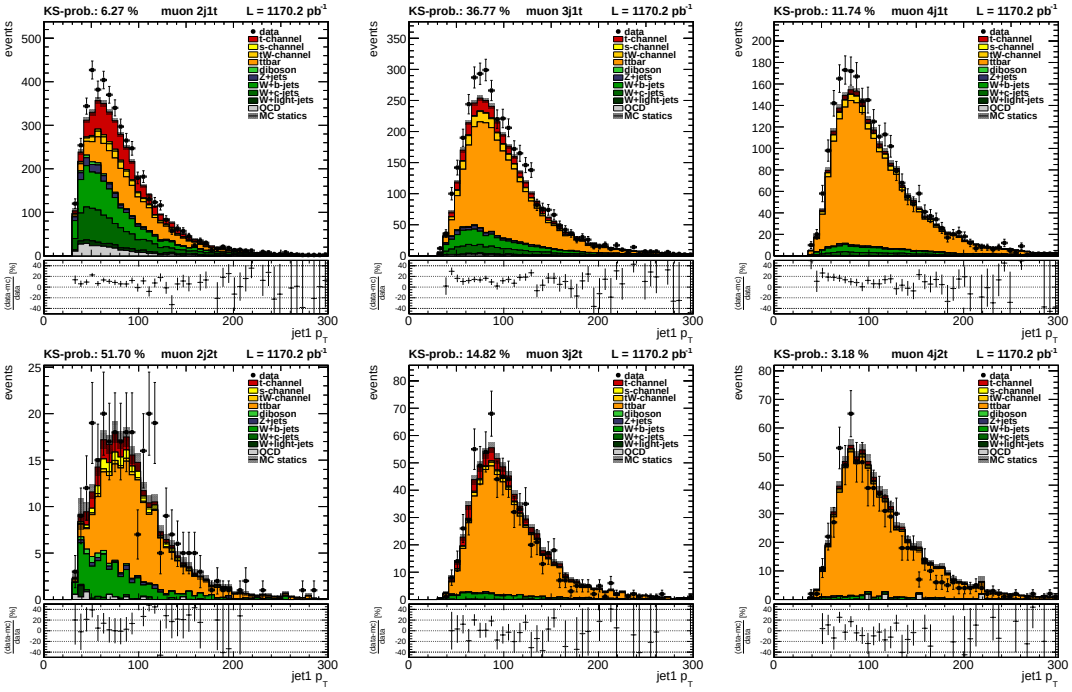


Figure A.7.: Transverse momentum p_T of first jet in muon channel

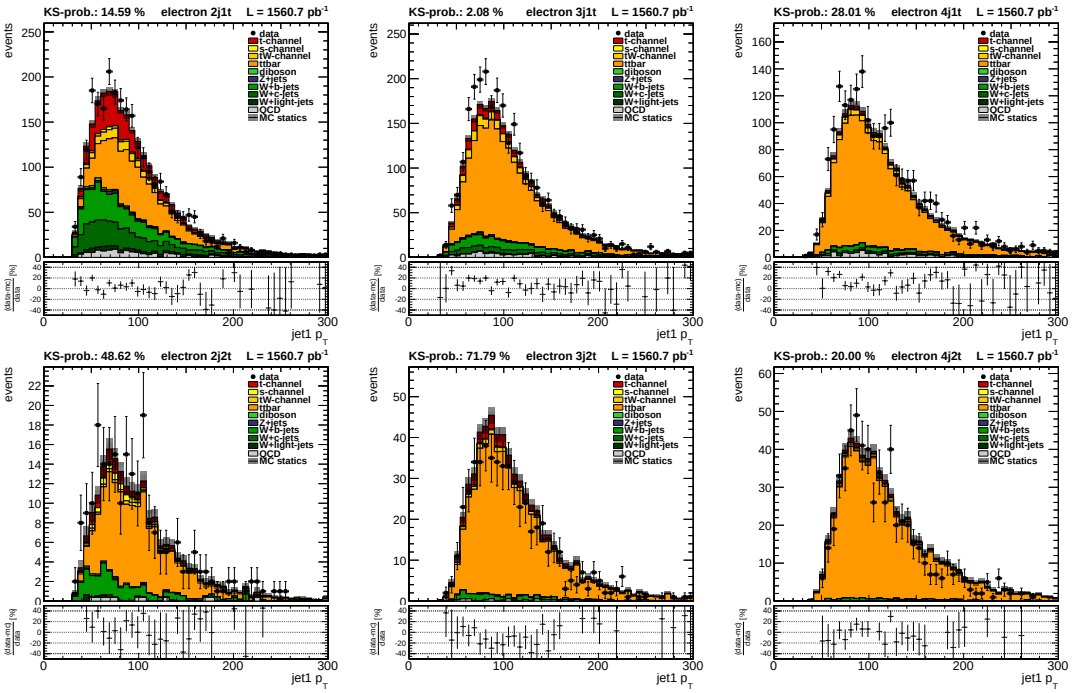


Figure A.8.: Transverse momentum p_T of first jet in electron channel

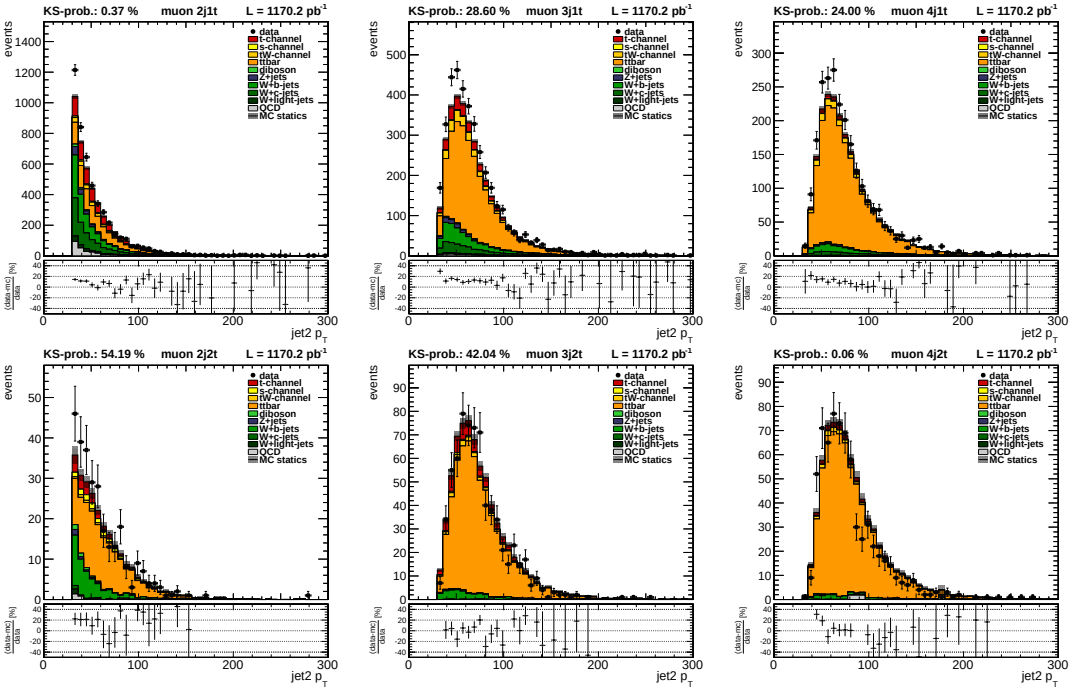


Figure A.9.: Transverse momentum p_T of second jet in muon channel

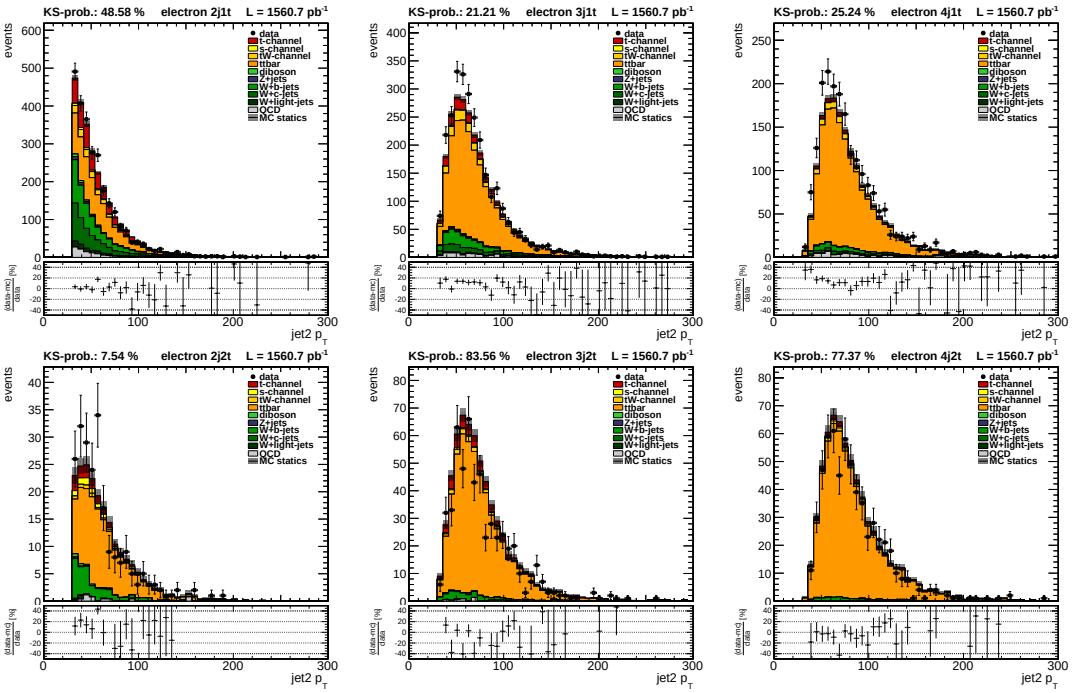


Figure A.10.: Transverse momentum p_T of second jet in electron channel

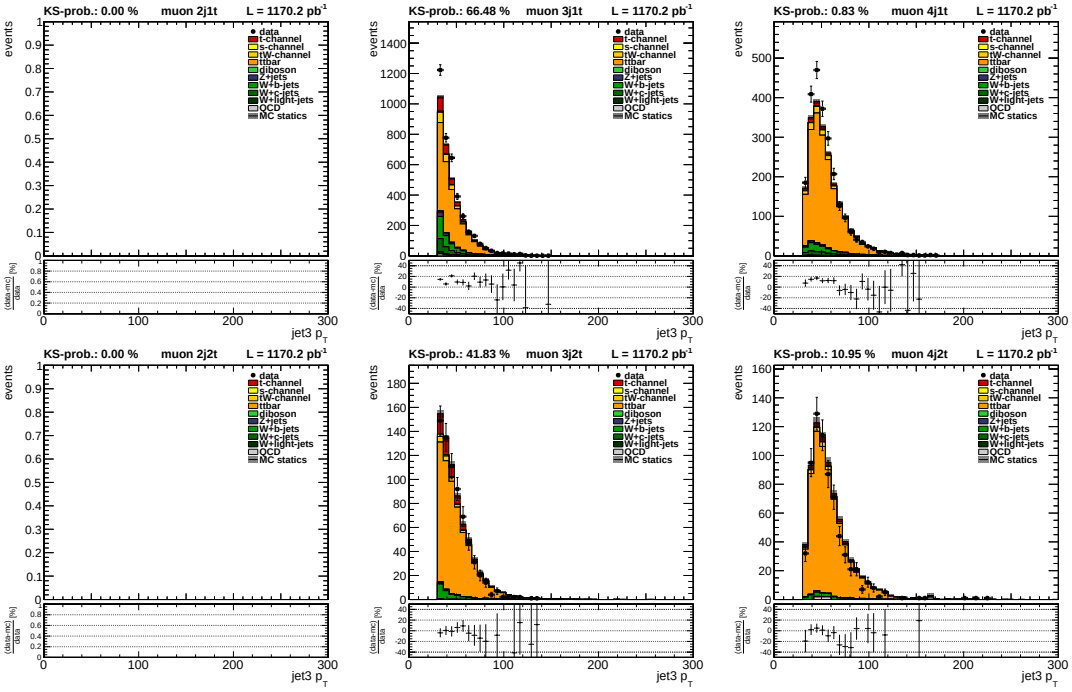


Figure A.11.: Transverse momentum p_T of third jet in muon channel

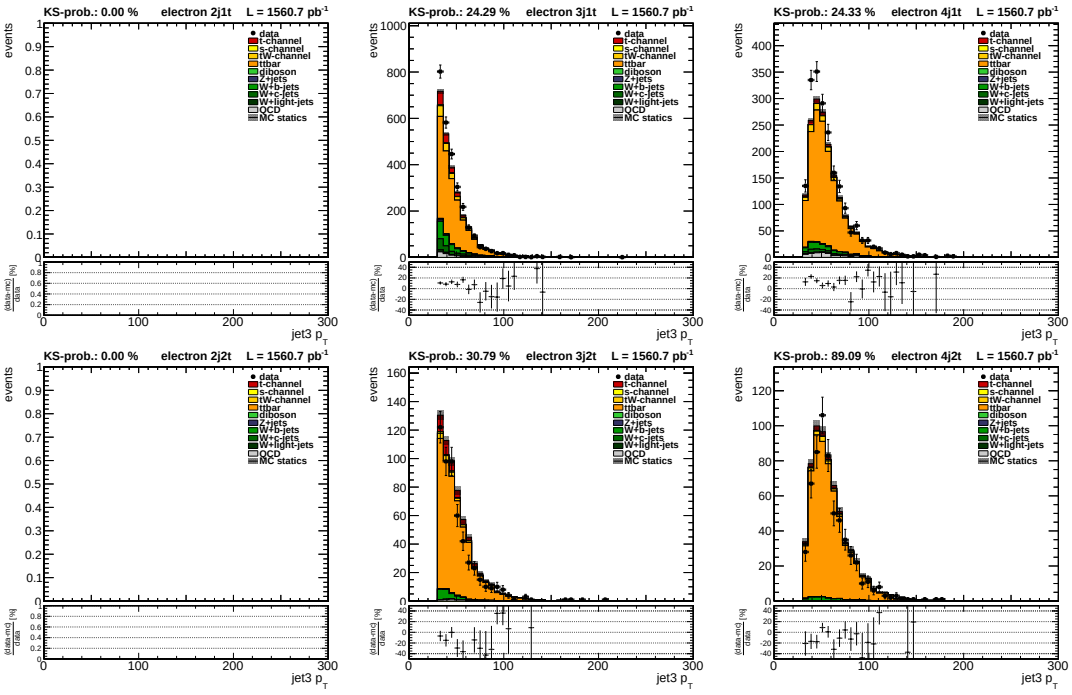


Figure A.12.: Transverse momentum p_T of third jet in electron channel

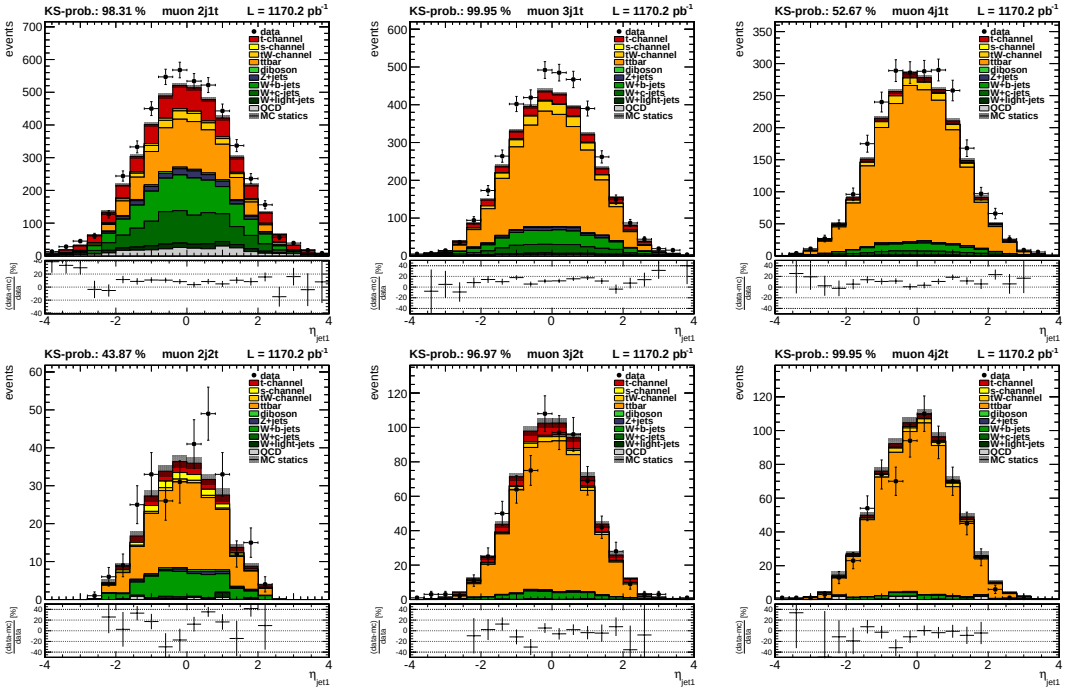


Figure A.13.: Pseudo rapidity η of first jet in muon channel

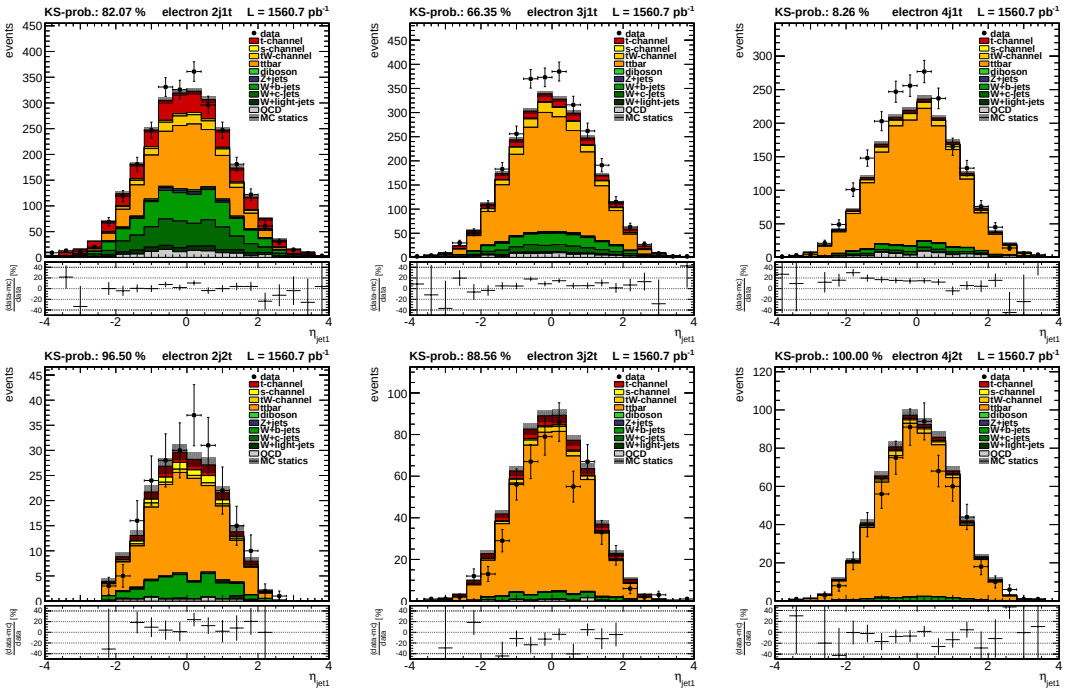


Figure A.14.: Pseudo rapidity η of first jet in electron channel

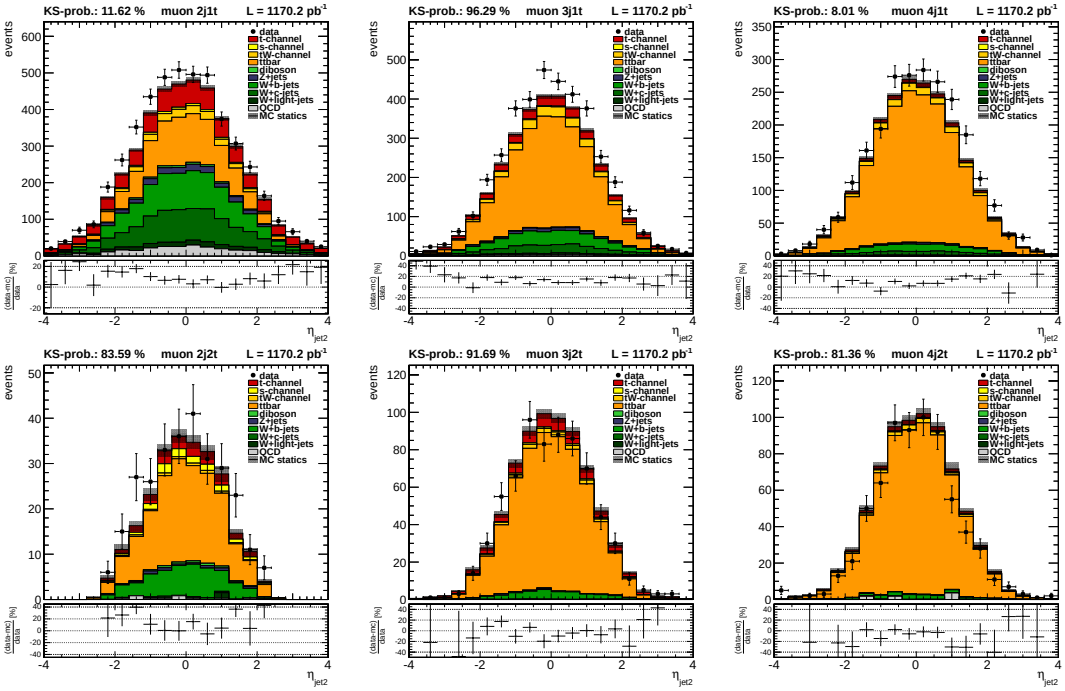


Figure A.15.: Pseudo rapidity η of second jet in muon channel

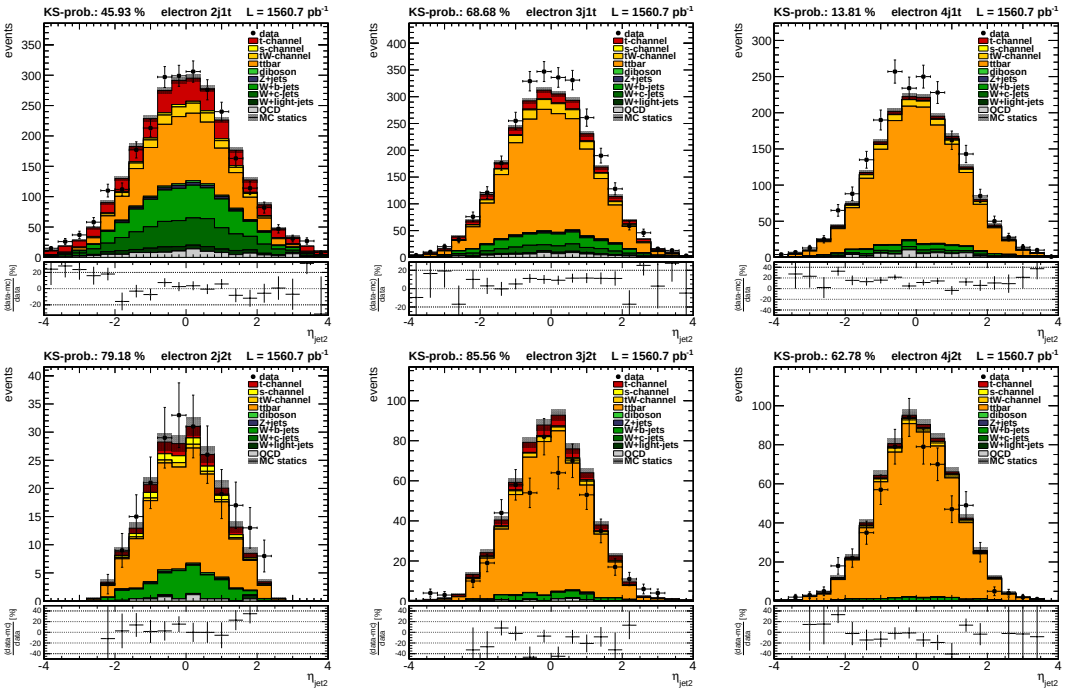


Figure A.16.: Pseudo rapidity η of second jet in electron channel

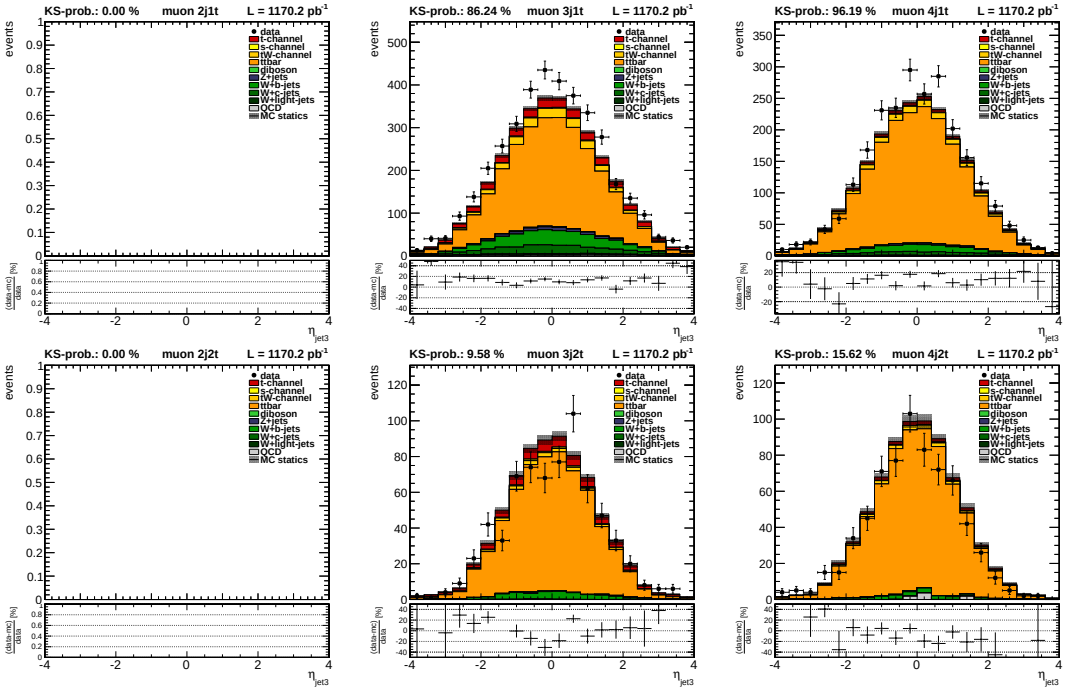


Figure A.17.: Pseudo rapidity η of third jet in muon channel

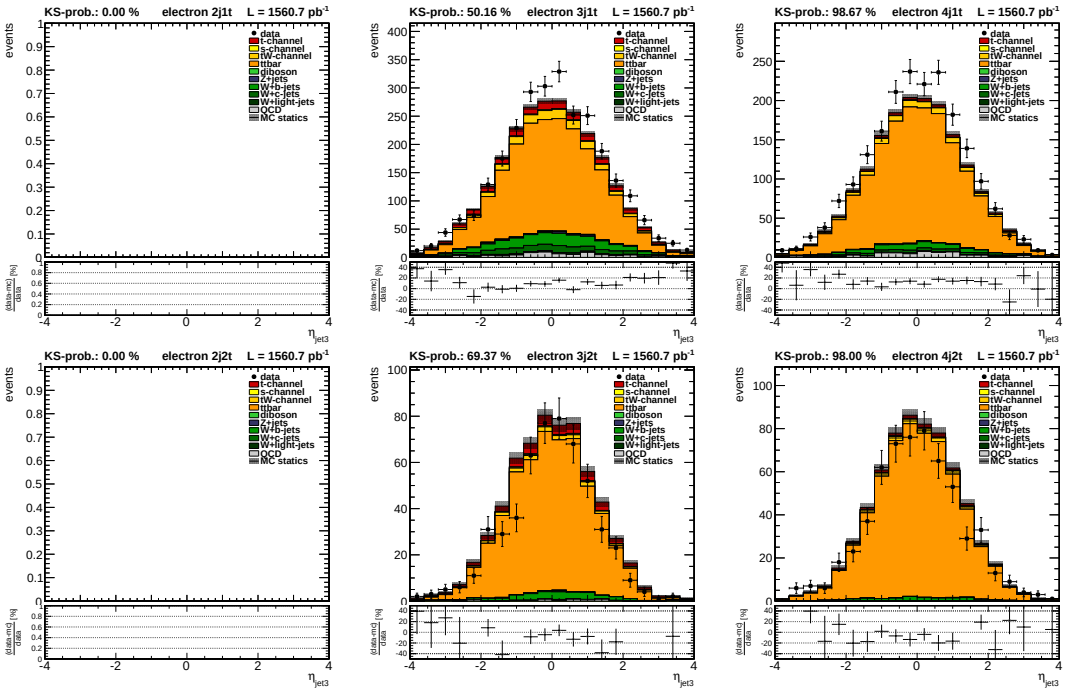


Figure A.18.: Pseudo rapidity η of third jet in electron channel

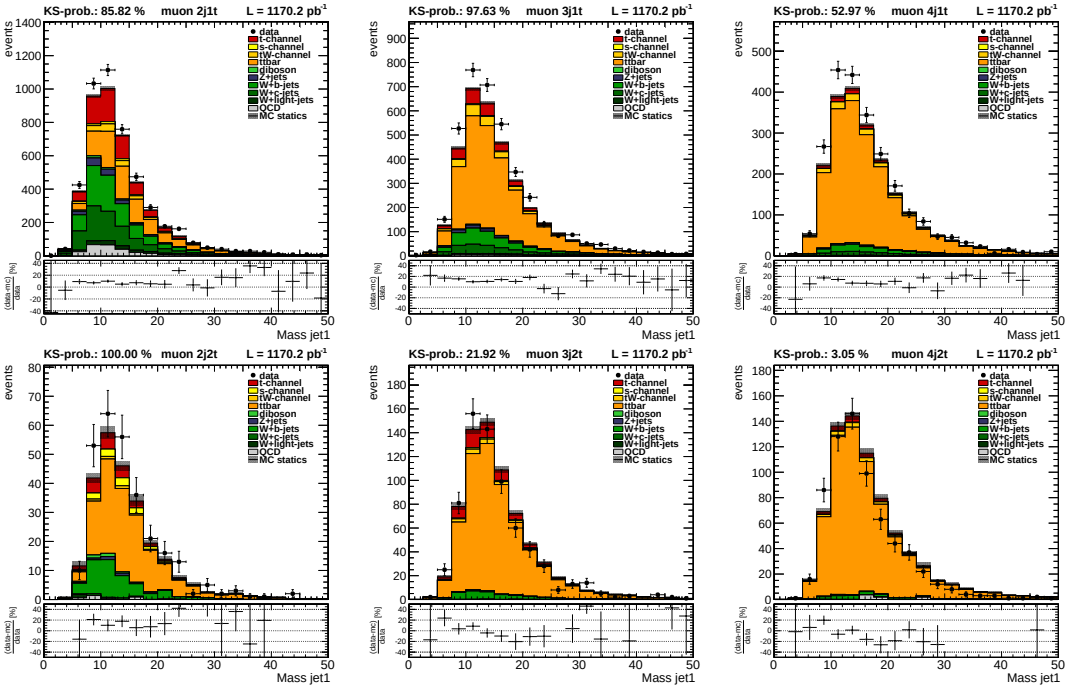


Figure A.19.: Invariant mass of first jet in muon channel

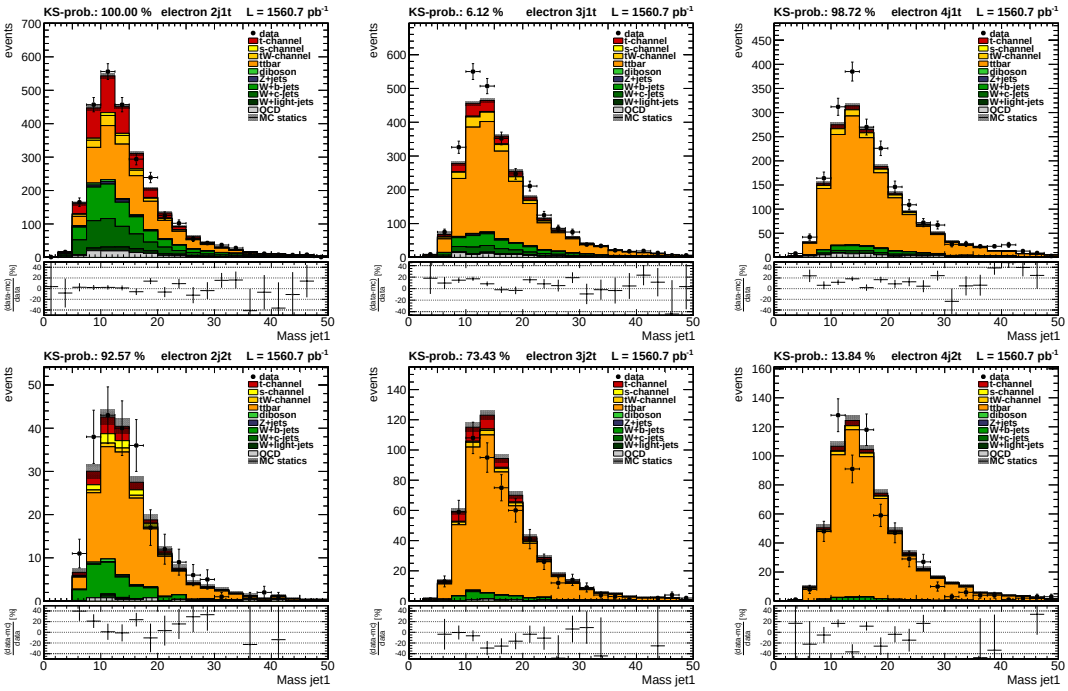


Figure A.20.: Invariant mass of first jet in electron channel

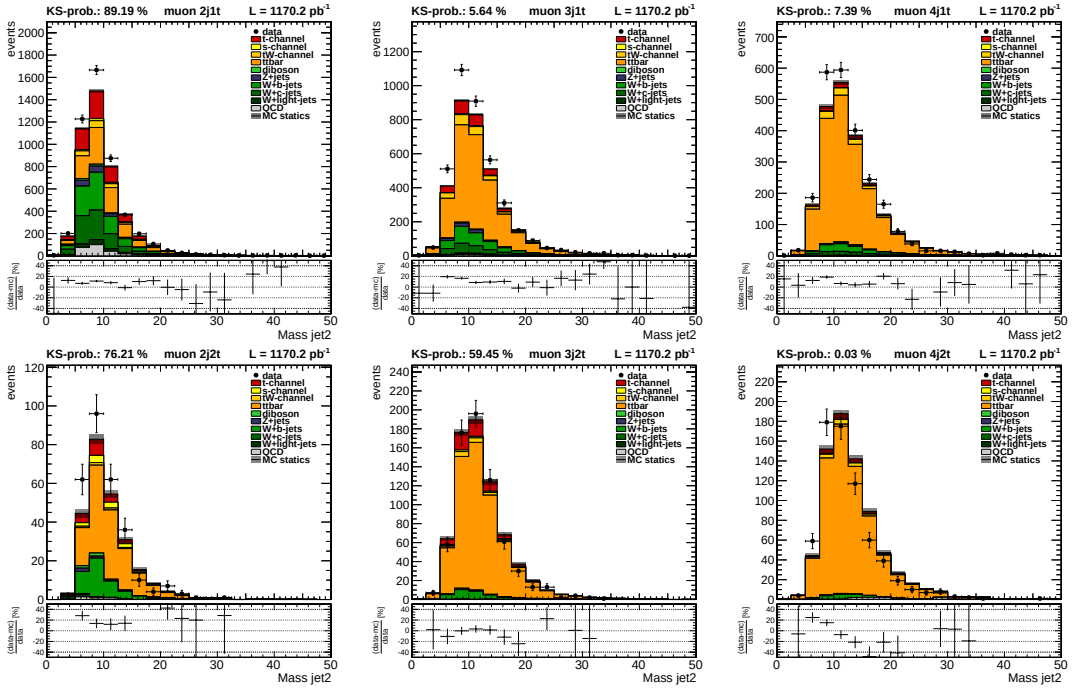


Figure A.21.: Invariant mass of second jet in muon channel

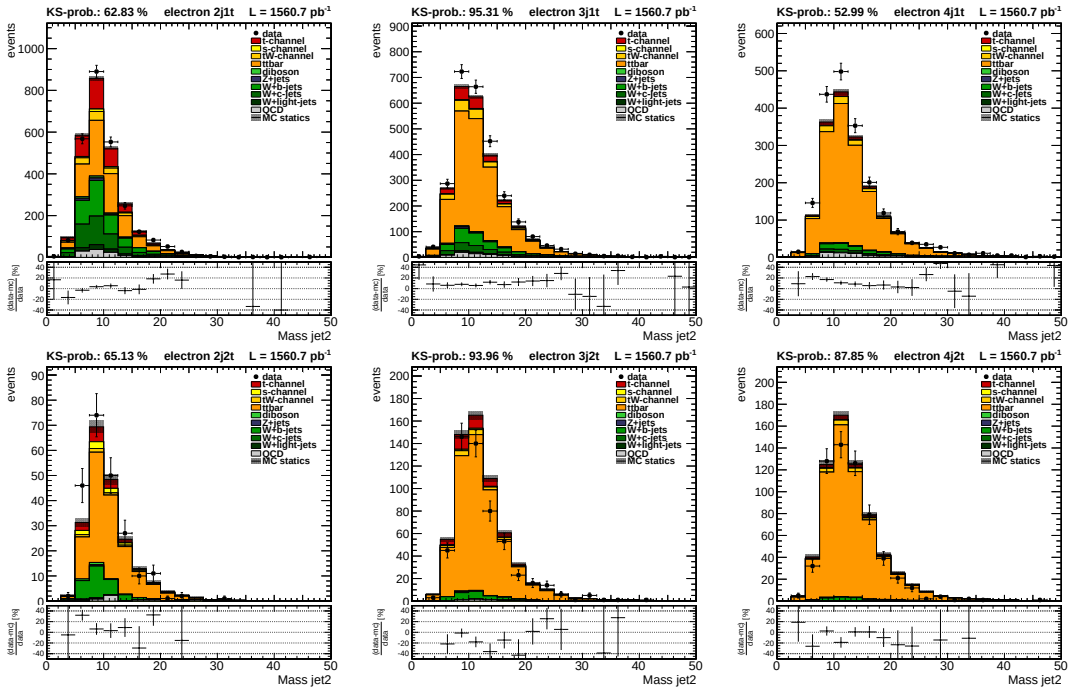


Figure A.22.: Invariant mass of second jet in electron channel

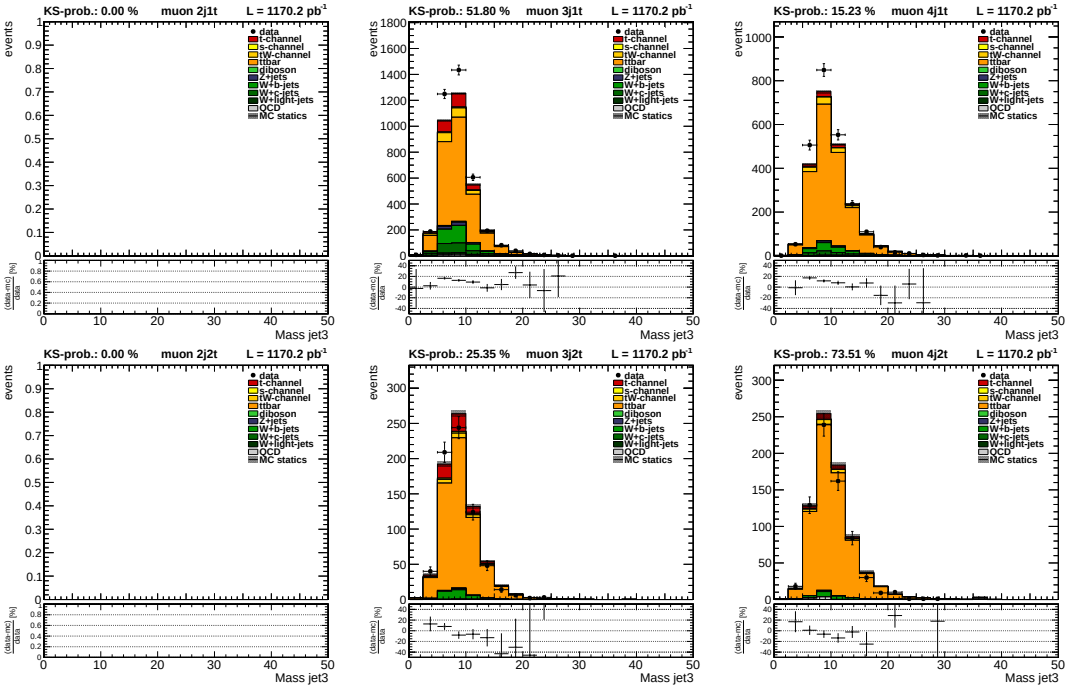


Figure A.23.: Invariant mass of third jet in muon channel

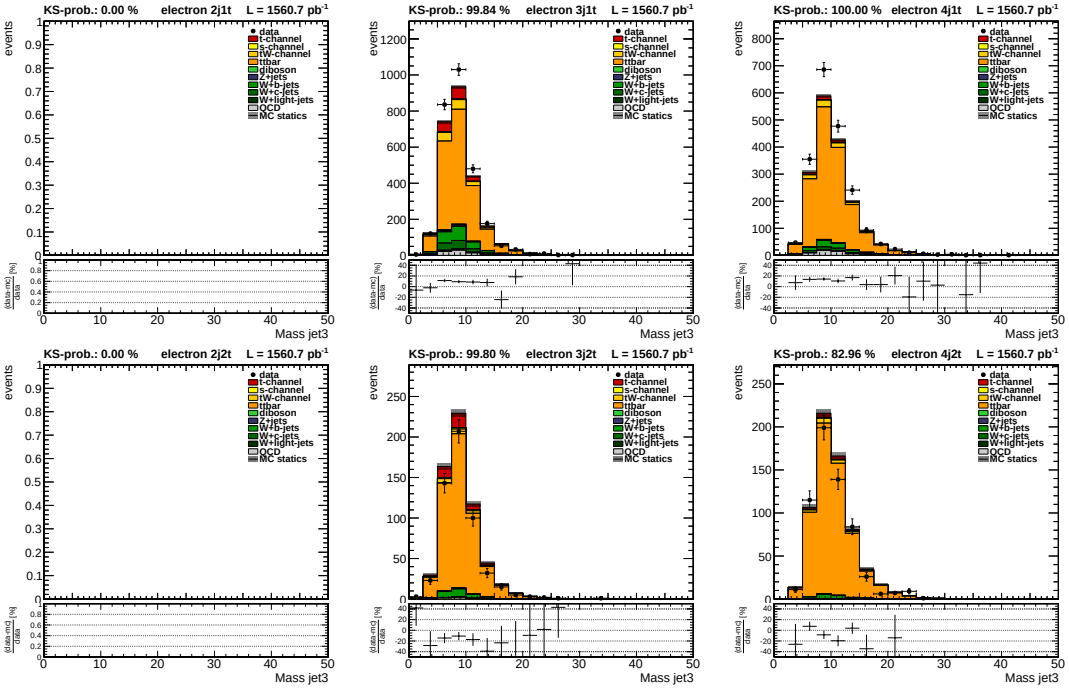


Figure A.24.: Invariant mass of third jet in electron channel

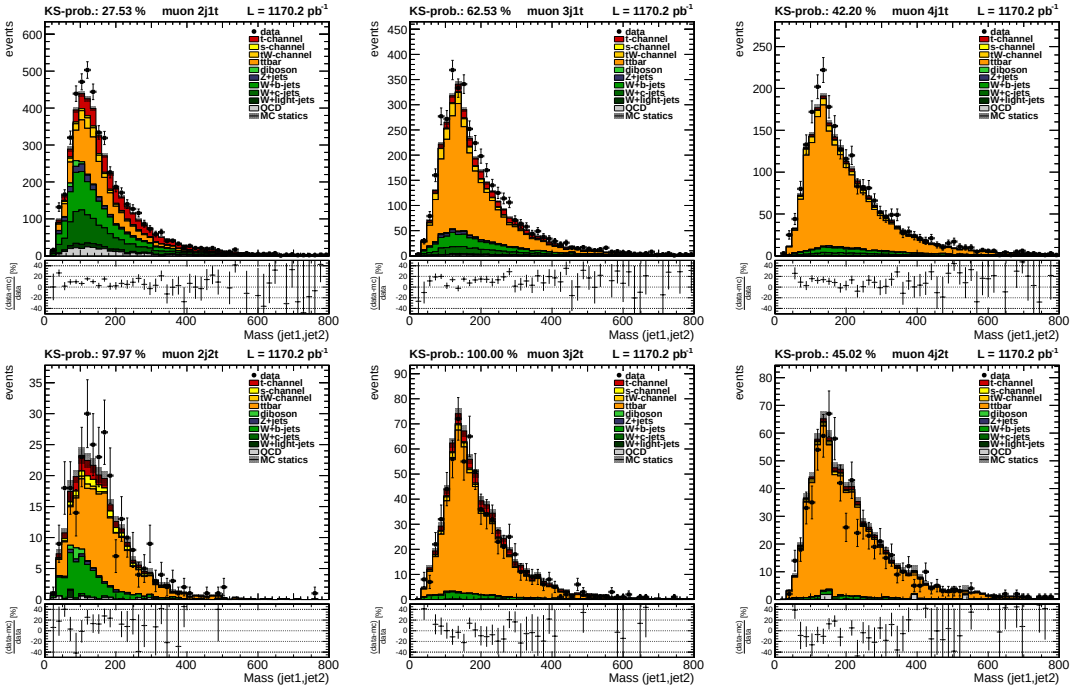


Figure A.25.: Invariant mass of first and second jet in muon channel

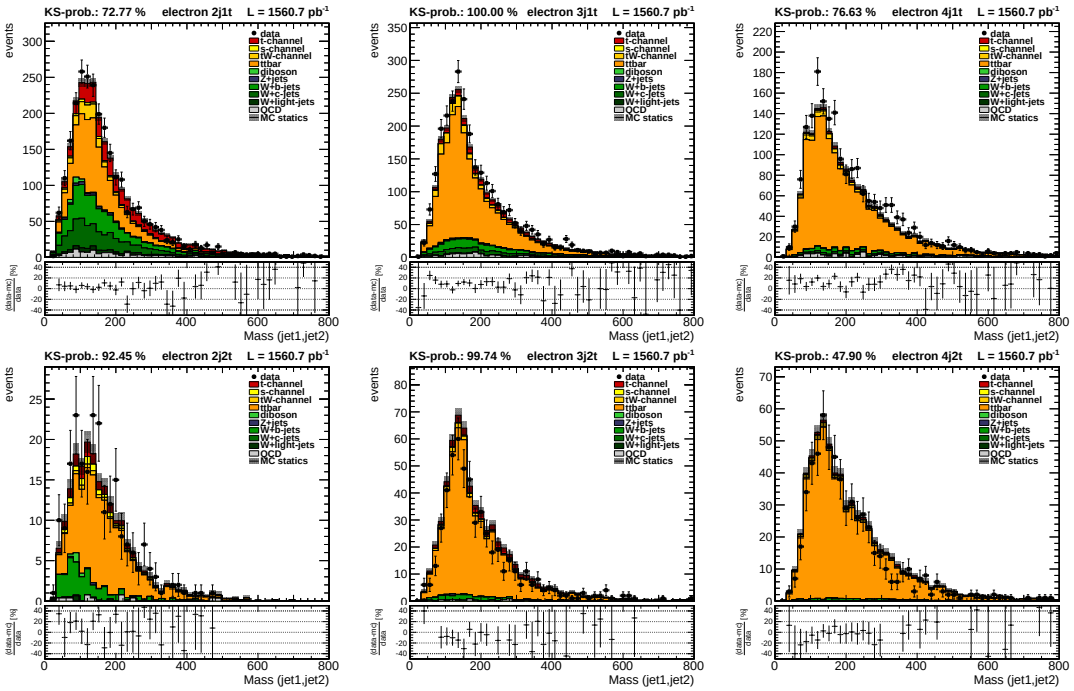


Figure A.26.: Invariant mass of first and second jet in electron channel

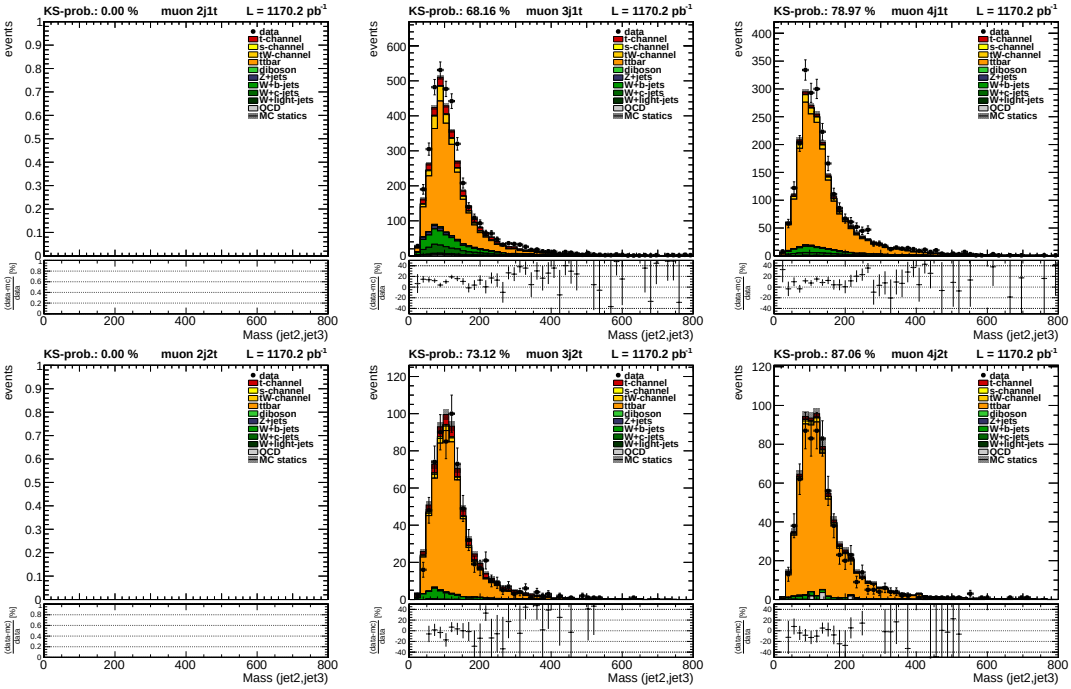


Figure A.27.: Invariant mass of second and third jet in muon channel

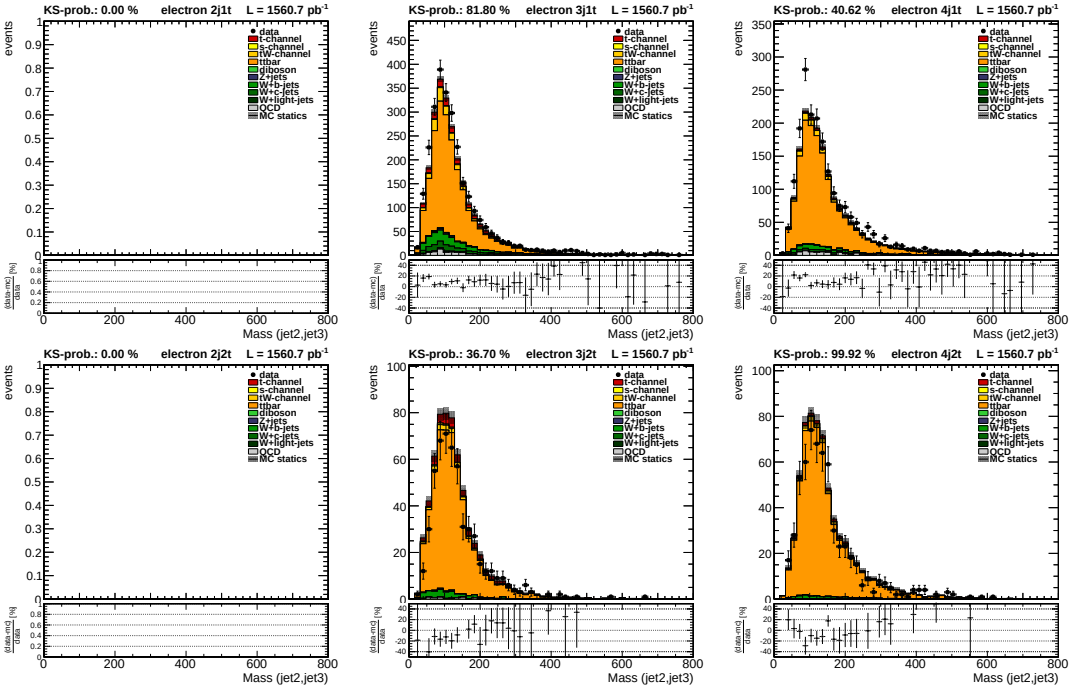


Figure A.28.: Invariant mass of second and third jet in electron channel

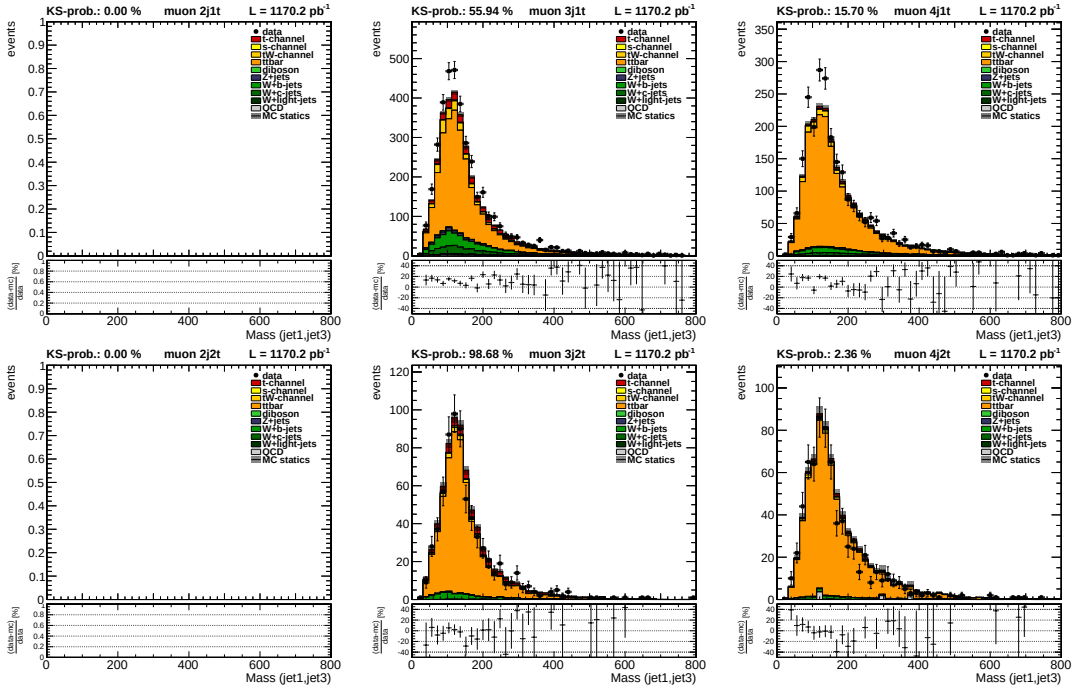


Figure A.29.: Invariant mass of first and third jet in muon channel

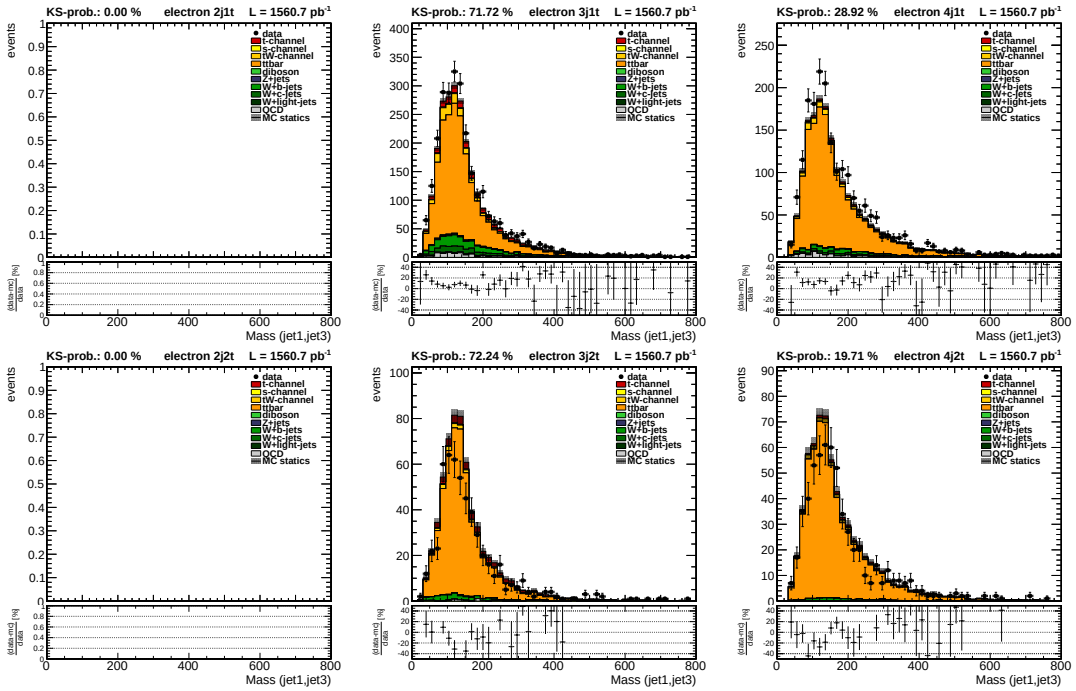


Figure A.30.: Invariant mass of first and third jet in electron channel

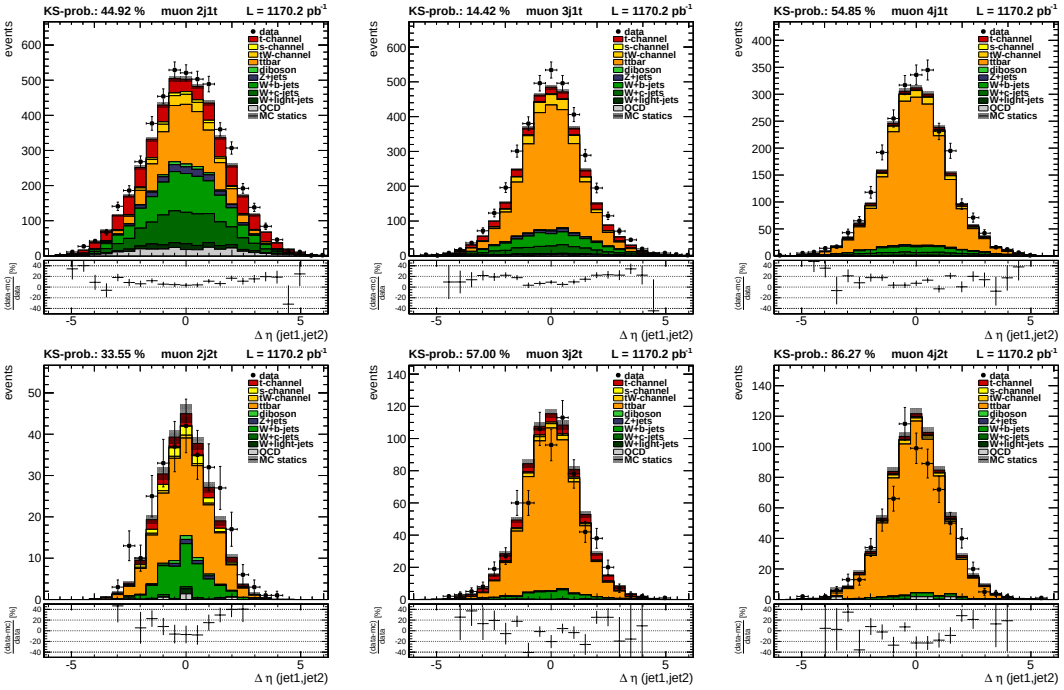


Figure A.31.: Difference in pseudo rapidity $\Delta\eta$ of first and second jet in muon channel

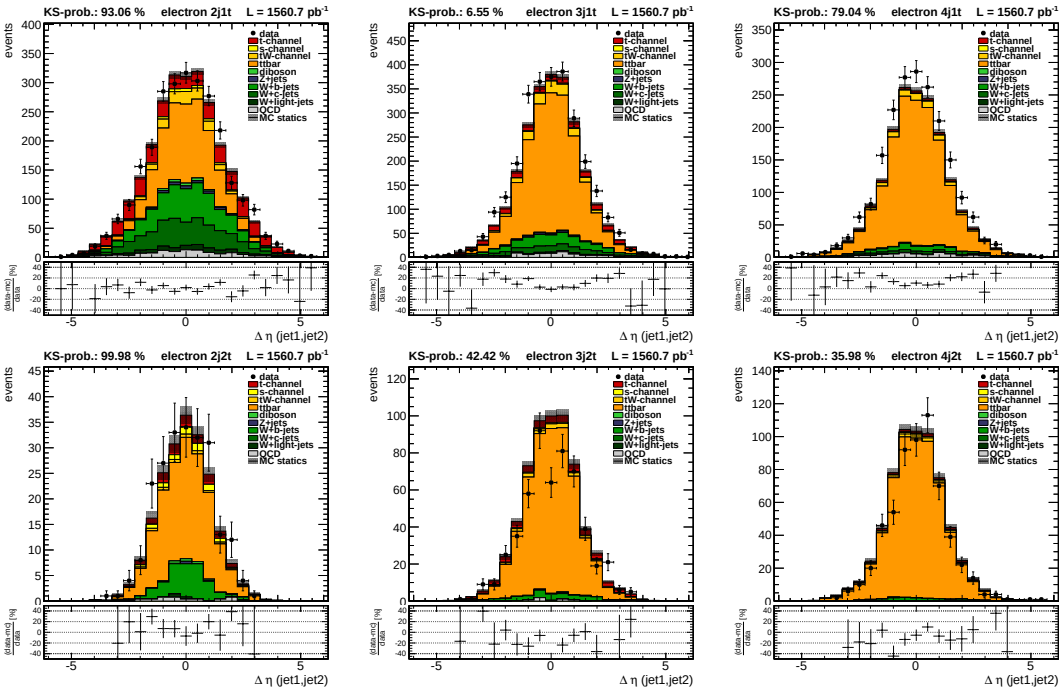


Figure A.32.: Difference in pseudo rapidity $\Delta\eta$ of first and second jet in electron channel

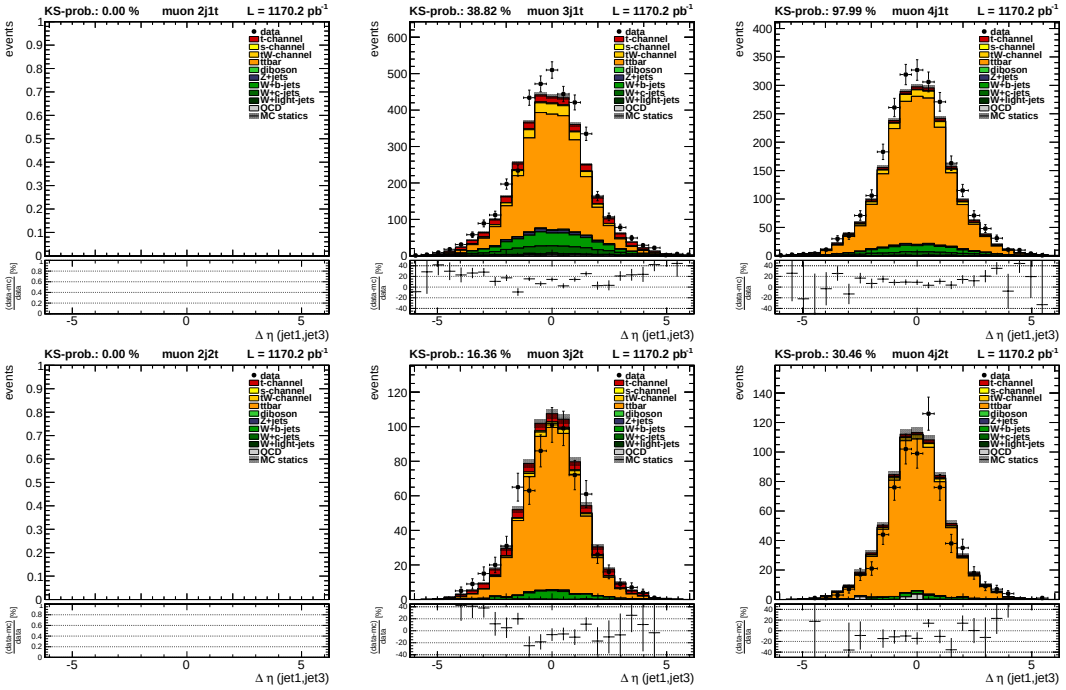


Figure A.33.: Difference in pseudo rapidity $\Delta\eta$ of first and third jet in muon channel

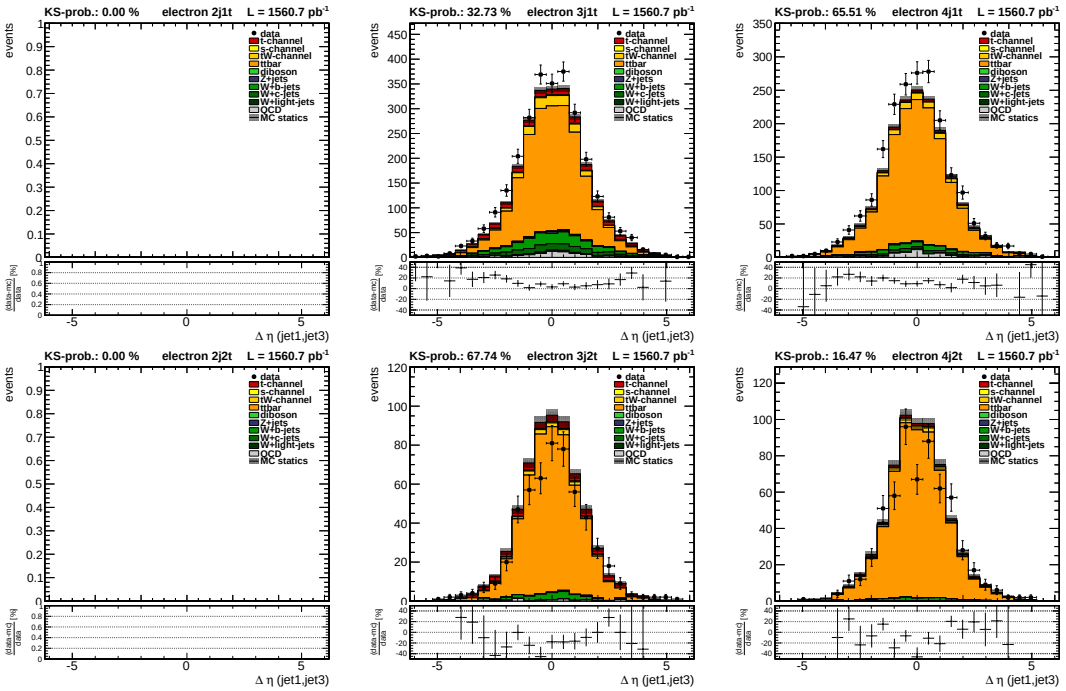


Figure A.34.: Difference in pseudo rapidity $\Delta\eta$ of first and third jet in electron channel

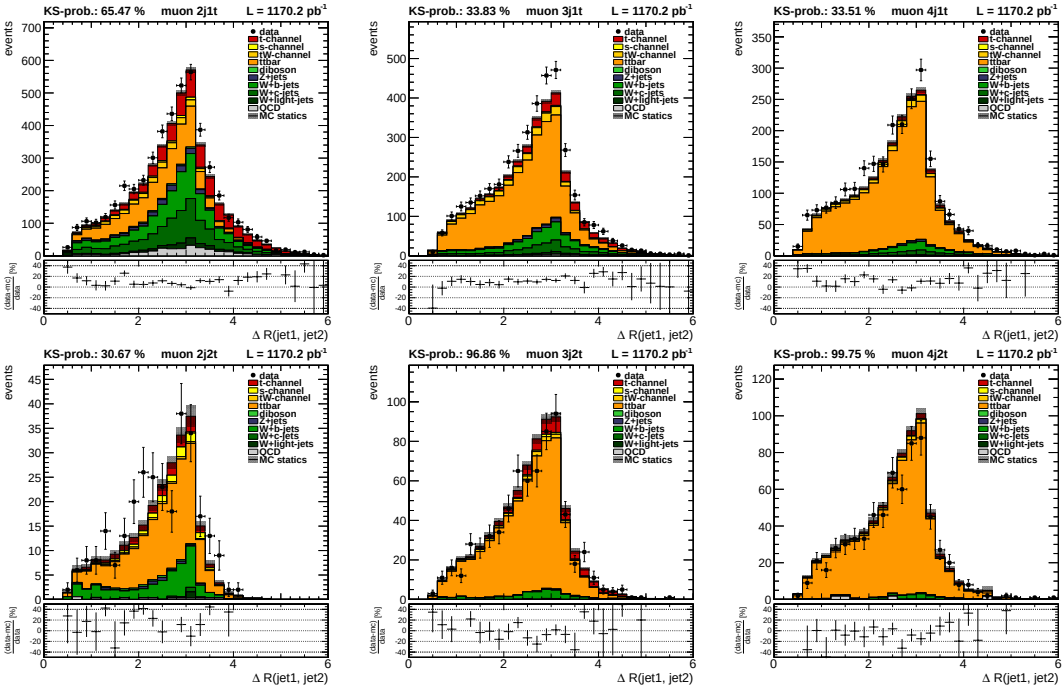


Figure A.35.: Distance in the $\eta-\phi$ plane ΔR between first and second jet in muon channel

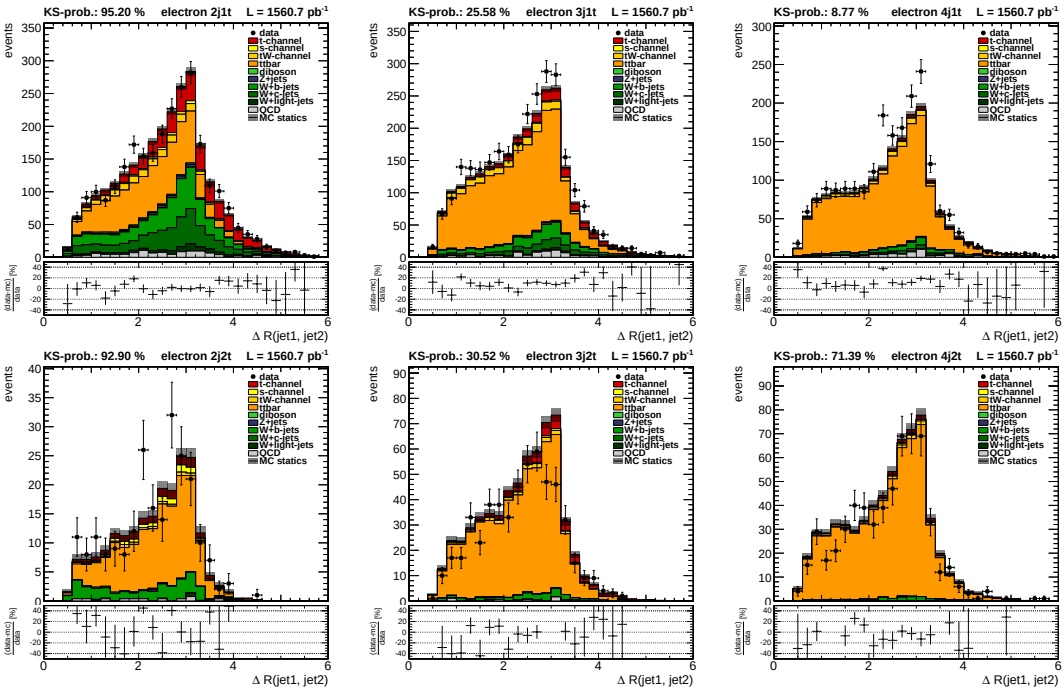


Figure A.36.: Distance in the $\eta - \phi$ plane ΔR between first and second jet in electron channel

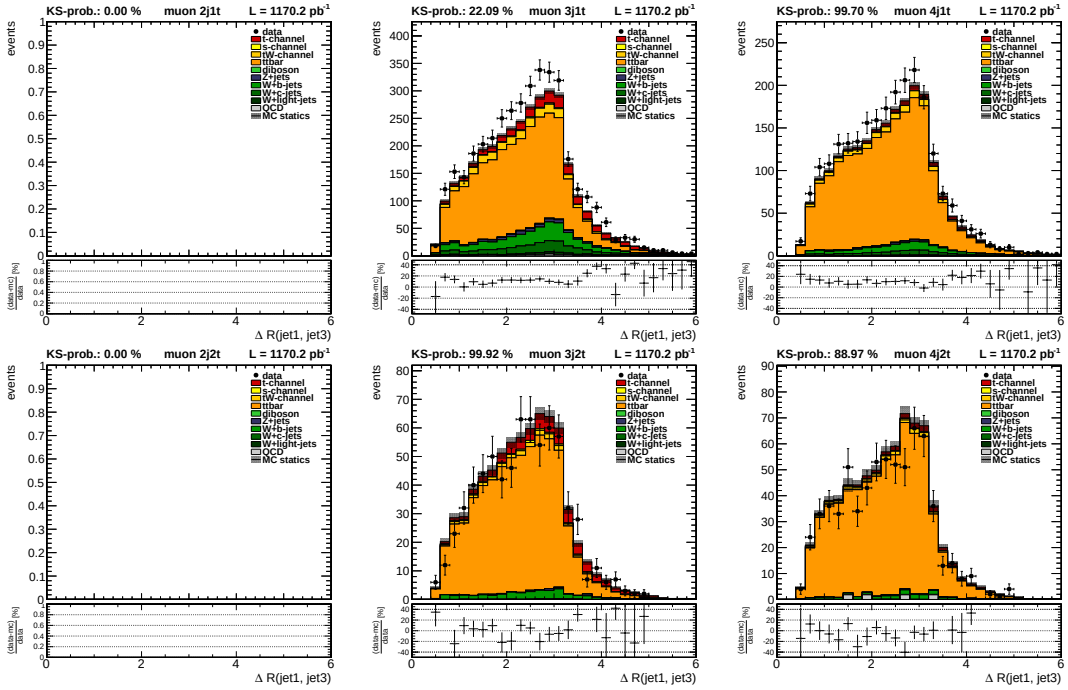


Figure A.37.: Distance in the $\eta - \phi$ plane ΔR between first and third jet in muon channel

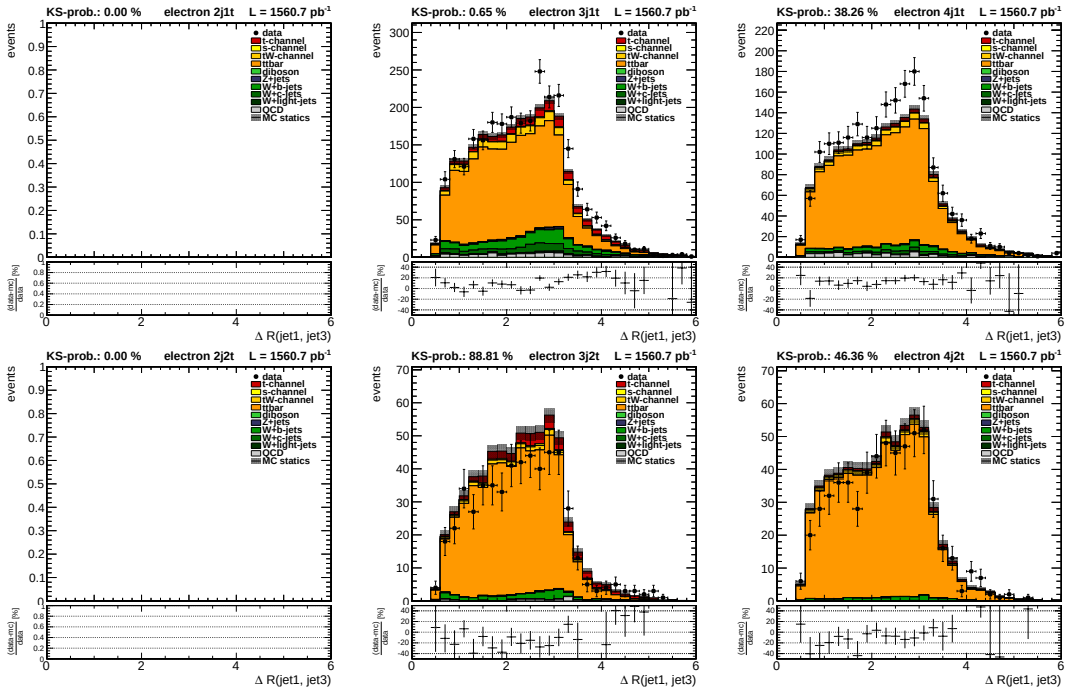


Figure A.38.: Distance in the $\eta - \phi$ plane ΔR between first and third jet in electron channel

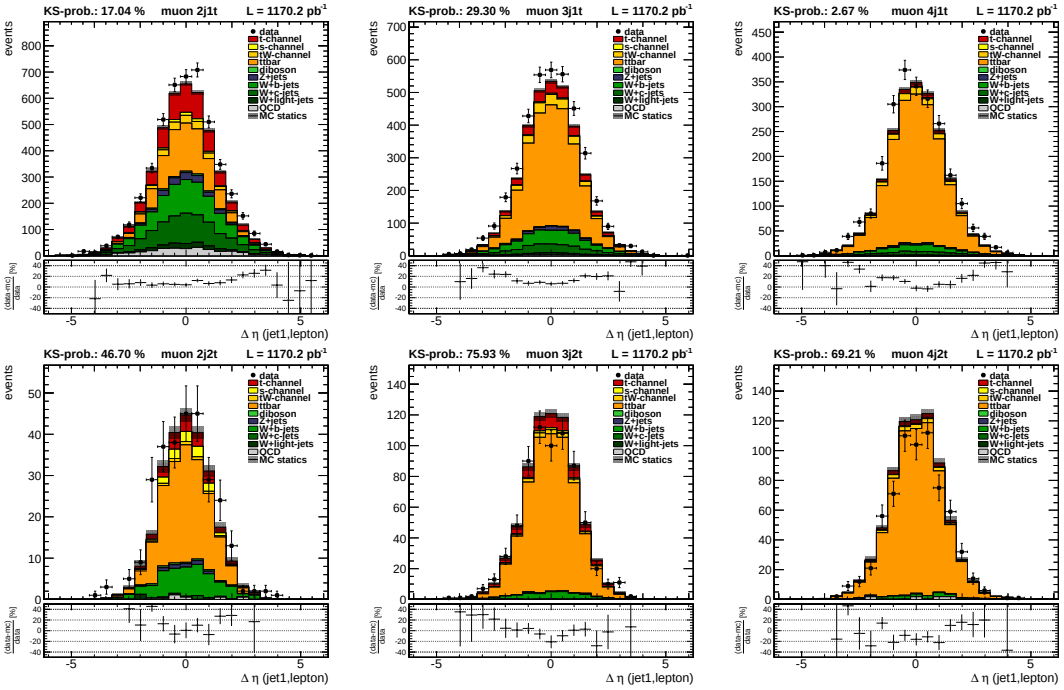


Figure A.39.: Difference in pseudo rapidity $\Delta\eta$ between first jet and lepton in muon channel

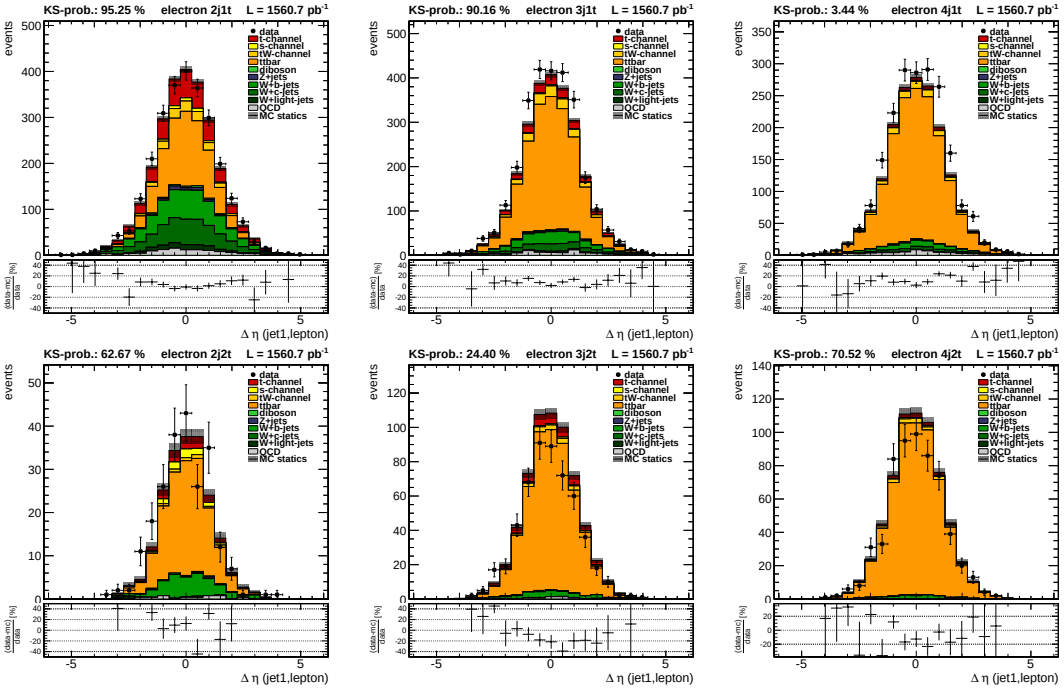


Figure A.40.: Difference in pseudo rapidity $\Delta\eta$ between first jet and lepton in electron channel

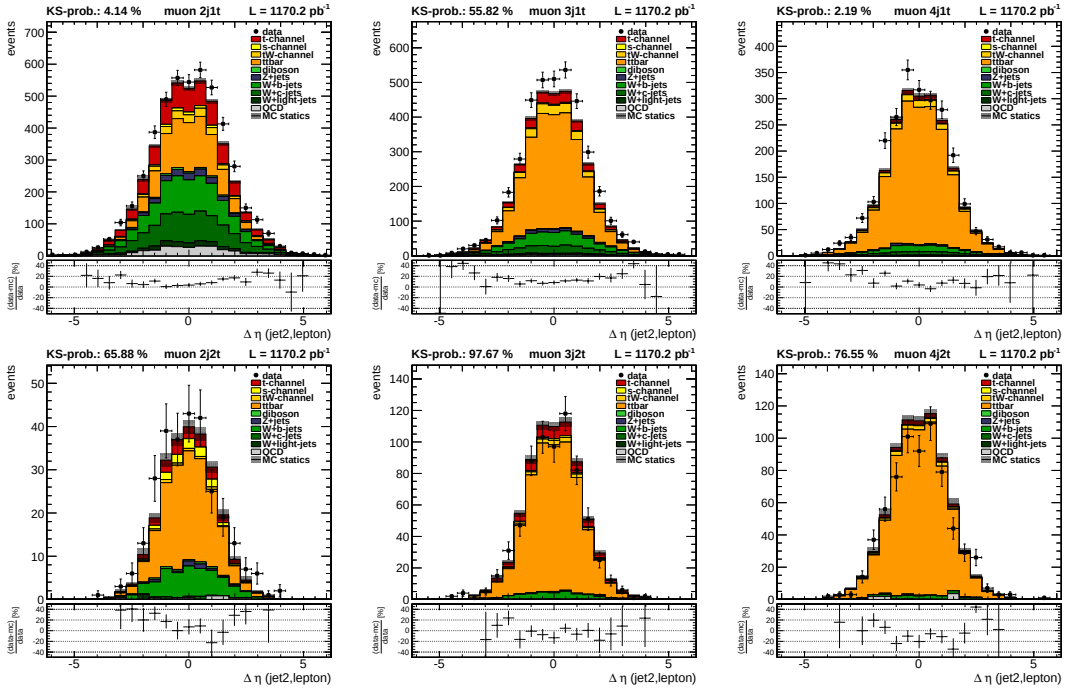


Figure A.41.: Difference in pseudo rapidity $\Delta\eta$ between second jet and lepton in muon channel

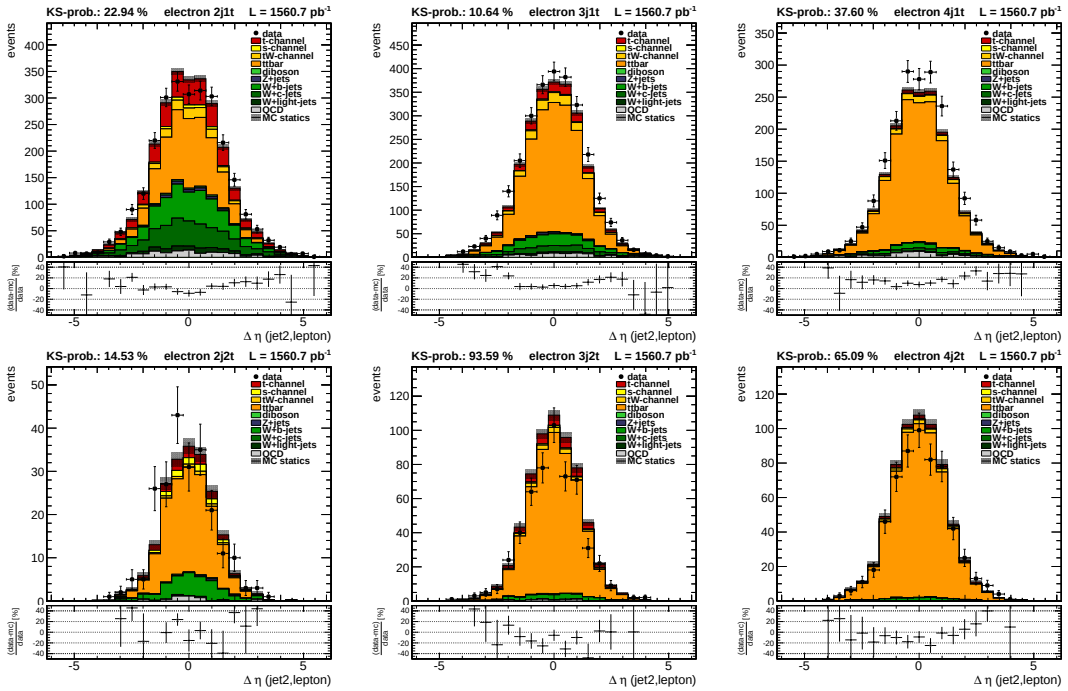


Figure A.42.: Difference in pseudo rapidity $\Delta\eta$ between second jet and lepton in electron channel

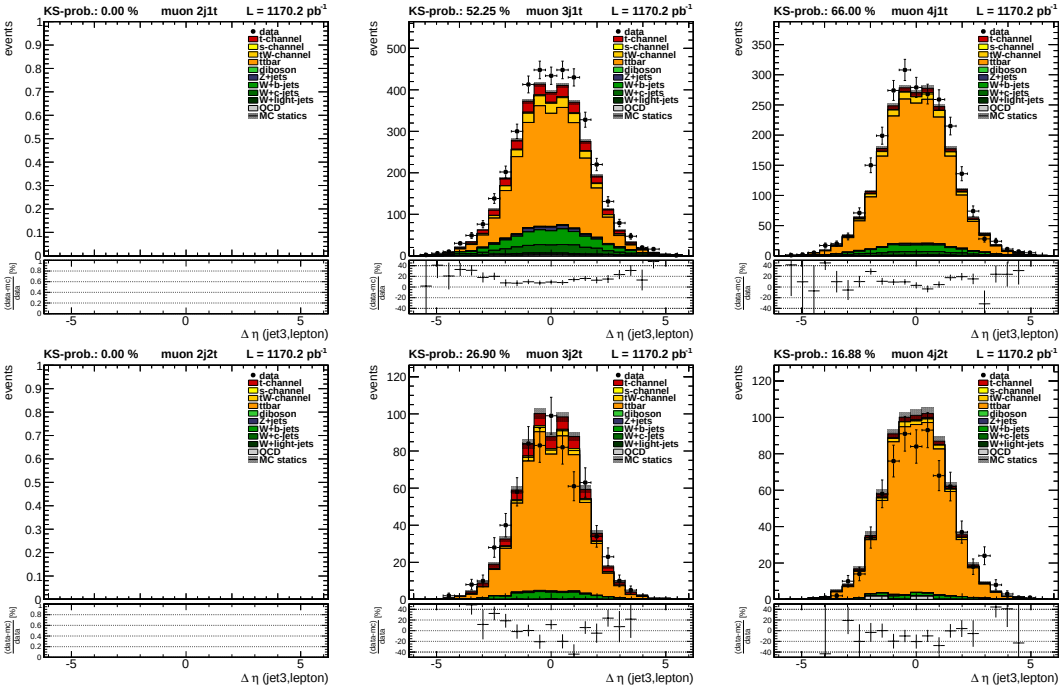


Figure A.43.: Difference in pseudo rapidity $\Delta\eta$ between third jet and lepton in muon channel

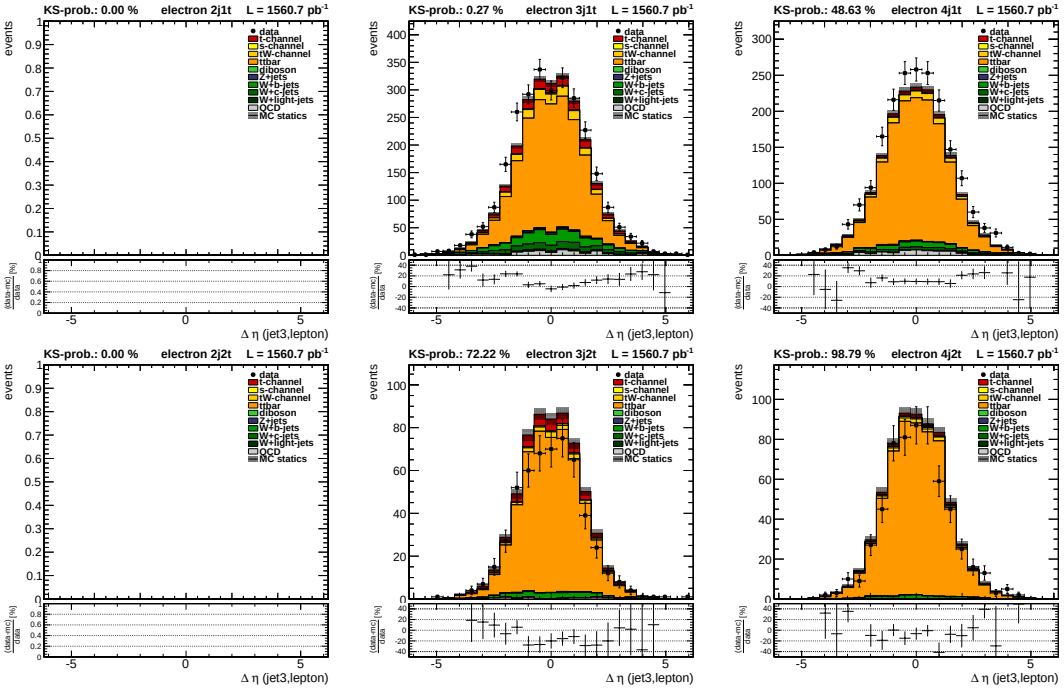


Figure A.44.: Difference in pseudo rapidity $\Delta\eta$ between third jet and lepton in electron channel

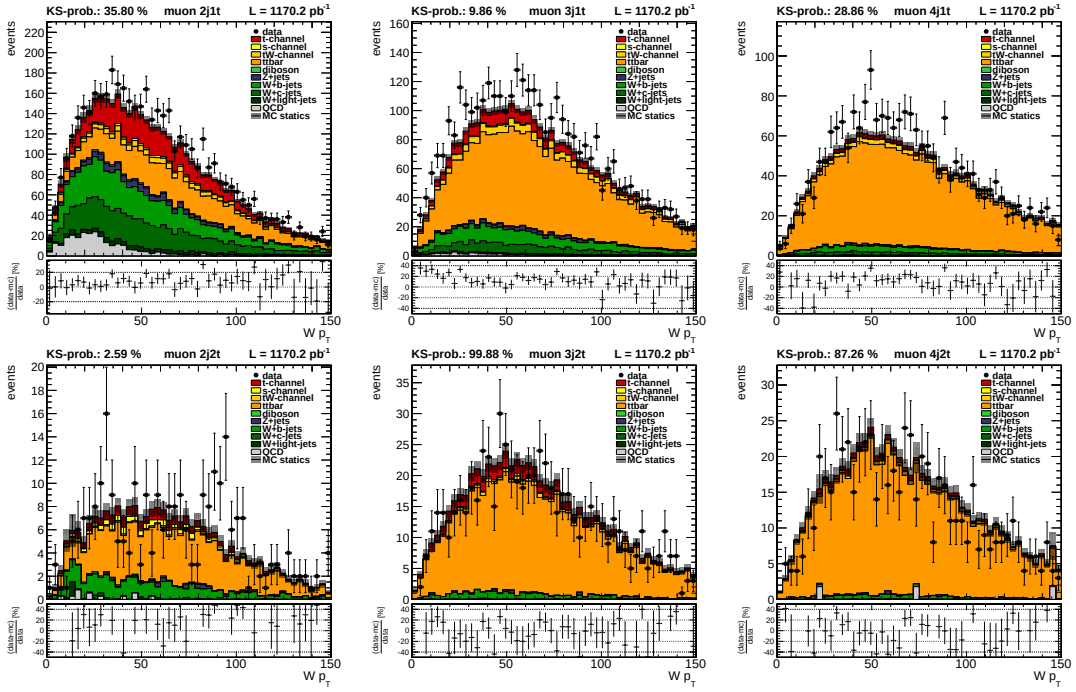


Figure A.45.: Reconstructed W boson candidate's transverse momentum p_T in muon channel

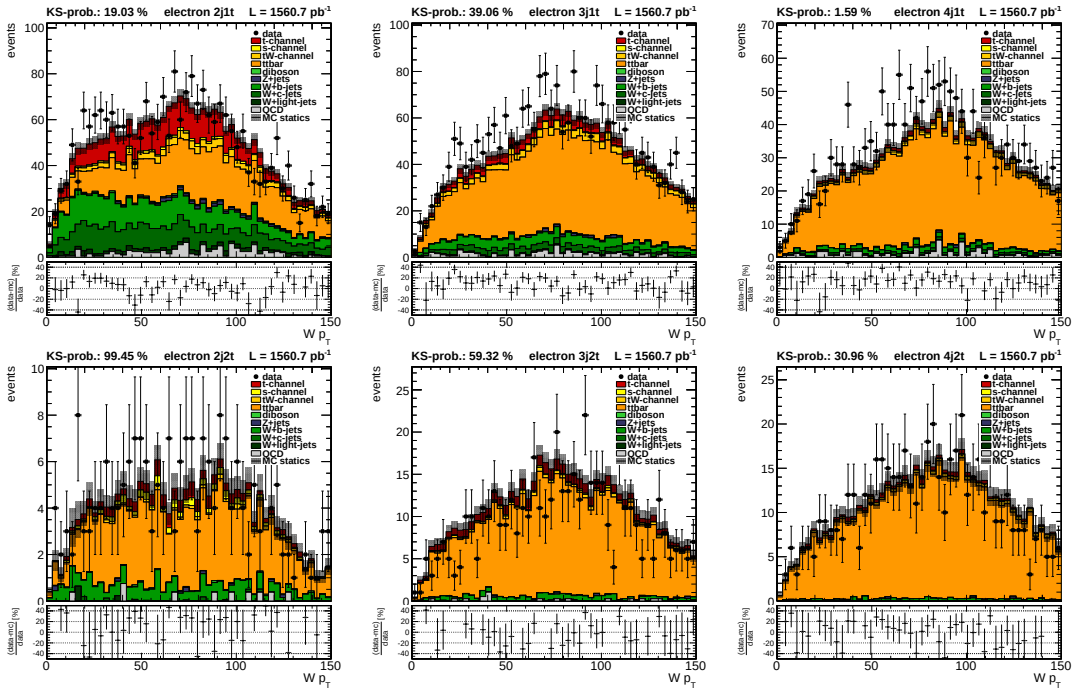


Figure A.46.: Reconstructed W boson candidate's transverse momentum p_T in electron channel

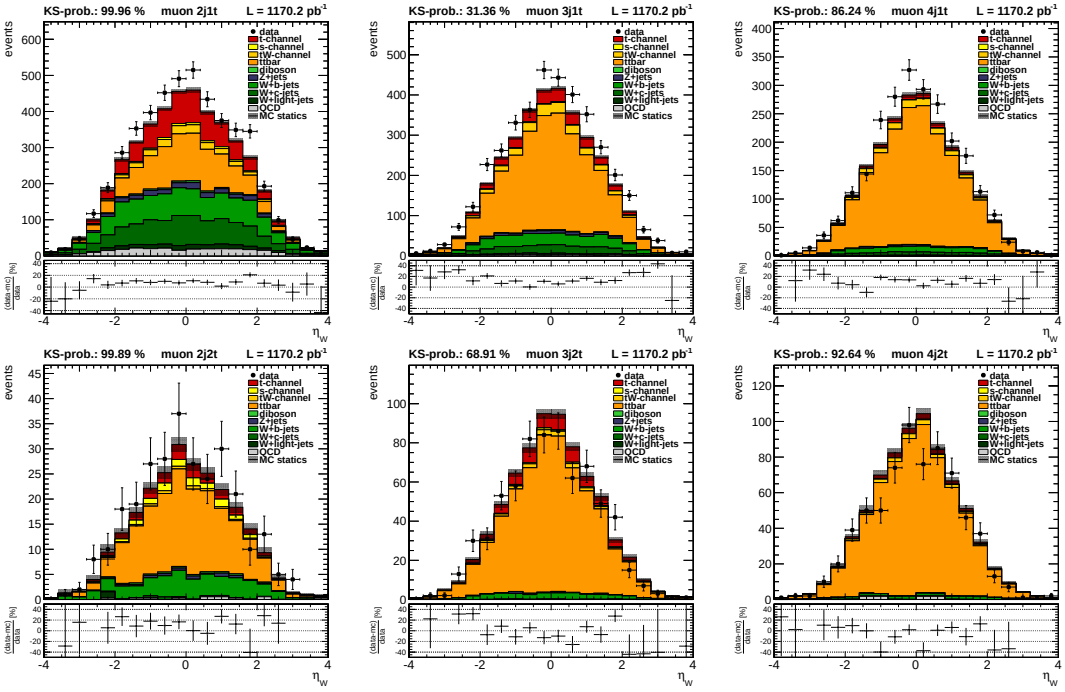


Figure A.47.: Reconstructed W boson candidate's pseudo rapidity η in muon channel

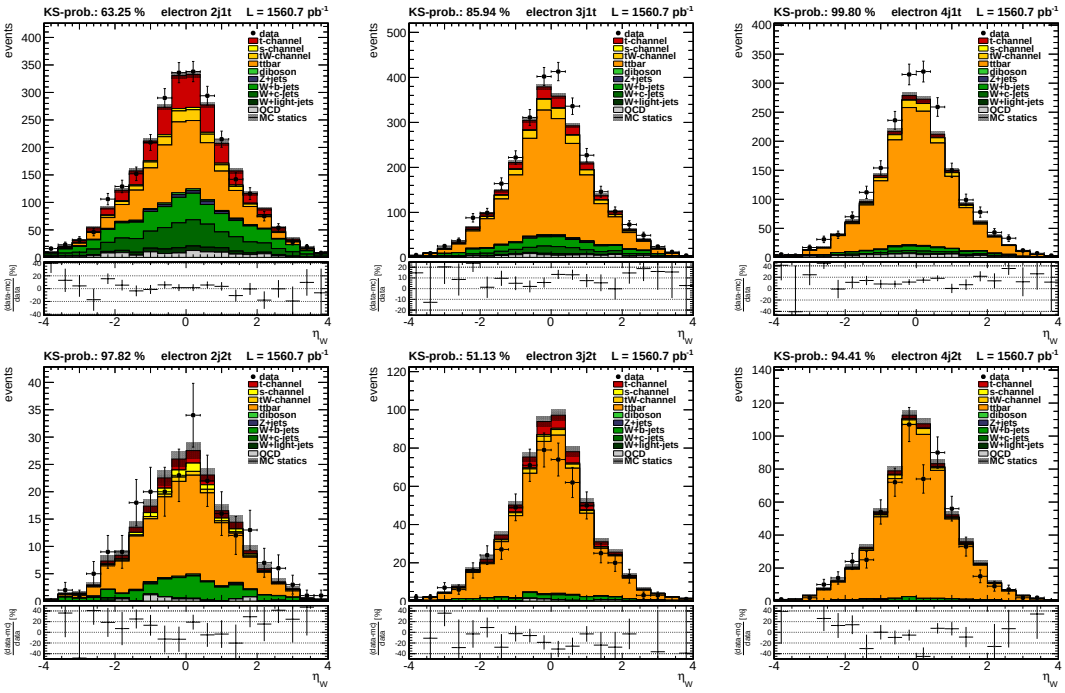


Figure A.48.: Reconstructed W boson candidate's pseudo rapidity η in electron channel

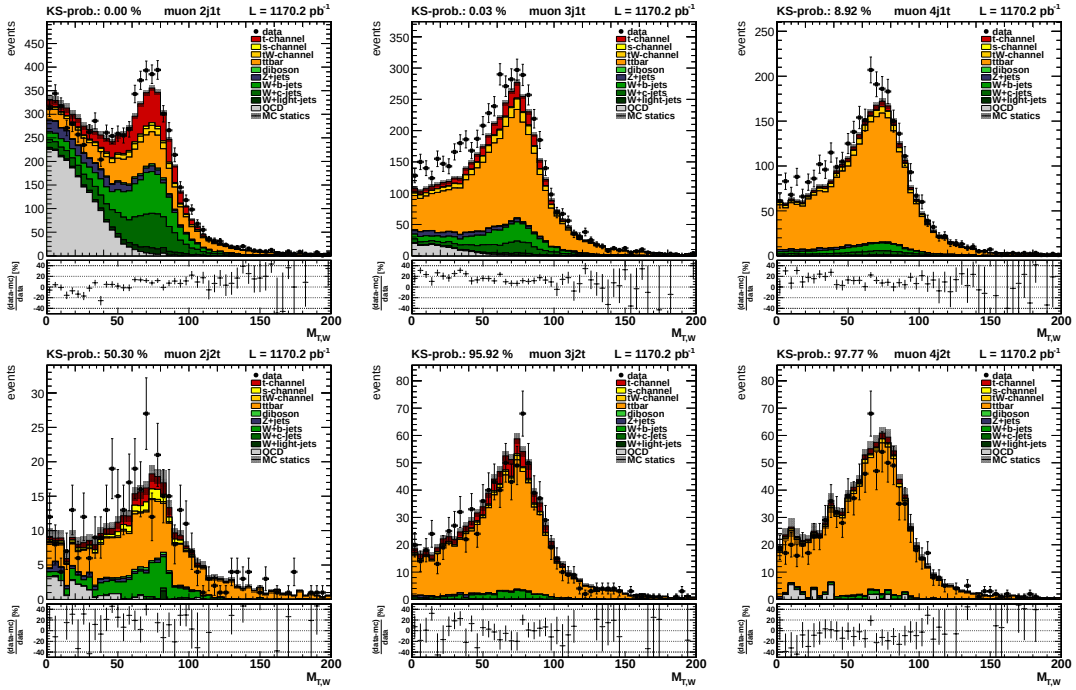


Figure A.49.: Transverse mass of the reconstructed W boson candidate MTW in the muon channel

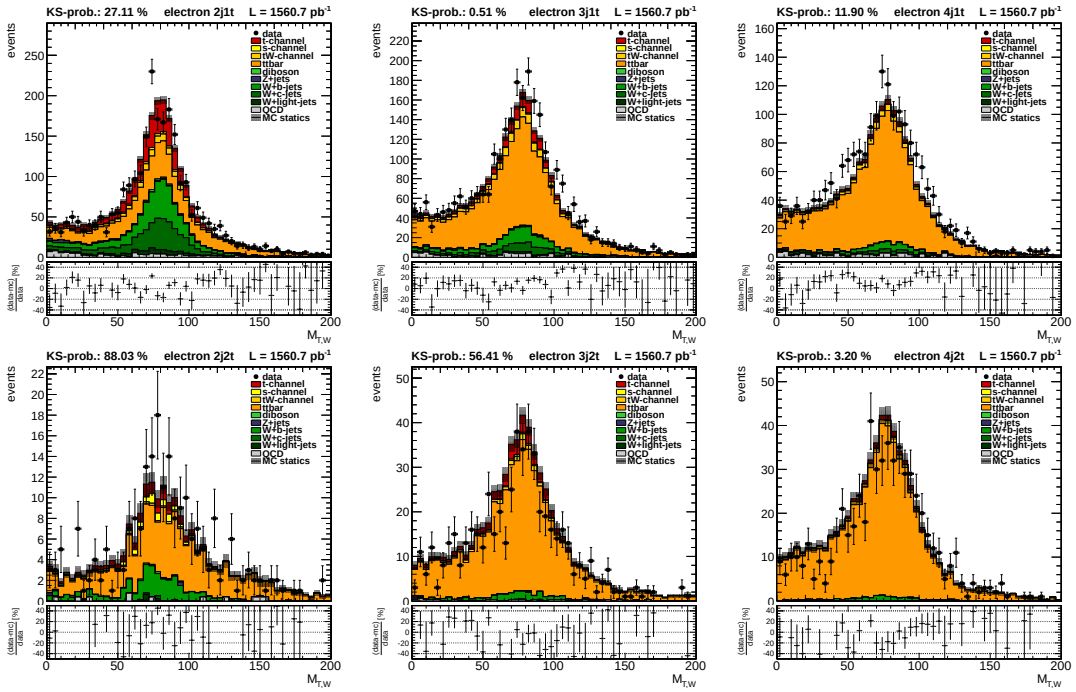


Figure A.50.: Transverse mass of the reconstructed W boson candidate MTW in the electron channel

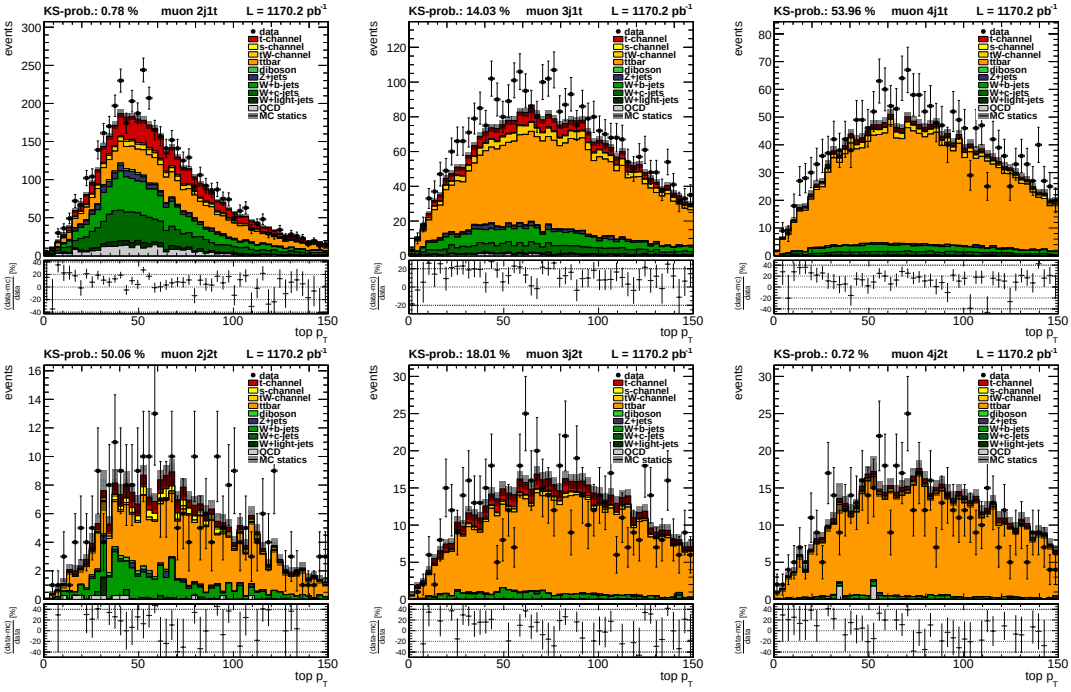


Figure A.51.: Transverse momentum of the reconstructed t quark candidate in the muon channel

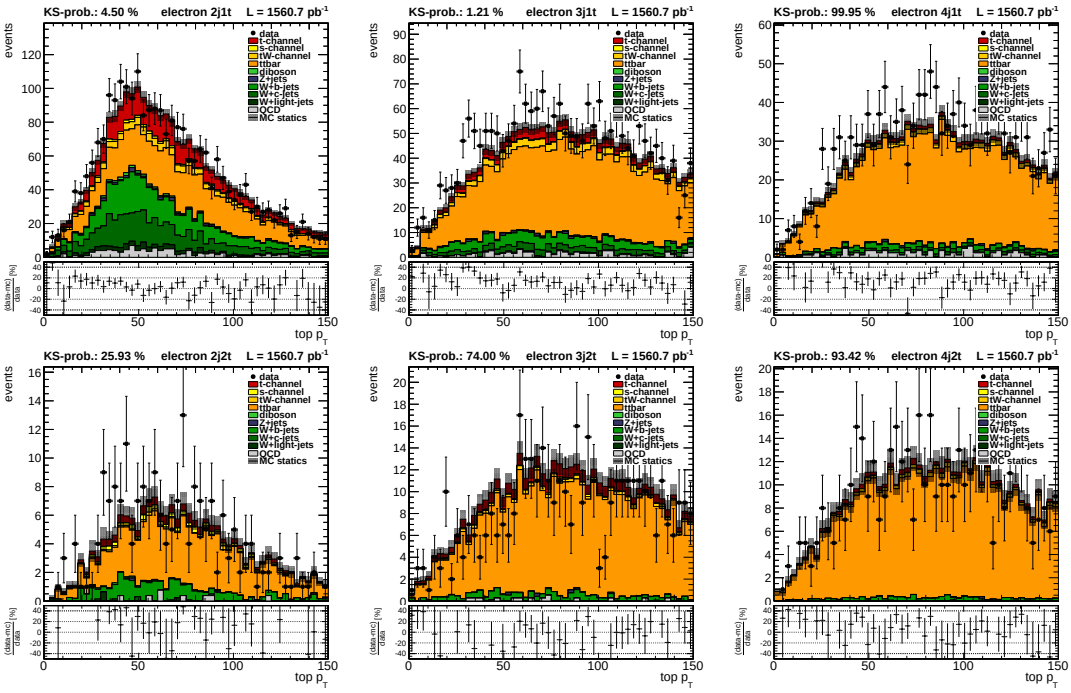


Figure A.52.: Transverse momentum of the reconstructed t quark candidate in the electron channel

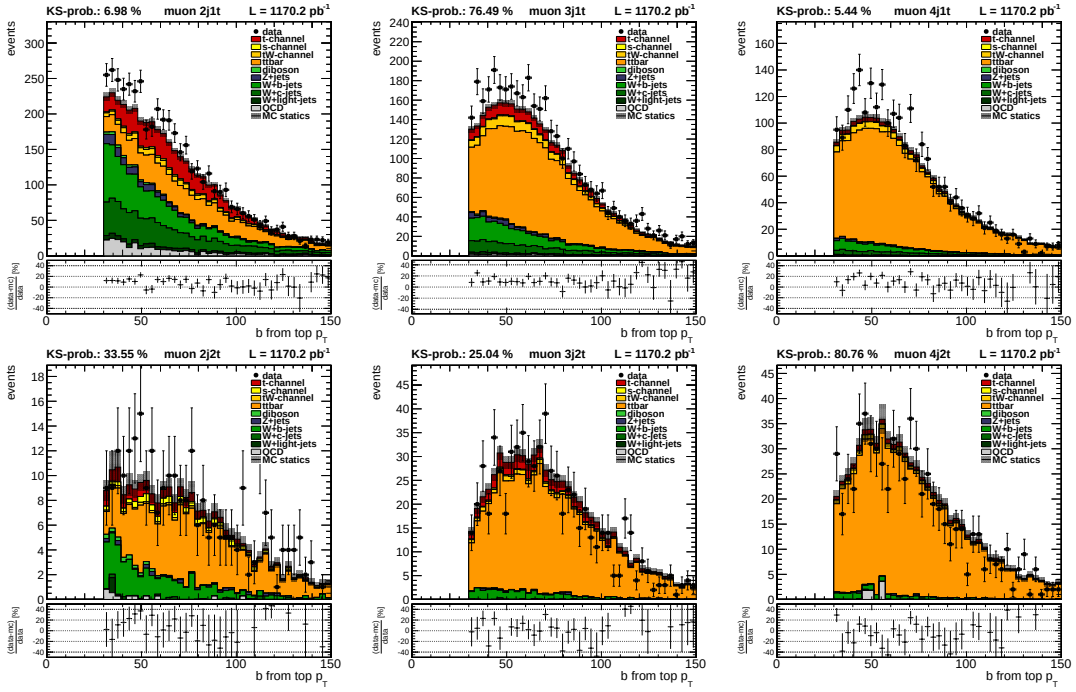


Figure A.53.: Transverse momentum of the b jet from the top decay in the muon channel

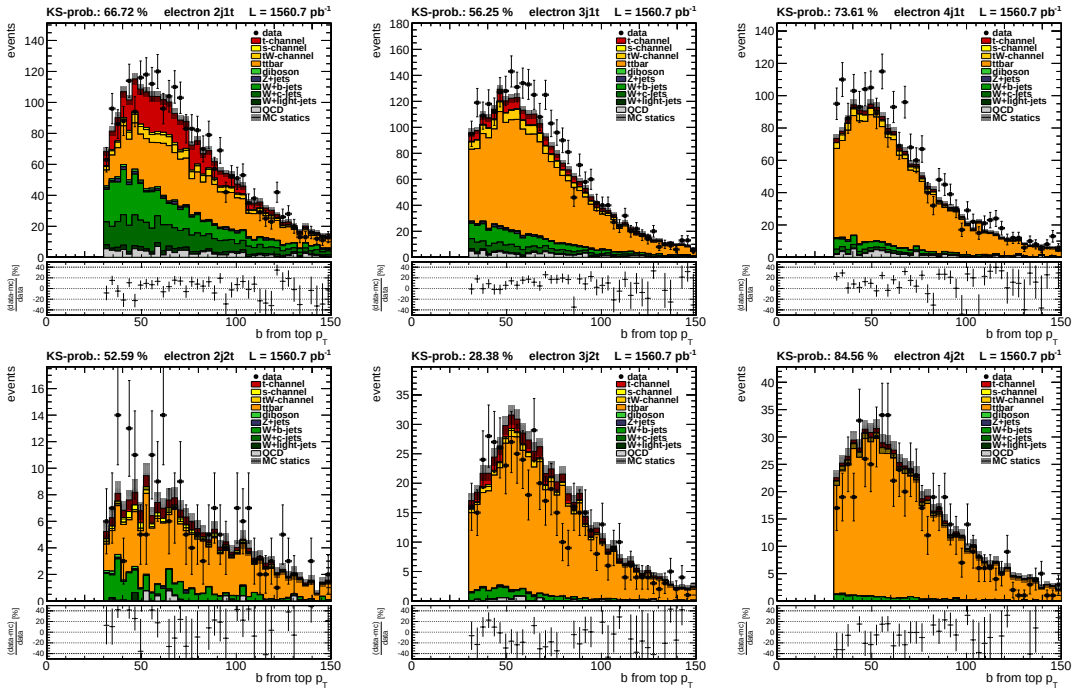


Figure A.54.: Transverse momentum of the b jet from the top decay in the electron channel

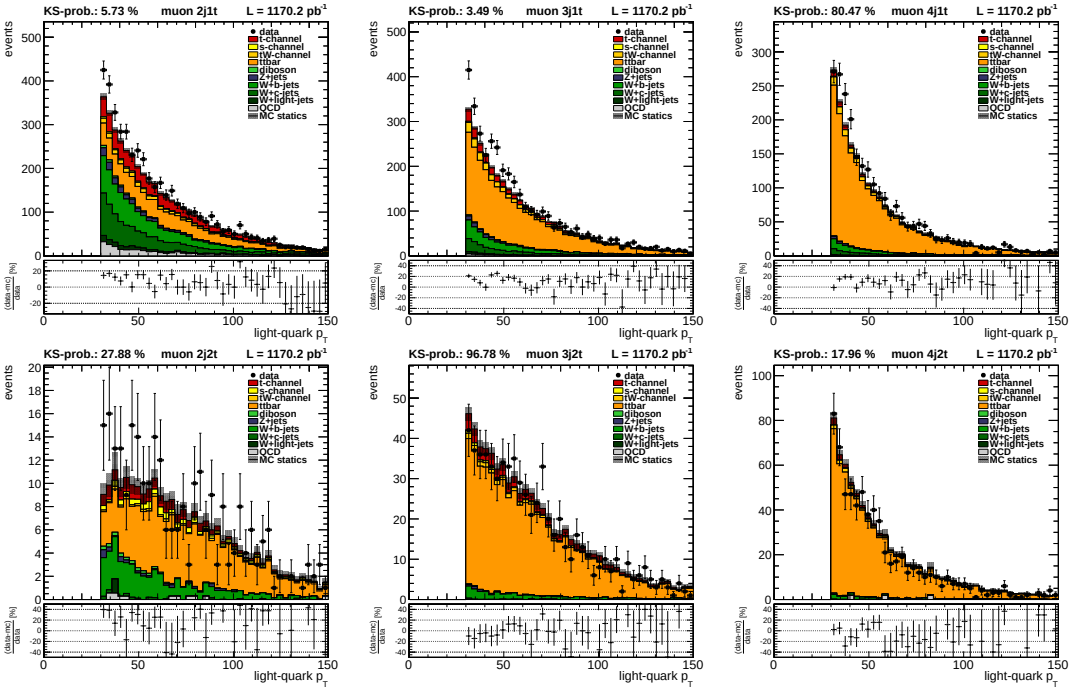


Figure A.55.: Transverse momentum of the light quark jet in the muon channel

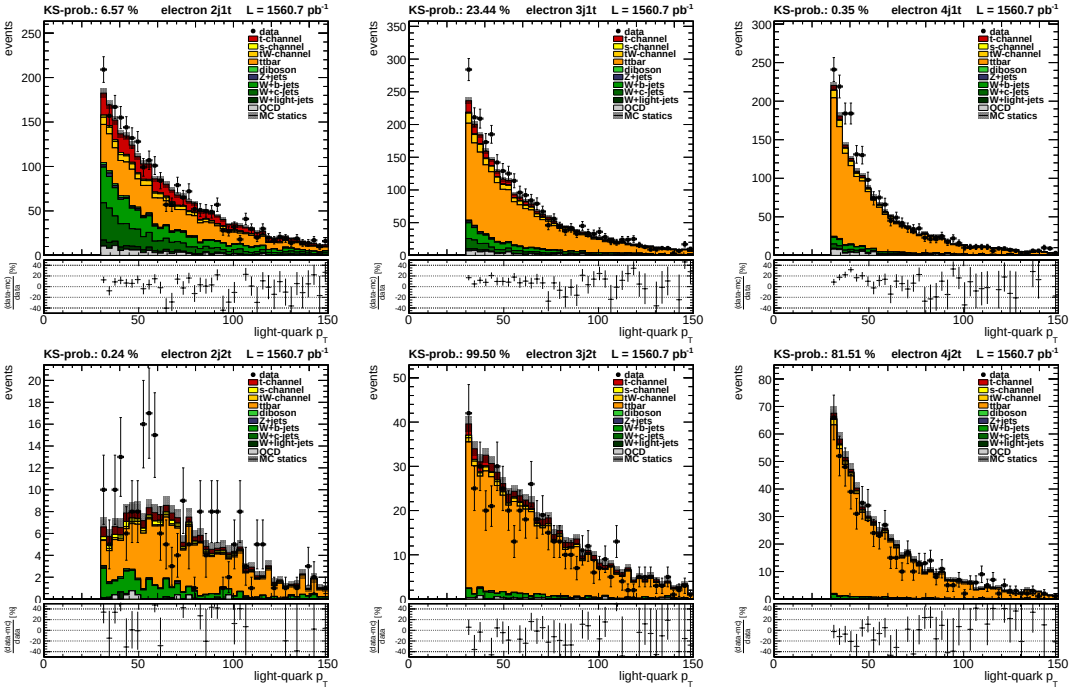


Figure A.56.: Transverse momentum of the light quark jet in the electron channel

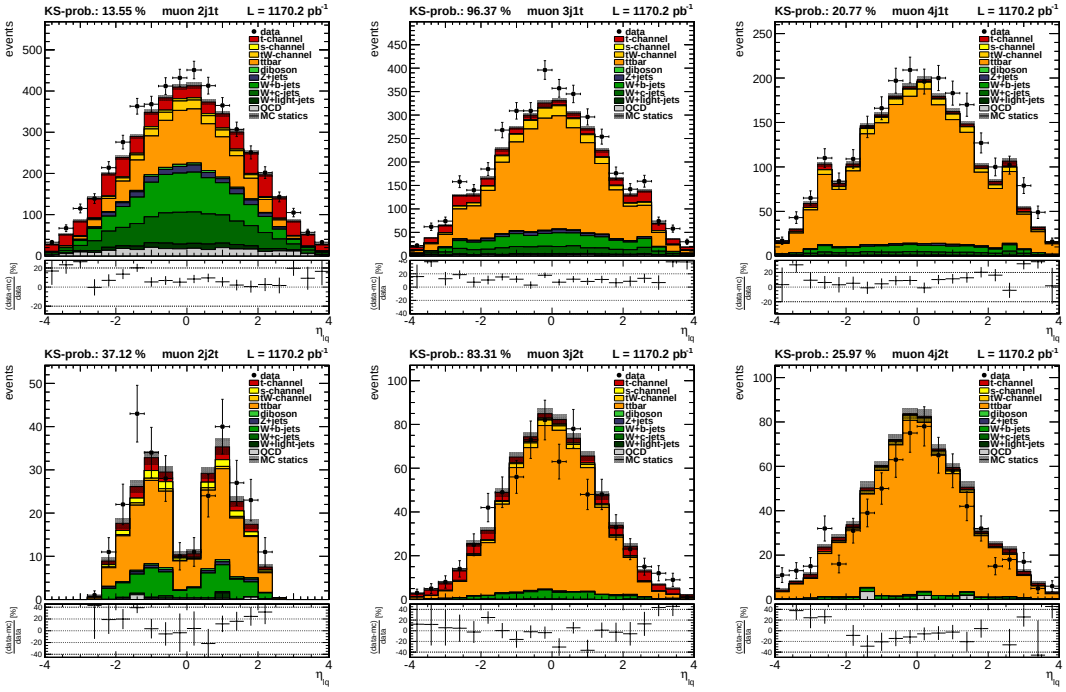


Figure A.57.: Pseudo rapidity η of the light quark jet in the muon channel

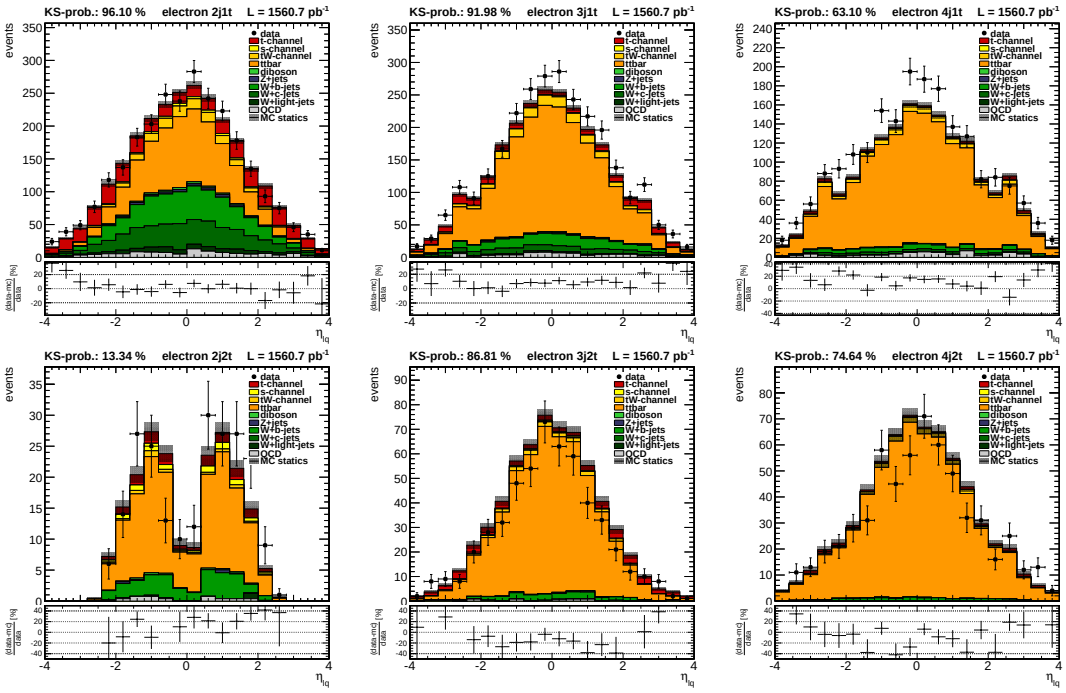


Figure A.58.: Pseudo rapidity η of the light quark jet in the electron channel

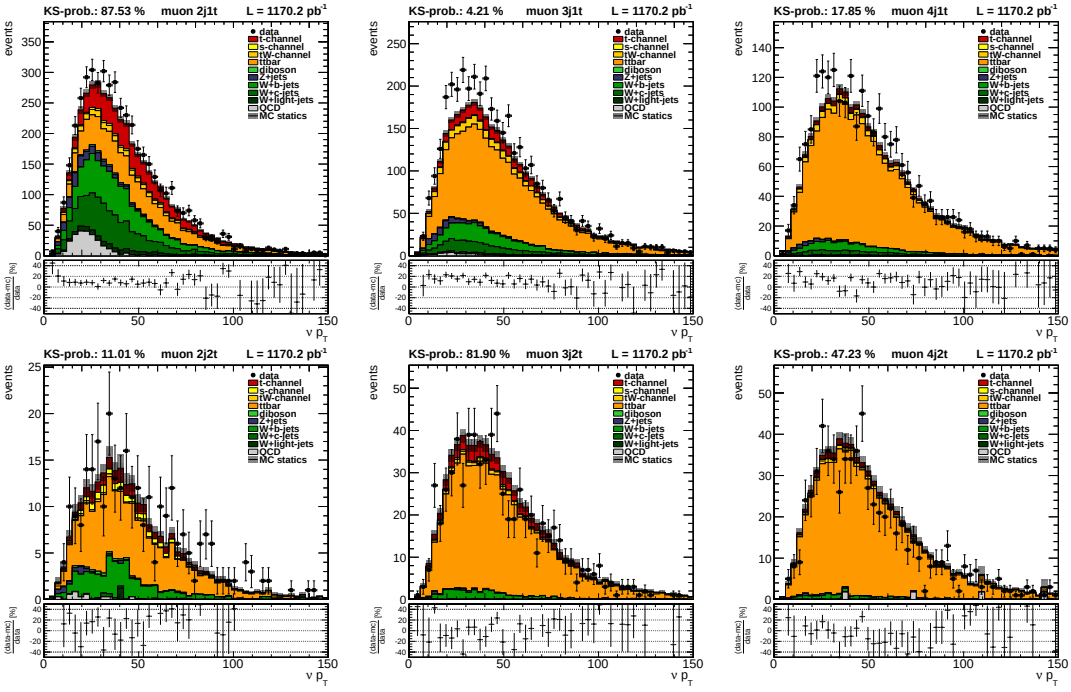


Figure A.59.: Transverse momentum p_T of the reconstructed neutrino candidate in the muon channel

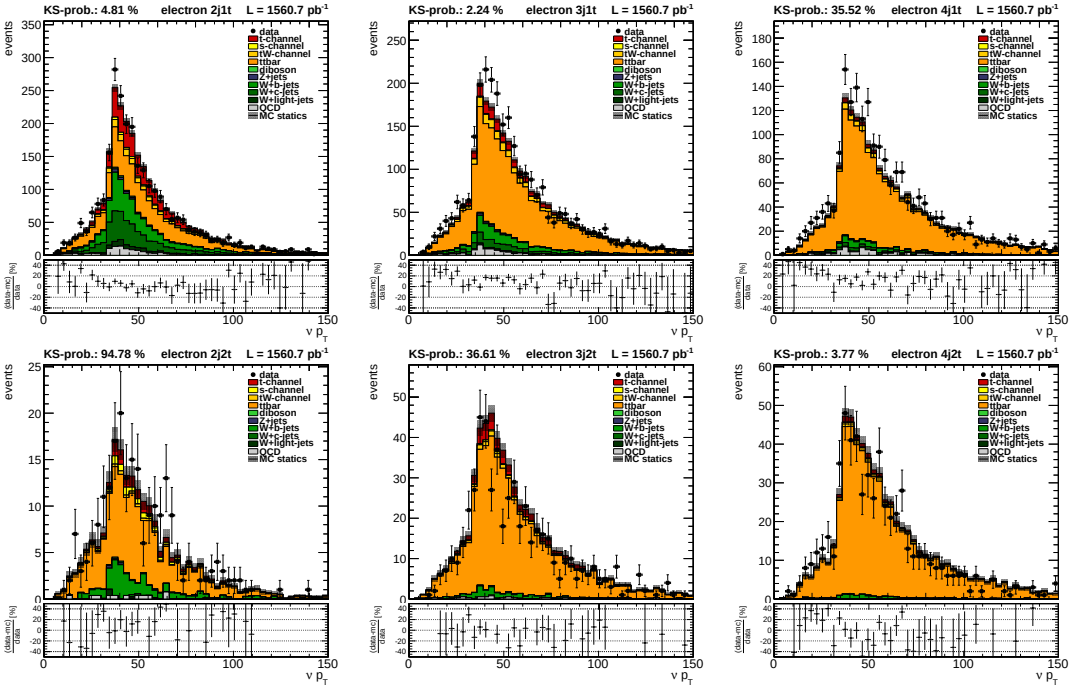


Figure A.60.: Transverse momentum p_T of the reconstructed neutrino candidate in the electron channel

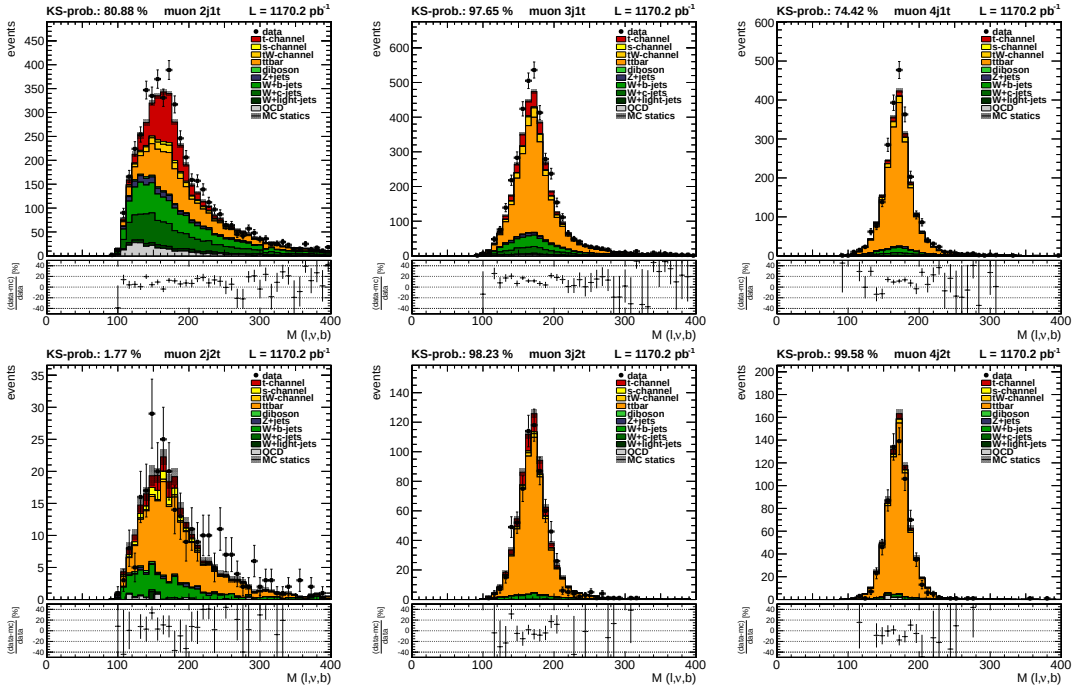


Figure A.61.: Invariant mass of the top quark decay products candidates μ, ν and b

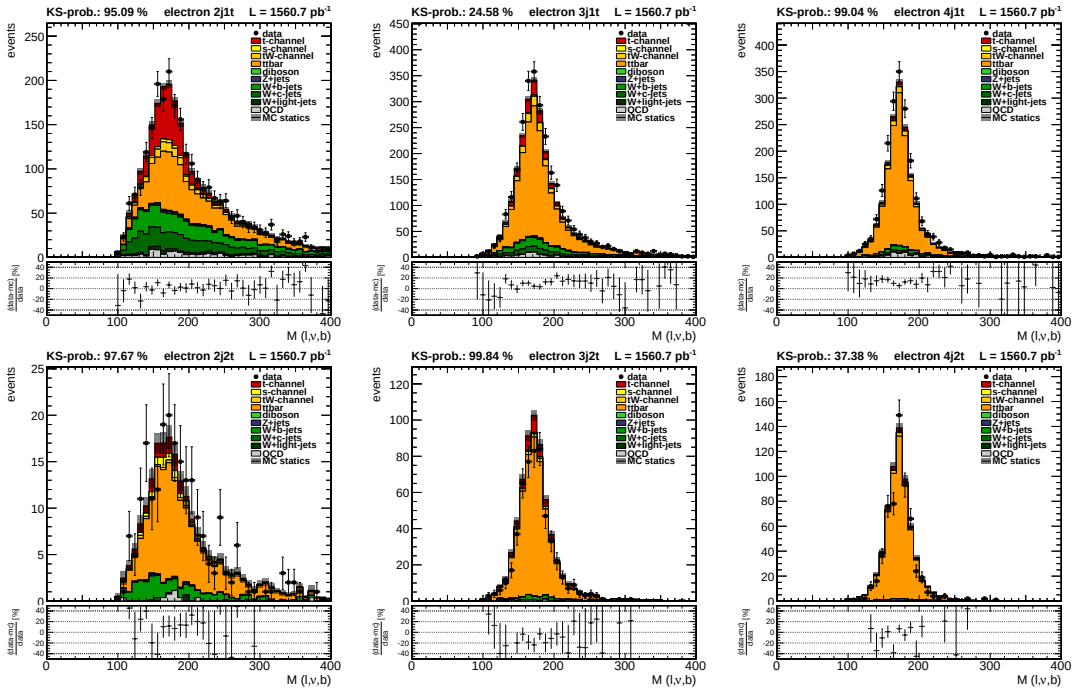


Figure A.62.: Invariant mass of the top quark decay products candidates e, ν and b

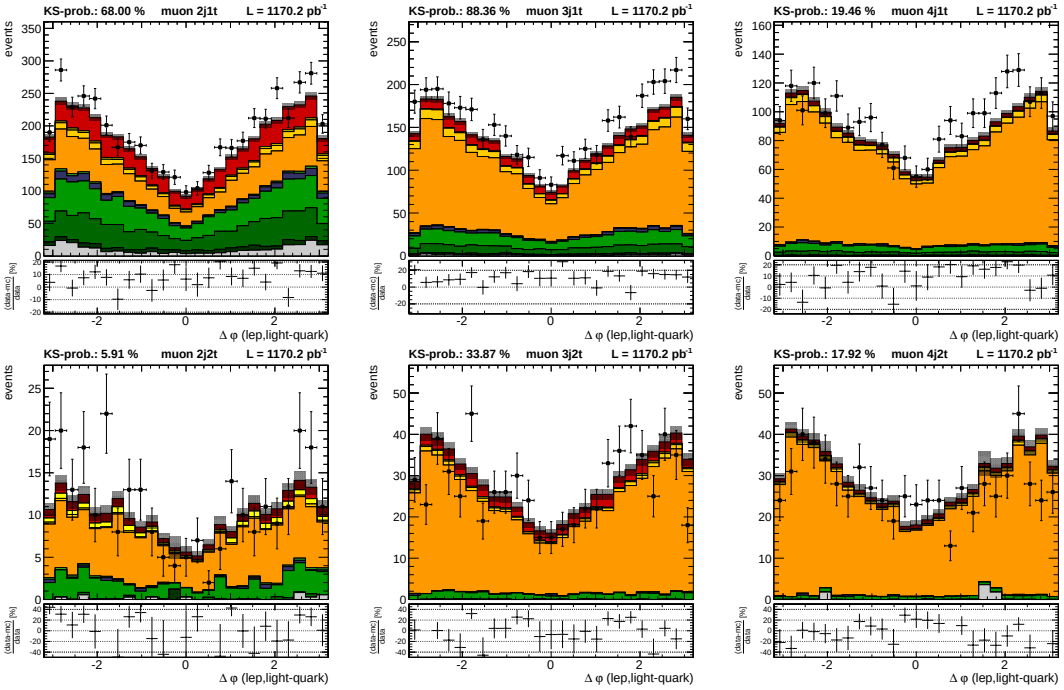


Figure A.63.: Difference in φ between the muon and the light quark jet

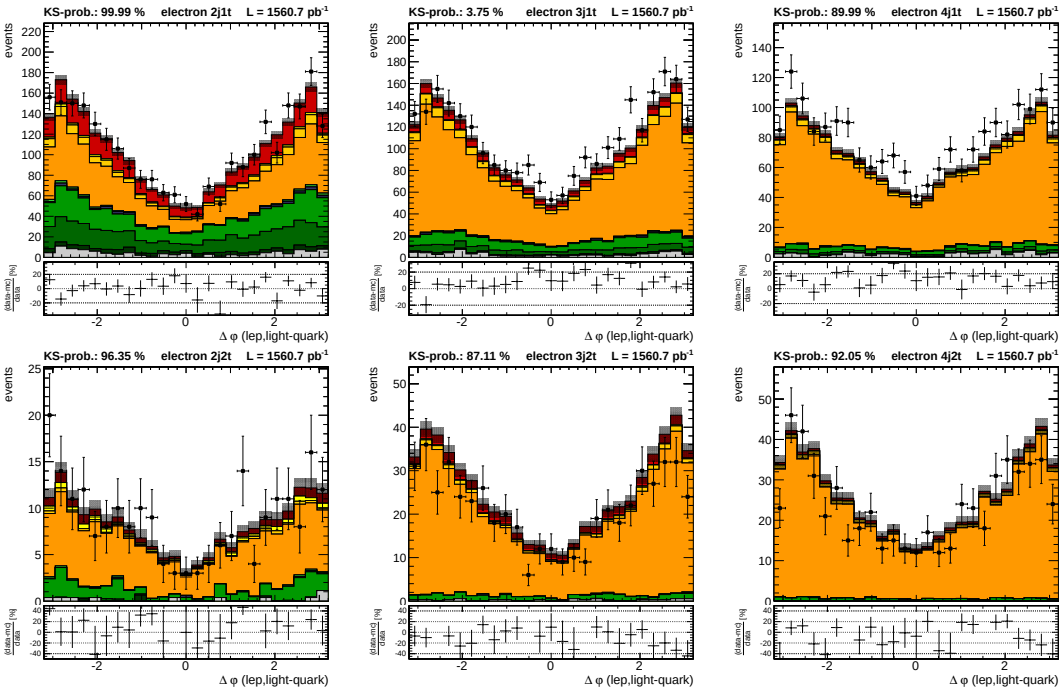


Figure A.64.: Difference in φ between the electron and the light quark jet

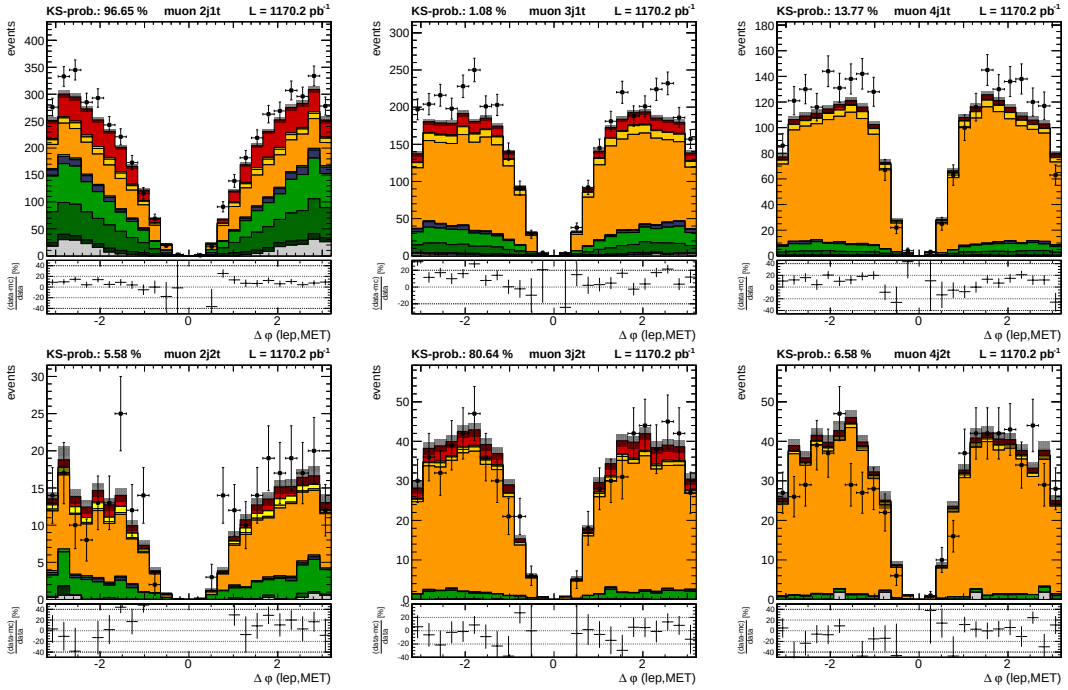


Figure A.65.: Difference in φ between the muon and the missing transverse energy MET

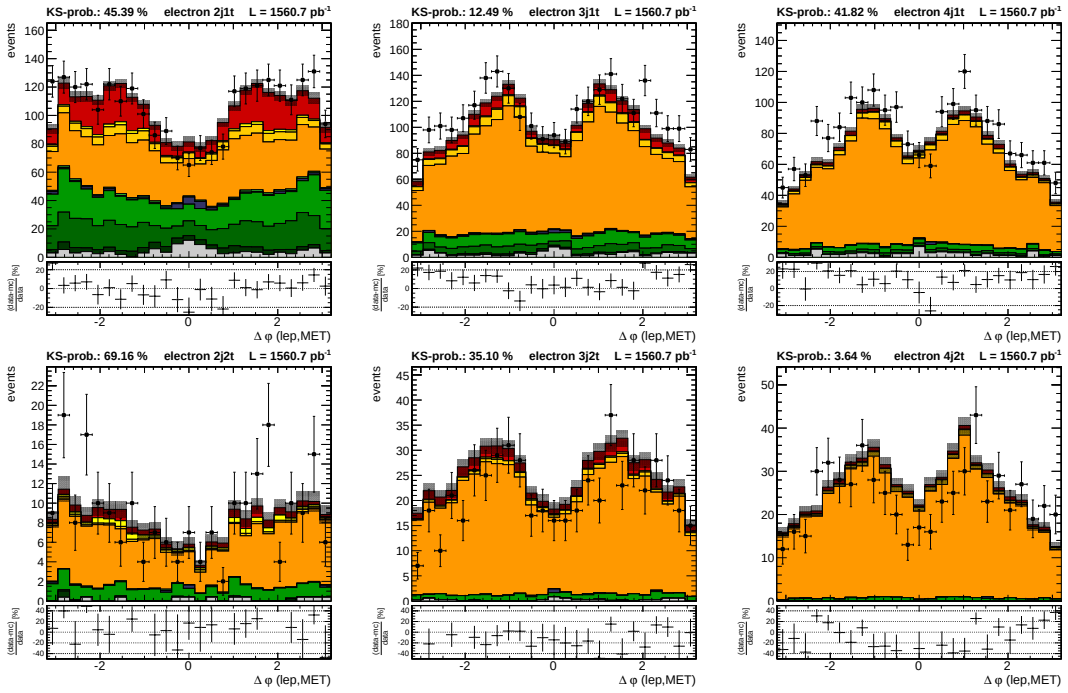


Figure A.66.: Difference in φ between the electron and the missing transverse energy MET

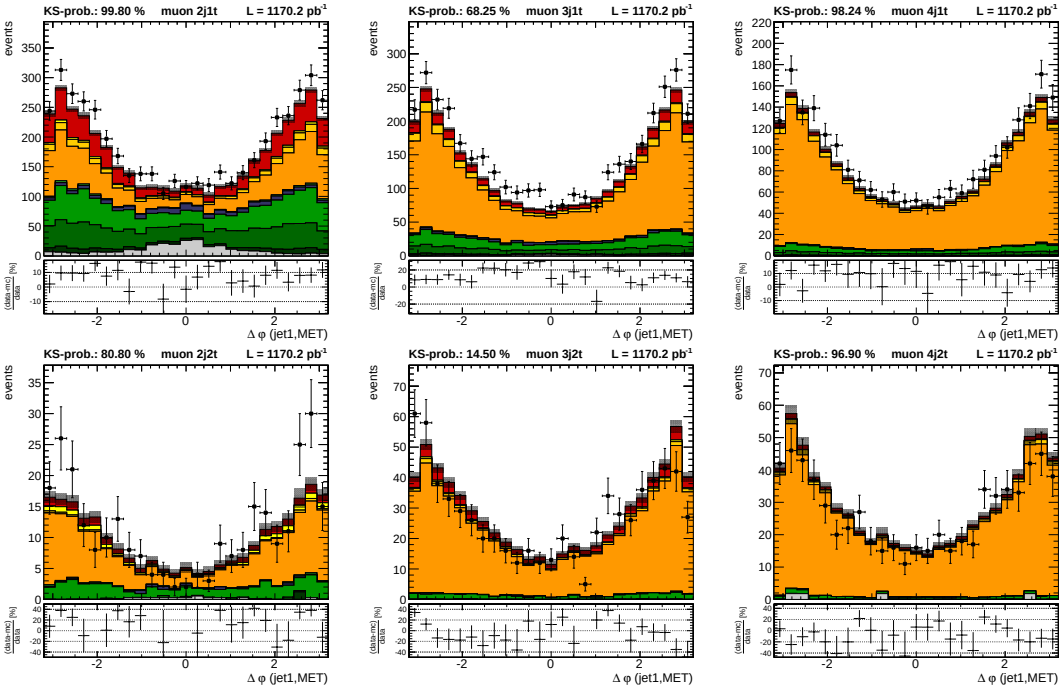


Figure A.67.: Difference in φ between the first jet and the missing transverse energy MET in the muon channel

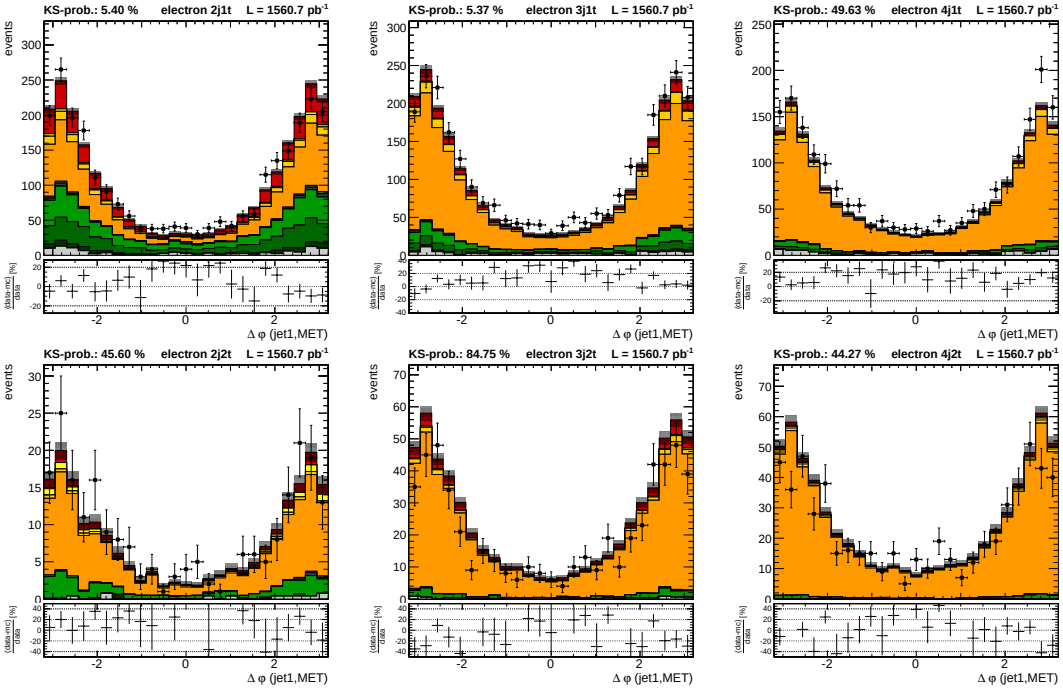


Figure A.68.: Difference in φ between the first jet and the missing transverse energy MET in the electron channel

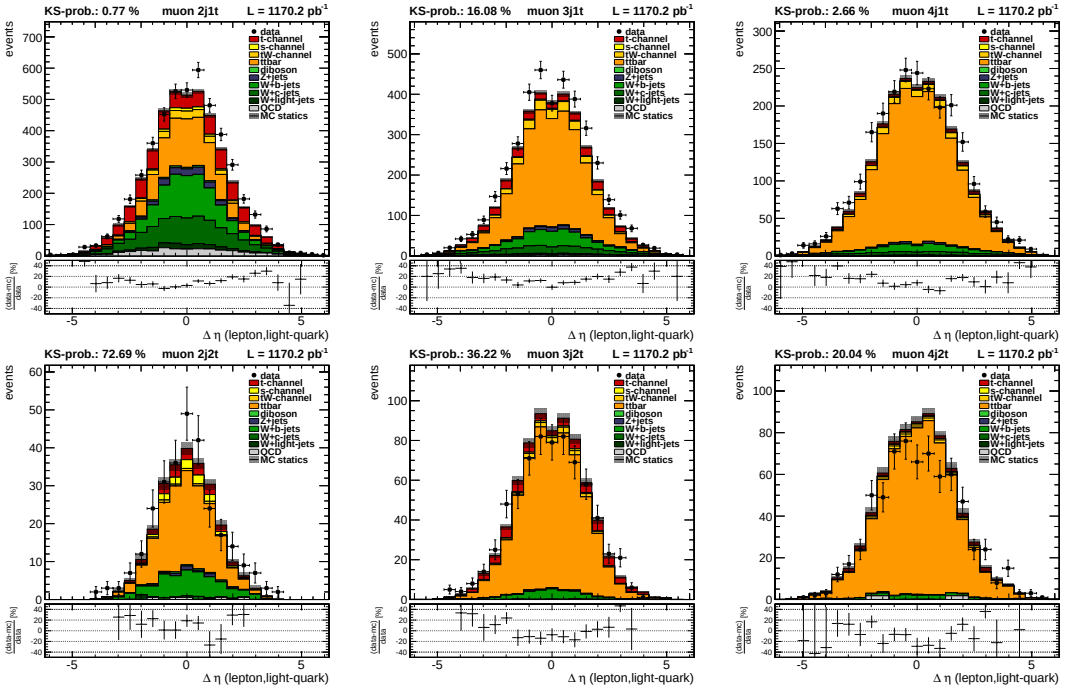


Figure A.69.: Difference in pseudo η between the muon and the light quark jet

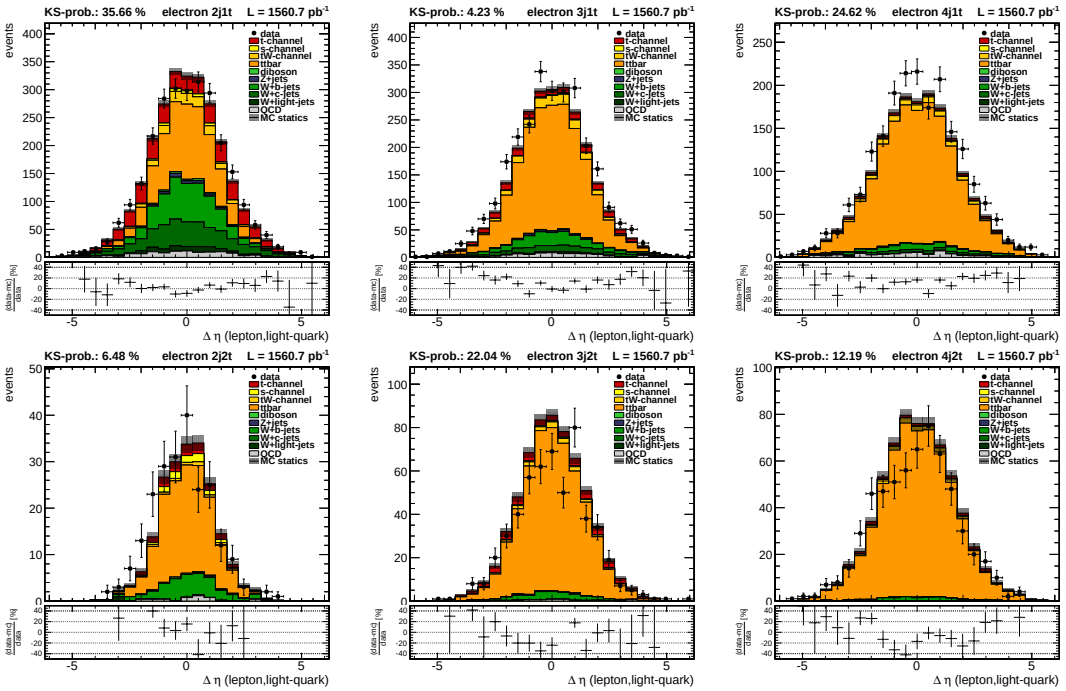


Figure A.70.: Difference in pseudo η between the electron and the light quark jet

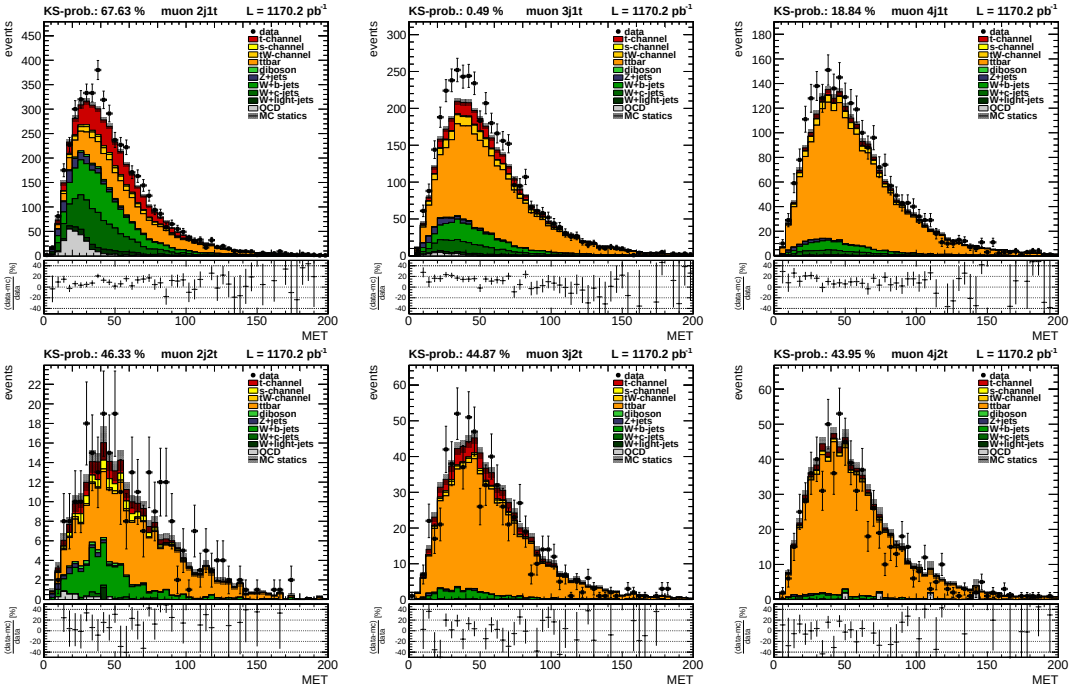


Figure A.71.: Missing transverse energy in the muon channel

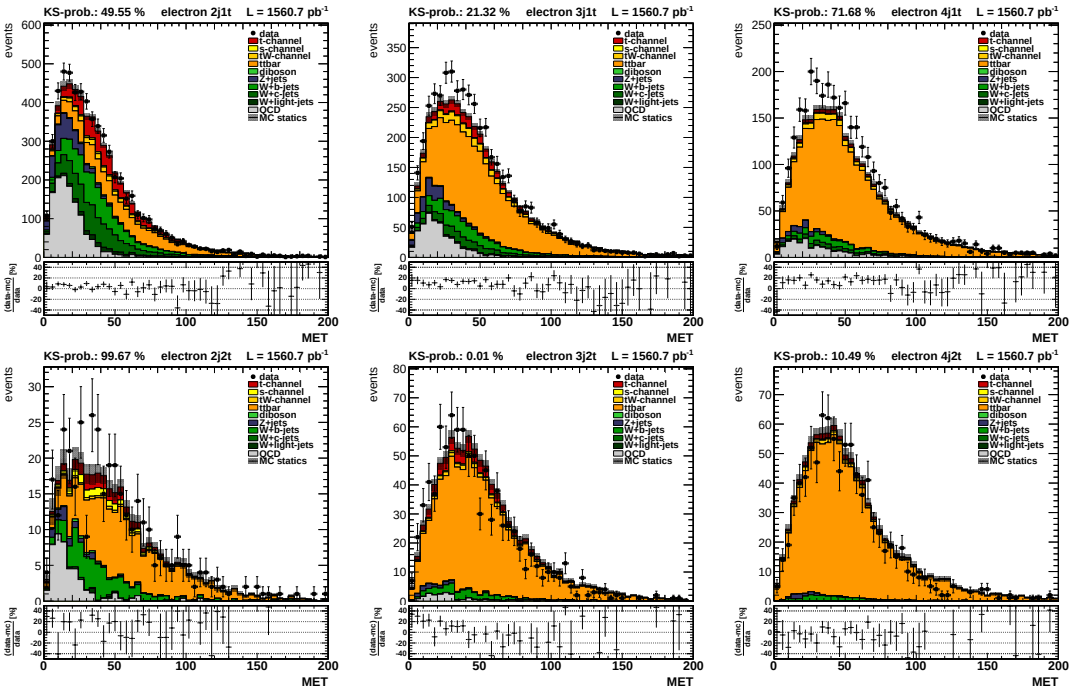


Figure A.72.: Missing transverse energy in the electron channel

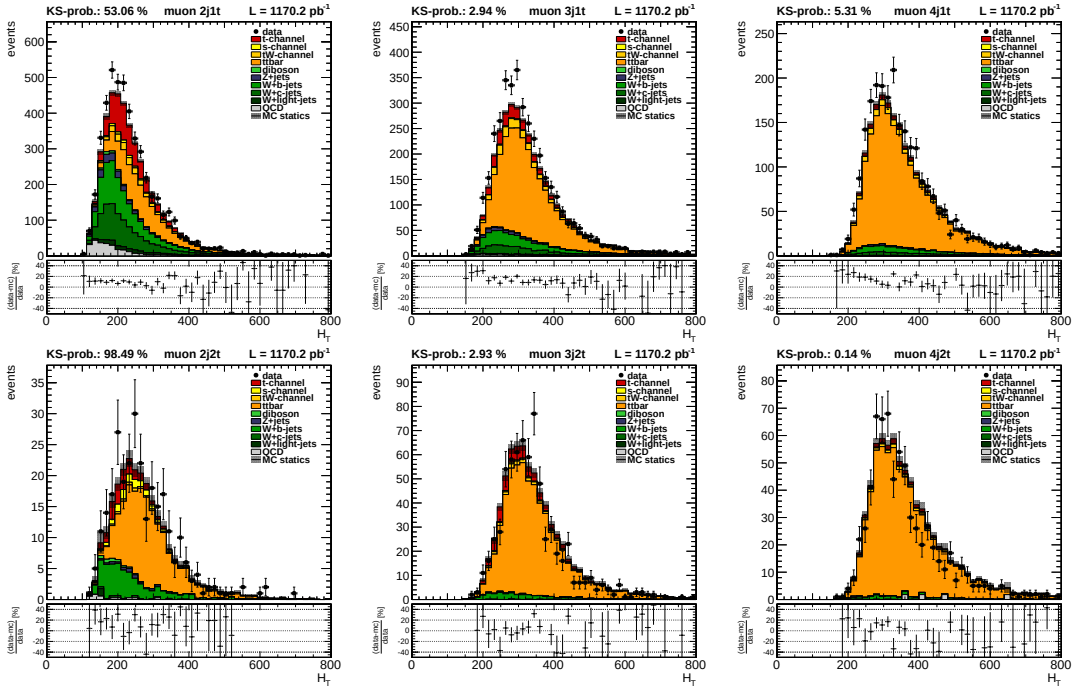


Figure A.73.: Absolute sum of all transverse momenta for all jets, muon and MET

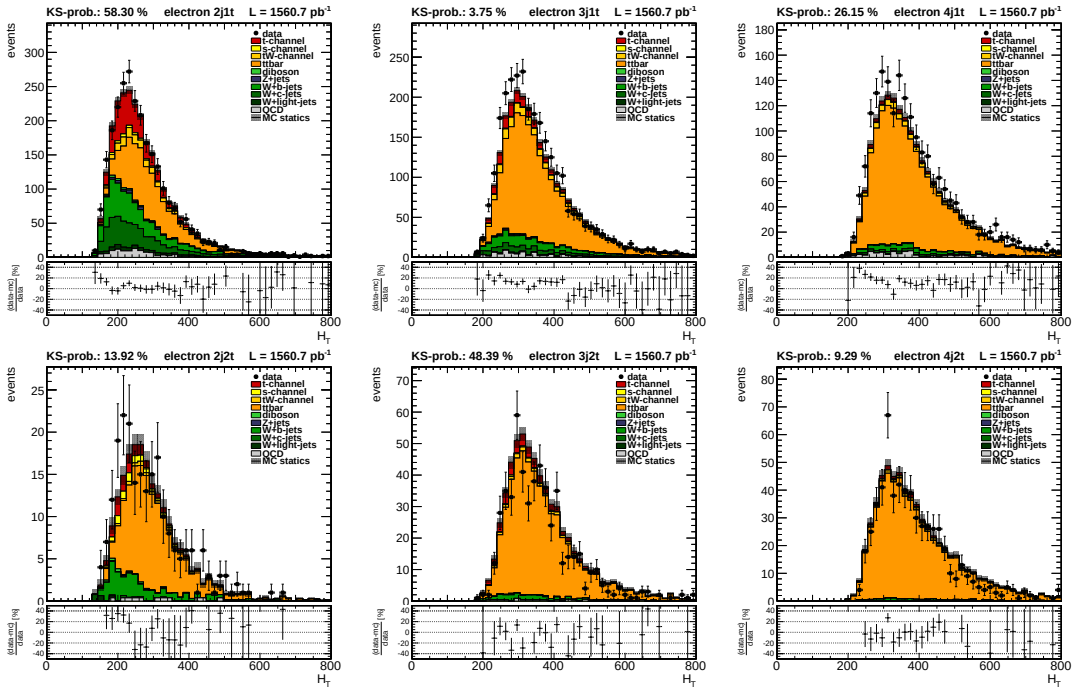


Figure A.74.: Absolute sum of all transverse momenta for all jets, electron and MET

B. NeuroBayes discriminator distributions

In this chapter the NeuroBayes discriminator distributions are given for three different sets of scale paramters for the process templates:

1. On pages 178 and 180 all templates are scaled to the expected number of events for each process. The QCD contribution is estimated from side band fits.
2. On pages 179 and 181 all templates are scaled and morphed to the best fitting shape, derived from individual fits in both lepton channels.
3. On pages 182 and 183 all templates are scaled and morphed to the best fitting shape, derived from a simulatenous (combined) fit in both lepton channels.

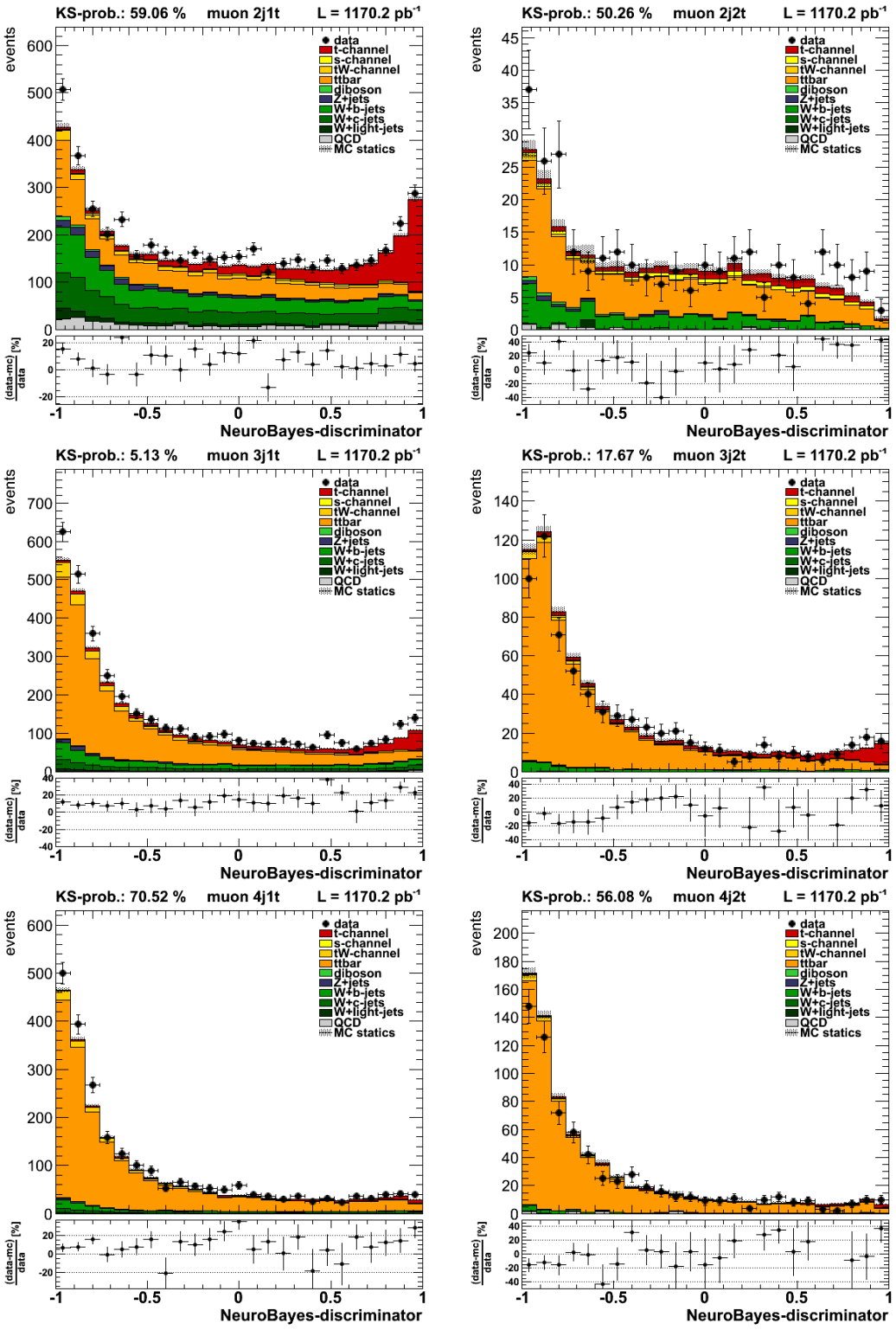


Figure B.1.: NeuroBayes discriminator distributions in the muon channel before the fit.

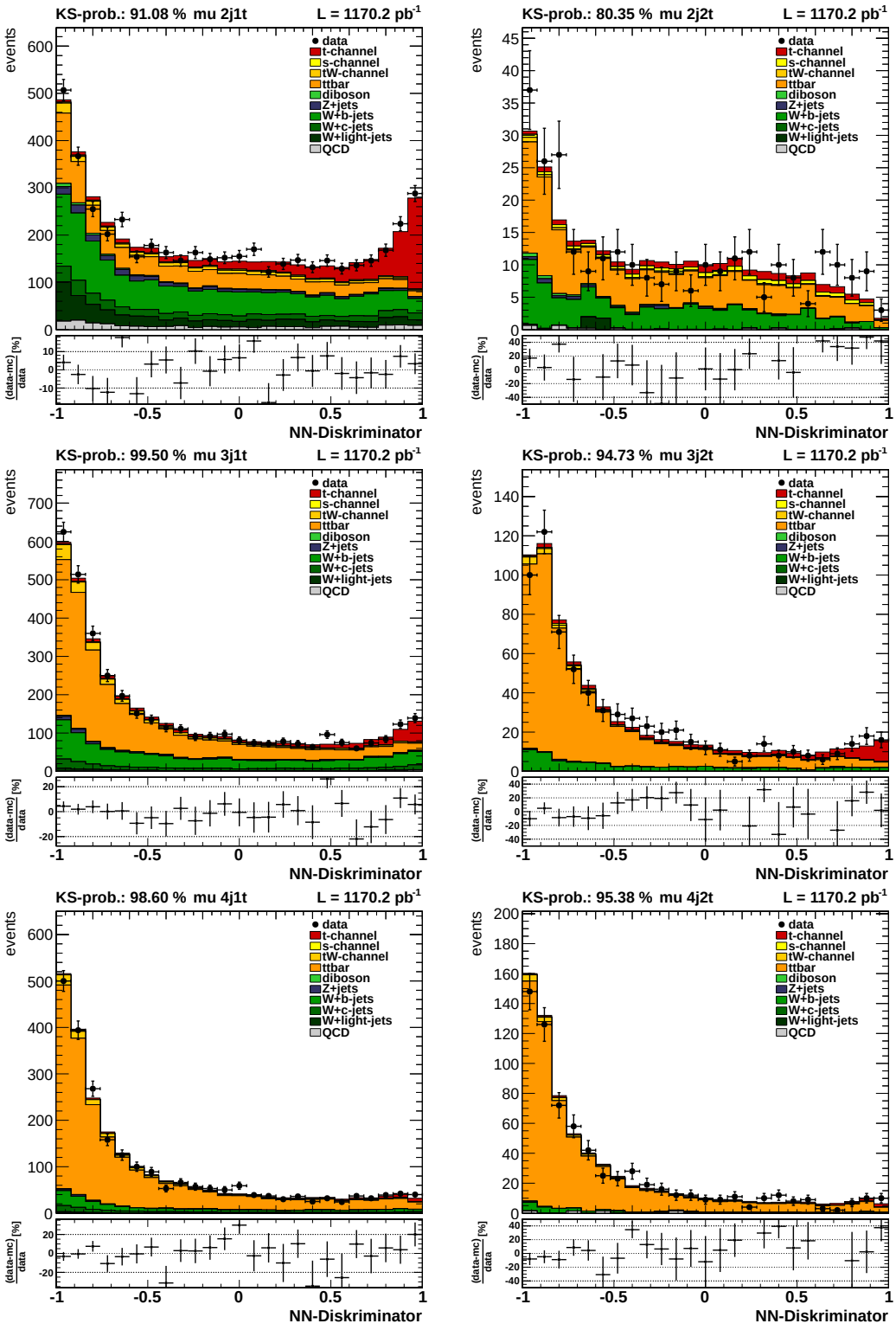


Figure B.2.: NeuroBayes discriminator distributions in the muon channel after the fit.

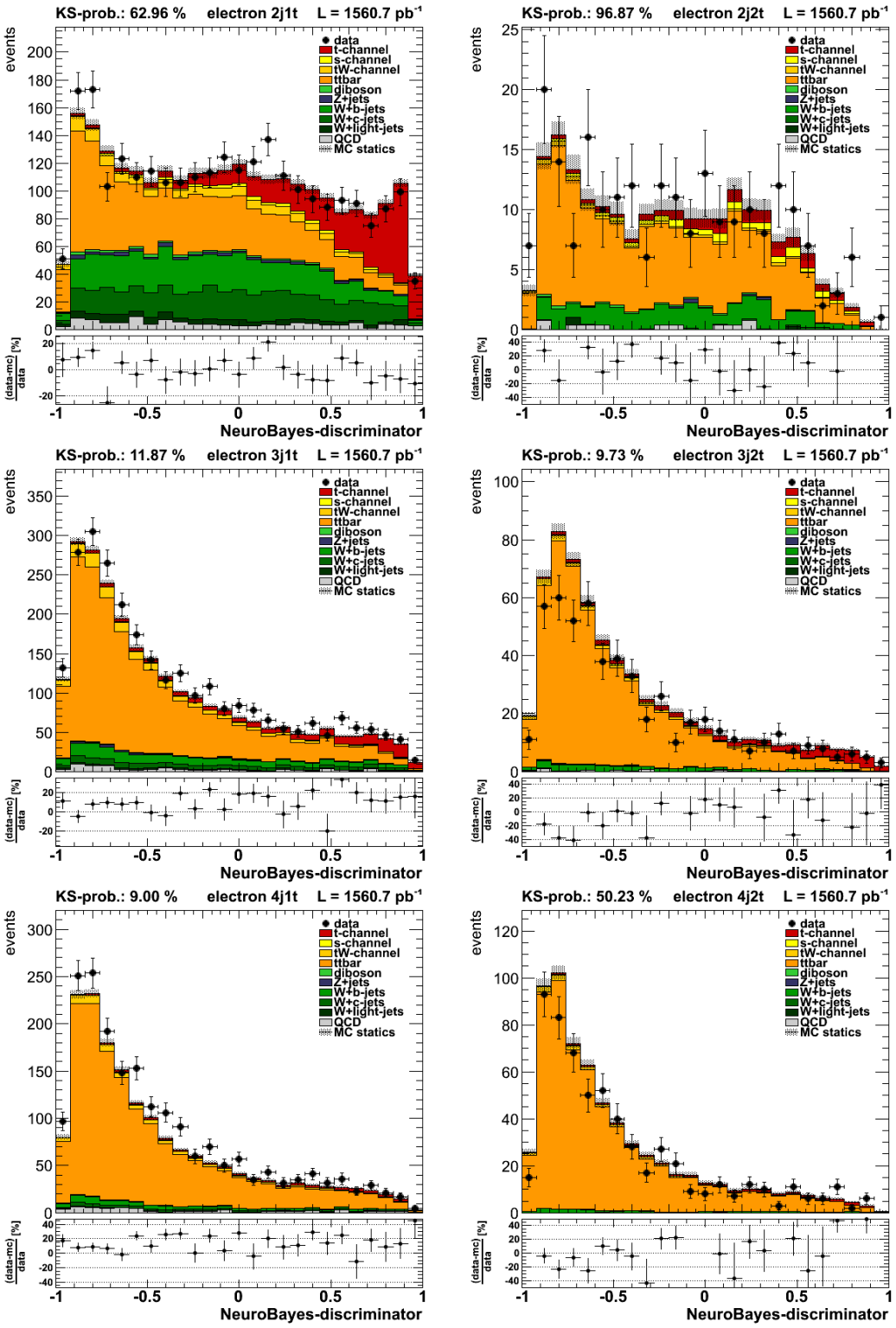


Figure B.3.: NeuroBayes discriminator distributions in the electron channel before the fit.

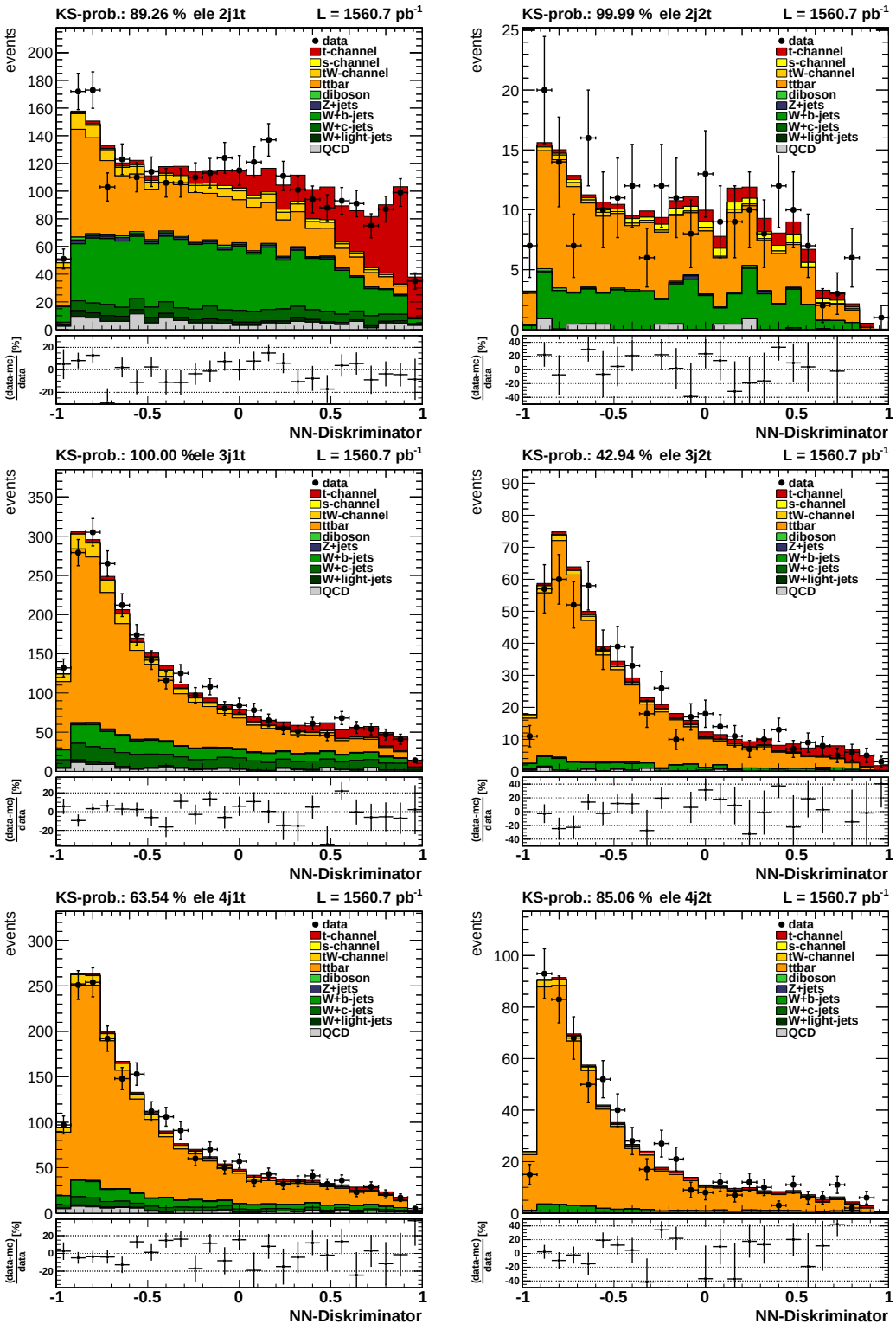


Figure B.4.: NeuroBayes discriminator distributions in the electron channel after the fit.

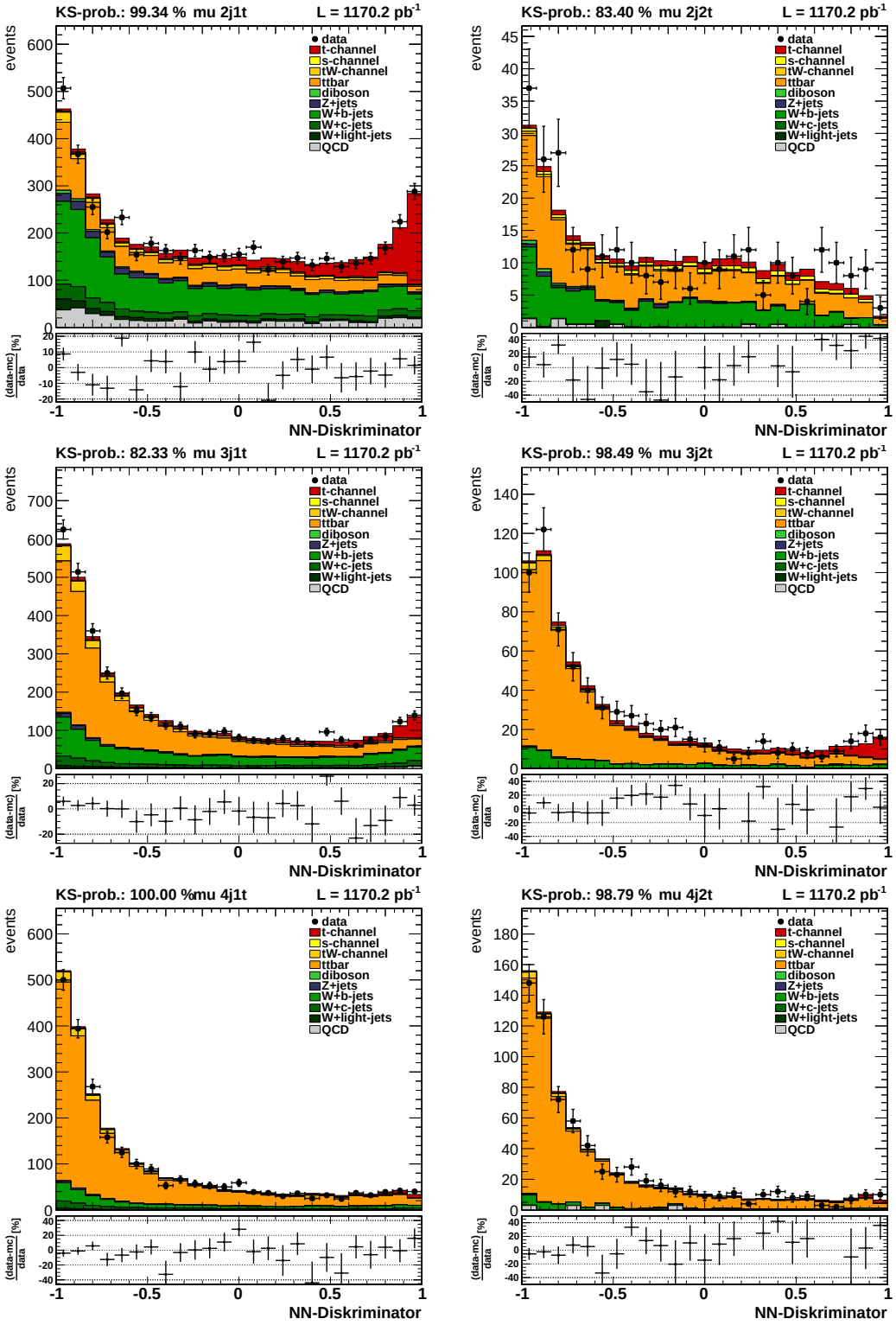


Figure B.5.: NeuroBayes discriminator distributions in the muon channel after the simulatenous fit.

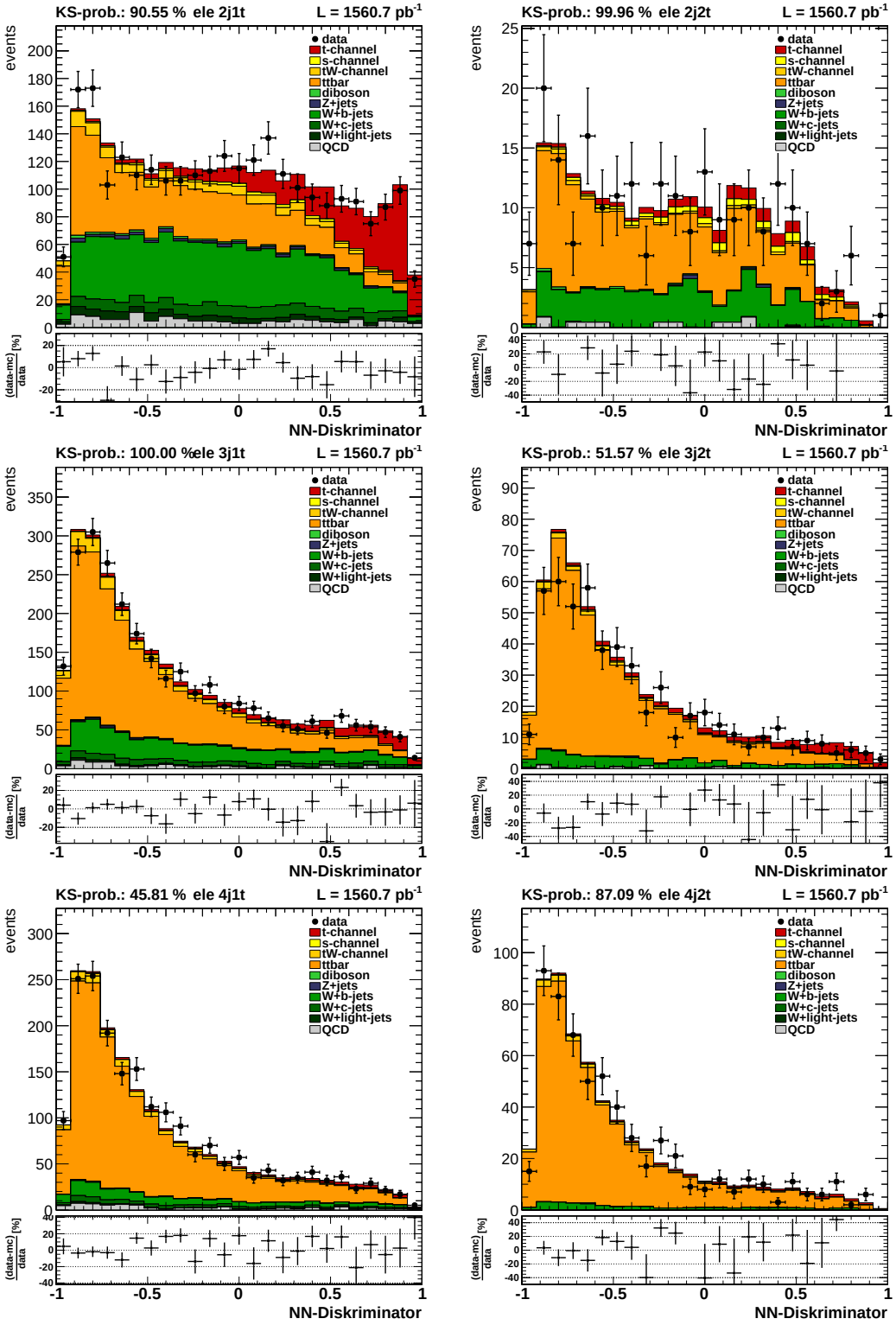


Figure B.6.: NeuroBayes discriminator distributions in the electron channel after the simulatenous fit.

List of Figures

List of Tables

Bibliography

[1] G. Aad et al. Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV. *Eur.Phys.J.*, C71:1577, 2011.

[2] G. Aad et al. Measurement of the t-channel single top-quark production cross section in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector. 2012.

[3] T. Aaltonen et al. First Observation of Electroweak Single Top Quark Production. *Phys.Rev.Lett.*, 103:092002, 2009.

[4] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 0803:S08002, 2008. doi: 10.1088/1748-0221/3/08/S08002,10.1088/1748-0221/3/08/S08002.

[5] F.D. Aaron et al. Combined Measurement and QCD Analysis of the Inclusive ep Scattering Cross Sections at HERA. *JHEP*, 01:109, 2010. doi: 10.1007/JHEP01(2010)109.

[6] S. Abachi et al. Observation of the Top Quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995. doi: 10.1103/PhysRevLett.74.2632. URL http://link.aps.org/doi/10.1103/PhysRevLett.74.2632.

[7] V. M. Abazov and other. Measurements of single top quark production cross sections and $|V_{tb}|$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. D*, 84:112001, Dec 2011. doi: 10.1103/PhysRevD.84.112001. URL http://link.aps.org/doi/10.1103/PhysRevD.84.112001.

[8] V. M. Abazov and other. Determination of the width of the top quark. *Phys. Rev. Lett.*, 106:022001, Jan 2011. doi: 10.1103/PhysRevLett.106.022001. URL http://link.aps.org/doi/10.1103/PhysRevLett.106.022001.

[9] V.M. Abazov et al. Observation of Single Top Quark Production. *Phys.Rev.Lett.*, 103:092001, 2009.

[10] F. Abe et al. Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.*, 74:2626–2631, 1995. doi: 10.1103/PhysRevLett.74.2626. URL http://link.aps.org/doi/10.1103/PhysRevLett.74.2626.

[11] W. Adam, R. Fruehwirth, A. Strandlie, and T. Todor. Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. Technical Report CMS-NOTE-2005-001. CERN-CMS-NOTE-2005-001, CERN, Geneva, 2005.

[12] W. Adam, B. Mangano, Th. Speer, and T. Todorov. Track Reconstruction in the CMS tracker. Technical Report CMS-NOTE-2006-041. CERN-CMS-NOTE-2006-041, CERN, Geneva, 2006.

[13] S. Agostinelli et al. GEANT4: A simulation toolkit. *Nucl. Instrum. Meth.*, A506: 250–303, 2003. doi: 10.1016/S0168-9002(03)01368-8.

[14] Z. Albrecht, T. Allmendinger, G. Barker, M. Feindt, C. Haag, et al. BSAURUS: A Package for inclusive B reconstruction in DELPHI. 2001. URL http://arxiv.org/abs/hep-ex/0102001.

[15] S. Alioli, P. Nason, C. Oleari, and E. Re. NLO single-top production matched with shower in POWHEG: s - and t -channel contributions. *Journal of High Energy Physics*, 2009(09):111, 2009. URL http://stacks.iop.org/1126-6708/2009/i=09/a=111.

[16] S. Alioli, P. Nason, C. Oleari, and E. Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 1006:043, 2010. doi: 10.1007/JHEP06(2010)043.

[17] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nuclear Physics B*, 126(2):298 – 318, 1977. ISSN 0550-3213. doi: 10.1016/0550-3213(77)90384-4. URL http://www.sciencedirect.com/science/article/pii/0550321377903844.

[18] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, A. Buckley, J. M. Butterworth, L. Dudko, S. Frixione, L. Garren, S. Gieseke, A. Gusev, I. Hinchliffe, J. Huston, B. Kersevan, F. Krauss, N. Lavesson, L. Lonnblad, E. Maina, F. Maltoni, M. L. Mangano, F. Moortgat, S. Mrenna, C. G. Papadopoulos, R. Pittau, P. Richardson, M. H. Seymour, A. Sherstnev, T. Sjostrand, P. Skands, S. R. Slabospitsky, Z. Wcas, B. R. Webber, M. Worek, and D. Zeppenfeld. A standard format for Les Houches Event Files, 2006. URL doi:10.1016/j.cpc.2006.11.010.

[19] D. Barge et al. Study of photon conversion rejection at CMS. Technical report, 2009. CMS AN-2009/159.

[20] R. Barlow. Fitting using finite Monte Carlo samples. *Computer Physics Communications*, 77(2):219–228, 1993. URL http://linkinghub.elsevier.com/retrieve/pii/001046559390005W.

[21] R. Barlow. Systematic errors: Facts and fictions. pages 134–144, 2002.

[22] J. Bauer. *Prospects for the Observation of Electroweak Top-Quark Production with the CMS Experiment*. PhD thesis, Karlsruher Institut für Technologie (KIT), 2010.

[23] E. Boos et al. CompHEP 4.4: Automatic computations from Lagrangians to events. *Nucl.Instrum.Meth.*, A534:250–259, 2004. doi: 10.1016/j.nima.2004.07.096.

[24] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Hoche, et al. General-purpose event generators for LHC physics. *Phys.Rept.*, 504:145–233, 2011. doi: 10.1016/j.physrep.2011.03.005.

[25] H. Byrd, P. Lu, and J. Nocedal. A limited memory algorithm for bound constrained optimization. Technical report, 1995.

[26] M. Cacciari, G.P. Salam, and G. Soyez. The anti- k t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, 2008. URL http://stacks.iop.org/1126-6708/2008/i=04/a=063.

[27] R. Campbell, J.M.and Frederix, F. Maltoni, and F. Tramontano. Nlo predictions for t-channel production of single top and fourth generation quarks at hadron colliders. *Journal of High Energy Physics*, 2009(10):042, 2009. URL http://stacks.iop.org/1126-6708/2009/i=10/a=042.

[28] J. Caron. Overall view of LHC experiments.. Vue d'ensemble des experiences du LHC. AC Collection. Legacy of AC. Pictures from 1992 to 2002., 1998.

[29] S. Catani, Y. Dokshitzer, M. Seymour, and B.R. Webber. Longitudinally-invariant $k_\perp$-clustering algorithms for hadron-hadron collisions. *Nucl. Phys. B*, 406(CERN-TH-6775-93. LU-TP-93-2):187–224, 1993.

[30] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber. QCD Matrix Elements + Parton Showers. *JHEP*, 0111:063, 2001. URL doi:10.1088/1126-6708/2001/11/063.

[31] CDF/PUB/TOP/PUBLIC/10793. *Measurement of Single Top Quark Production in 7.5 fb^{-1} of CDF Data Using Neural Networks*, 2012. The CDF Collaboration. URL http://www-cdf.fnal.gov/physics/new/top/2012/SingleTopPublic/index.html.

[32] *The ROOT project*. CERN. URL http://root.cern.ch.

[33] CERN. LHC machine outreach. Website, 2008. URL http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/.

[34] *oai:cds.cern.ch:1347440. LHC Lumi Days: LHC Workshop on LHC Luminosity Calibration*, Geneva, 2011. CERN, CERN.

[35] S. Chatrchyan et al. Measurement of the t-channel single top quark production cross section in pp collisions at sqrt(s) = 7 TeV. *Phys.Rev.Lett.*, 107:091802, 2011.

[36] Chrislb. Diagram of an artificial neuron. Wikipedia, 2005.

[37] Th. Chwalek. Measurement of the W-Boson Helicity-Fractions in Top-Quark Decays with the CDF II Experiment and Prospects for an Early $t\bar{t}$ Cross-Section Measurement with the CMS Experiment. *PhD Thesis, Karlsruhe Institute of Technology,*, IEKP-KA/2010-5, 2010.

[38] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003, 2008. URL http://stacks.iop.org/1748-0221/3/i=08/a=S08003.

[39] The CMS collaboration. *CMS Physics, Technical Design Report Detector Performance and Software*, volume 1. 2006. URL http://cmsdoc.cern.ch/cms/cpt/tdr/.

[40] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, 2008. URL http://stacks.iop.org/1748-0221/3/i=08/a=S08004.

[41] The CMS Collaboration. Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. (CMS-PAS-PFT-09-001), 2009.

[42] The CMS Collaboration. Algorithms for b Jet identification in CMS. (CMS-PAS-BTV-09-001), 2009.

[43] The CMS Collaboration. Commissioning of the particle-flow event reconstruction with the first lhc collisions recorded in the cms detector. (CMS-PAS-PFT-10-001), 2010.

[44] The CMS Collaboration. Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7 tev. (CMS-PAS-PFT-09-002), 2010.

[45] The CMS Collaboration. Particle-flow commissioning with muons and electrons from j/psi and w events at 7 tev. (CMS-PAS-PFT-09-003), 2010.

[46] The CMS Collaboration. Tracking and primary vertex results in first 7 tev collisions. 2010.

[47] The CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *Journal of Instrumentation*, 6:11002, 2011. doi: 10.1088/1748-0221/6/11/P11002.

[48] The CMS Collaboration. Performance of the b-jet identification in CMS. (CMS-PAS-BTV-11-001), 2011. URL http://cdsweb.cern.ch/record/1366061.

[49] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *Journal of Instrumentation*, 6(11):P11002, 2011. URL http://stacks.iop.org/1748-0221/6/i=11/a=P11002.

[50] The CMS Collaboration. b-jet identification in the cms experiment. (CMS-PAS-BTV-11-004), 2012. URL http://cdsweb.cern.ch/record/1427247.

[51] The CMS collaboration. Absolute calibration of the luminosity measurement at cms: Winter 2012 update. 2012.

[52] The CMS Collaboration and A. Hunt. Measurement of cms luminosity, 2010. URL http://indico.cern.ch/getFile.py/access?contribId=21&sessionId=2&resId=0&materialId=slides&confId=94815.

[53] The LHCb Collaboration. The LHCb Detector at the LHC. *Journal of Instrumentation*, 3(08):S08005, 2008. URL http://stacks.iop.org/1748-0221/3/i=08/a=S08005.

[54] The LHCf Collaboration. The LHCf detector at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08006, 2008. URL http://stacks.iop.org/1748-0221/3/i=08/a=S08006.

[55] The TOTEM Collaboration. The TOTEM Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08007, 2008. URL http://stacks.iop.org/1748-0221/3/i=08/a=S08007.

[56] M.A. Dobbs, S. Frixione, Eric Laenen, K. Tollefson, H. Baer, et al. Les Houches guidebook to Monte Carlo generators for hadron collider physics. pages 411–459, 2004. Compiled by the Working Group on Quantum Chromodynamics and the Standard Model.

[57] Y. Dokshitzer, G.D. Leder, S. Moretti, and B.R. Webber. Better jet clustering algorithms. *JHEP*, 9708:001, 1997.

[58] Y.L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov.Phys.JETP*, 46:641–653, 1977.

[59] The CMS electron γ physics objects group. Electron identification based on simple cuts. URL https://twiki.cern.ch/twiki/bin/view/CMS/SimpleCutBasedEleID.

[60] M. Feindt. A Neural Bayesian Estimator for Conditional Probability Densities. *ArXiv Physics e-prints*, 2004. URL http://arxiv.org/abs/physics/0402093.

[61] M. Feindt and U. Kerzel. The neurobayes neural network package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 559(1):190 – 194, 2006. ISSN 0168-9002. doi: 10.1016/j.nima.2005.11.166. URL http://www.sciencedirect.com/science/article/pii/S0168900205022679.

[62] R. P. Feynman. Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction. *Phys. Rev.*, 80:440–457, 1950. doi: 10.1103/PhysRev.80.440. URL http://link.aps.org/doi/10.1103/PhysRev.80.440.

[63] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. Advantages of the Color Octet Gluon Picture. *Phys.Lett.*, B47:365–368, 1973. doi: 10.1016/0370-2693(73)90625-4. URL http://inspirehep.net/record/83736/export/hx.

[64] S. Frixione, P. Nason, and C. Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007. doi: 10.1088/1126-6708/2007/11/070. URL http://www.slac.stanford.edu/spires/find/hep/www?key=7337256&FORMAT=WWWBRIEFBIBTEX.

[65] R. Fruehwirth, W. Waltenberger, and P. Vanlaer. Adaptive vertex fitting. Technical Report CMS-NOTE-2007-008, CERN, Geneva, Mar 2007.

[66] M. Gell-Mann. A schematic model of baryons and mesons. *Physics Letters*, 8 (3):214 – 215, 1964. ISSN 0031-9163. doi: 10.1016/S0031-9163(64)92001-3. URL http://www.sciencedirect.com/science/article/pii/S0031916364920013.

[67] D. Green. "rediscovering" the Standard Model at Cms. *Modern Physics Letters A*, 26:309–317, 2011. doi: 10.1142/S0217732311035134.

[68] V.N. Gribov and L.N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov.J.Nucl.Phys.*, 15:438–450, 1972.

[69] The Particle Data Group. Review of Particle Physics, 2010-2011. Review of Particle Properties. *J. Phys. G*, 37(7A):075021, 2010.

[70] St. Hoeche, F. Krauss, N. Lavesson, L. Lonnblad, M. Mangano, et al. Matching parton showers and matrix elements. 2006.

[71] A. Iorio. Measurement of single top production in pp collisions at 7 TeV with the CMS detector. Technical Report CMS-CR-2011-277, CERN, Geneva, 2011.

[72] CMS jet energy correction group. Jet energy scale uncertainty sources. URL https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources.

[73] V. Khachatryan et al. First Measurement of the Cross Section for Top-Quark Pair Production in Proton-Proton Collisions at sqrt(s)=7 TeV. *Phys.Lett.*, B695: 424–443, 2011.

[74] N. Kidonakis. Top quark pair and single top production at Tevatron and LHC energies. PoS ICHEP2010:059,2010, 2010. URL http://arxiv.org/abs/1008.2460.

[75] N. Kidonakis. Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production. *Phys. Rev. D*, 83:091503, May 2011. doi: 10.1103/PhysRevD.83.091503. URL http://link.aps.org/doi/10.1103/PhysRevD.83.091503.

[76] N. Kidonakis. Top Quark Theoretical Cross Sections and pT and Rapidity Distributions, 2011. URL http://arxiv.org/abs/1109.3231.

[77] M. Kobayashi and T. Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog.Theor.Phys.*, 49:652–657, 1973. doi: 10.1143/PTP.49.652.

[78] D. Konstantinov, J.R. Komaragiri, and Th. Peiffer. SingleTopTurnOnCurves, 2012. URL https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopTurnOnCurves.

[79] J.H. Kühn, A. Scharf, and P. Uwer. Weak interaction effects in top-quark pair production at hadron colliders. *The European Physical Journal C - Particles and Fields*, 51:37–53, 2007. ISSN 1434-6044. URL http://dx.doi.org/10.1140/epjc/s10052-007-0275-x. 10.1140/epjc/s10052-007-0275-x.

[80] *The minuit package*. LCG, 2006. URL http://seal.web.cern.ch/seal/work-packages/mathlibs/minuit/.

[81] G. Mahlon and S. Parke. Single Top Quark Production at the LHC: Understanding Spin. Phys.Lett. B476 (2000) 323-330, 1999.

[82] F. Maltoni and T. Stelzer. MadEvent: Automatic event generation with MadGraph. *JHEP*, 0302:027, 2003.

[83] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur.Phys.J.*, C63:189–285, 2009. doi: 10.1140/epjc/s10052-009-1072-5.

[84] D. Martschei. Developement of a soft electron based b-jet tagger for the CMS experiment. Master's thesis, 2009. URL http://www-ekp.physik.uni-karlsruhe.de/pub/web/thesis/iekp-ka2009-06.pdf.

[85] E. Migliore. CMS Tracker alignment and material budget measurement. Technical Report CMS-CR-2011-164, CERN, Geneva, 2011.

[86] Th. Müller, J. Ott, and J. Wagner-Kuhr. `theta` website. URL http://www.theta-framework.org.

[87] Th. Müller, J. Ott, and J. Wagner-Kuhr. *theta a framework for template-based modeling and inference*. Instiut für Experimentelle Kernphysik, 2010. URL http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/files/theta.pdf.

[88] The CMS muon physics objects group. CMS twiki, 2011. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId#Basline_muon_selections_for_2011.

[89] P. Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 0411:040, 2004. doi: 10.1088/1126-6708/2004/11/040.

[90] J. Ott. Proposal for treatment of systematics in t-channel analyses and discussion with btv and jetmet pog contacts, 2012. URL https://indico.cern.ch/contributionDisplay.py?contribId=1&confId=178372.

[91] J. Ott. In situ constraining of systematic uncertainties. private communications, May 2012.

[92] PDG. *Nuclear and Particle Physics*, chapter 11. THE CKM QUARK-MIXING MATRIX. Particle Data Group, 2011. URL http://pdg.lbl.gov/2011/reviews/rpp2011-rev-ckm-matrix.pdf.

[93] Th. Peiffer. Q2 systematic MC samples re-weighting, 2012. URL http://indico.cern.ch/getFile.py/access?contribId=5&resId=0&materialId=slides&confId=177717.

[94] A. Pukhov, E. Boos, M. Dubinin, V. Edneral, V. Ilyin, et al. CompHEP: A Package for evaluation of Feynman diagrams and integration over multiparticle phase space. 1999.

[95] E. Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur.Phys.J.*, C71:1547, 2011. doi: 10.1140/epjc/s10052-011-1547-z.

[96] S. Röcker. Prospects for the Observation of Single Top Quark Production with the CMS Experiment using Neural Networks. Master's thesis, Karlsruher Institut für Technologie (KIT), 2011.

[97] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Parallel distributed processing: explorations in the microstructure of cognition, vol. 1. chapter Learning internal representations by error propagation, pages 318–362. MIT Press, Cambridge, MA, USA, 1986. ISBN 0-262-68053-X. URL http://dl.acm.org/citation.cfm?id=104279.104293.

[98] G.P. Salam and G. Soyez. A practical seedless infrared-safe cone jet algorithm. *Journal of High Energy Physics*, 2007(05):086, 2007. URL http://stacks.iop.org/1126-6708/2007/i=05/a=086.

[99] R. Schwienhorst. Single top-quark production with the ATLAS detector in pp collisions at sqrt(s)=7TeV. 2011.

[100] The CMS single top group. Systematics for the single top analysis 2011 (specific to t-channel), 2012. URL https://twiki.cern.ch/twiki/bin/view/CMS/SingleTopTChanSyst.

[101] T. Stelzer and W.F. Long. Automatic generation of tree level helicity amplitudes. *Comput.Phys.Commun.*, 81:357–371, 1994. doi: 10.1016/0010-4655(94)90084-1.

[102] M. Strovink. Direct Measurement of the Top Quark Mass, 1997. URL http://www-d0.fnal.gov/public/pubs/topmass_prl_sum.html.

[103] CDF Tevatron Electroweak Working Group and D0 Collaborations. Combination of CDF and D0 results on the mass of the top quark using up to 5.8 fb-1 of data. 2011.

[104] S. van der Meer. Calibration of the effective beam height in the isr. oai:cds.cern.ch:296752. Technical Report CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, Geneva, 1968.

[105] B. Walsh. Markov Chain Monte Carlo and Gibbs sampling. Lecture notes for EEB 581, 2004. URL http://web.mit.edu/~wingated/www/introductions/mcmc-gibbs-intro.pdf.

[106] M. Whalley and A. Buckley. LHAPDF the Les Houches Accord PDF Interface, 2012. URL http://lhapdf.hepforge.org/.

[107] M.R. Whalley, D. Bourilkov, and R.C. Group. The Les Houches accord PDFs (LHAPDF) and LHAGLUE. 2005.

[108] M. Wobisch and T. Wengler. Hadronization corrections to jet cross-sections in deep inelastic scattering. 1998.

Danksagung

An dieser Stelle möchte ich all jenen Personen danken, die direkt oder indirekt zum Gelingen dieser Arbeit beigetragen haben.

Daher gilt mein Dank vor allem meinem Doktorvater Prof. Michael Feindt der mir diese Arbeit ermöglicht hat. Ich bin sehr dankbar für die Freiheiten und Möglichkeiten die ich während meiner Promotion hatte und für alles was ich in der Zeit gelernt habe. Außerdem danke ich ihm für die einmalige Möglichkeit neben meiner Promotion in der Firma <phi-t> zu Arbeiten und dadurch eine sehr abwechslungsreiche Zeit zu erleben und Industrieerfahrung zu sammeln.

Ebenso danke ich Prof. Thomas Müller für die Übernahme des Korefferats und die Möglichkeit diese Arbeit als Kooperation zwischen zwei Arbeitssgruppen zu beenden. Erst dadurch hatte ich Zugang zu einer breiten (Wissens-)Basis innerhalb der CMS-Kollaboration und habe die nötige Unterstützung erfahren.

Ein besonderer Dank gilt daher auch Dr. Jeannine Wagner-Kuhr, die während der letzten Jahre meiner Arbeit meine direkte Betreuerin war und diese entscheidend geprägt und voran getrieben hat. Durch ihre hohen Anforderugen habe ich in kurzer Zeit sehr viel geschafft und gelernt, was ohne den nötigen Antrieb sicher länger gedauert hätte.

Ebenso danke ich meinen Kollegen Steffen Röcker, Jens Hansen und Jochen Ott für die gute Zusammenarbeit und Hauke Held für die forschungsübergreifende Kooperation auch abseits der Physik. Die freundschaftliche Arbeit mit Ihnen hat auch die stressigen Zeiten noch sehr angenehm gemacht und den Spaß an der Arbeit immer wieder zurück gebracht.

Mein Dank gilt auch allen anderen Kollegen des Instituts für Experimentelle Kernphysik für die tolle Arbeitsatmosphäre und Ihre Unterstützung speziell während der letzten Monate meiner Arbeit. Auch bin ich sehr froh, dass am EKP die Computerinfrastruktur durch die Doktoranden selbst adminsistriert wird. Dadurch hat man die Möglichkeit vieles über den Betrieb eines Computerclusters zu lernen und auszuprobieren.

Abschließend möchte ich mich noch bei meiner Familie für die jahrelange Unterstützung bedanken. Daneben danke ich meinen Freunden und Bekannten, besonders meinen ehemaligen Kommilitonen, für die seelische und moralische Unterstützung und auch so manchen Motivationsschub bedanken. Ohne diesen Rückhalt wäre mir diese Arbeit wohl wesentlich schwerer gefallen. Dafür sei allen herzlichst gedankt.