



BERGISCHE UNIVERSITÄT WUPPERTAL
FACHBEREICH C
Mathematik und Naturwissenschaften
FACHGRUPPE PHYSIK

**Search for the decay of a top quark superpartner
into a top quark and a neutralino using multivariate
techniques at the ATLAS experiment**

Master Thesis
DANIELA BÖRNER

Abstract

This master thesis describes a search for Supersymmetry at the ATLAS experiment at the LHC. For this search it is necessary to know the differences between the analysed decay of the top quark superpartner (stop quark) into a top quark and a neutralino and the corresponding backgrounds. Therefore distributions with different behaviours for these decays are needed. Separating variables are for example the missing transverse energy, the transverse mass, the perpendicular component of the missing transverse energy to the leptonically decaying top quark and some angular distributions between the decay products of the top quarks, the missing transverse energy, or the leading jets. Some of these variables are also used in the official ATLAS search.

These variables are used in a neural network which is trained to differentiate between the stop decay and the two main backgrounds – $t\bar{t}$ and $W+jets$. This new method was optimised in order to reach a better signal to background ratio than the simple cut analysis. The outcome distribution does not show any significant excess in the comparison of data and the Standard Model prediction. Therefore the distribution is used to set limits on the stop and neutralino masses. The limits using the neural network are improved compared to those using simple cuts especially for the bulk and the diagonal region.

Contents

1. Introduction	1
2. Theory of particle physics and the LHC	3
2.1. The Standard Model	3
2.1.1. Elementary particles	3
2.1.2. Interactions	3
2.1.3. Open questions of the Standard Model	4
2.2. Supersymmetry	6
2.3. The Large Hadron Collider	8
2.4. The ATLAS experiment	9
3. Motivation	13
4. General Information	17
4.1. Simulated data samples	17
4.2. Triggers	17
4.3. Object definition and selection	18
4.4. General event preselection	19
5. Distinction of signal and background	21
5.1. Variables used in the official ATLAS analysis	21
5.2. Additional variables	28
5.3. Conclusion	33
6. Evaluation and optimisation of multivariate techniques	35
6.1. Multivariate methods	35
6.1.1. Multilayer Perceptron	35
6.1.2. Boosted Decision Tree	37
6.2. Optimising the neural networks	39
6.3. Results for the three regions	44
6.3.1. Bulk region	44
6.3.2. Diagonal region	48
6.3.3. High mass region	54
6.4. Conclusion	57
7. Validation of the simulation	59
7.1. First comparison of data and simulation	59
7.2. Determination of correct normalisation factors	60
7.3. Resulting comparison of data and simulation	63
7.4. Shape dependence of the neural network	65
7.5. Resulting neural network output distributions	67
8. Limits on simplified models	69
8.1. Theoretical procedure	69

8.2. Limit setting with pseudo data	70
8.3. Limit setting using data	78
8.4. Comparison to official results	80
8.5. Conclusion	82
9. Preparation for run II and reoptimisation	83
9.1. Variable distributions	83
9.2. Optimisation	85
9.3. Optimisation with a different preselection	88
9.4. Reoptimisation for the high mass region using $\sqrt{s} = 8 \text{ TeV}$	90
9.5. Results and conclusion	91
10. Conclusion and outlook	93
References	95
A. Appendix	99
A.1. Simulated data samples	99
A.2. Cuts from the official analysis	101
A.3. Distributions with and without an ISR jet	101
A.4. Optimisation results for the muon channel	108
A.5. Neural network distributions for all backgrounds	114
A.6. Determination of normalisation factors	115
A.7. Validation of the simulation	118
A.8. Limits on simplified models	125
A.9. Preparation for run II and reoptimisation of the high mass region	127
Erklärung	133

1. Introduction

In the Standard Model of particle physics the basic interactions of fundamental particles are described. The LHC, so far the largest and most powerful particle accelerator, is installed at CERN. There are four big experiments – ATLAS, CMS, ALICE and LHCb – which are installed at CERN at the interaction regions of the LHC in order to analyse the particle collisions. Using these detectors, the Standard Model is investigated and new physics beyond the Standard Model is searched for.

One possible extension of the Standard Model is Supersymmetry, the symmetry of fermions and bosons. In this theory a supersymmetric partner is introduced for every Standard Model particle. For fermions this supersymmetric partner is a boson, for bosons this partner is a fermion. Therefore the supersymmetric partner of the top quark, which is a fermion, is the bosonic stop quark (\tilde{t}). The decay of this stop quark is analysed in this master thesis. This analysis assumes, among other things, that the lightest supersymmetric squark is the stop quark. Therefore it does not decay into other squarks. The neutralino is here assumed to be the lightest supersymmetric particle, which should be stable.

In the simplified model considered in this analysis, the stop quarks are only produced in pairs. Therefore the $t\bar{t}^*$ pair decays into two top quarks and two neutralinos. The neutralinos are only measured as missing transverse energy, as they only interact weakly. The Standard Model $t\bar{t}$ background is the dominant background for the $t\tilde{t}^*$ decay. The aim of this thesis is not only the separation between the $t\tilde{t}^*$ and the $t\bar{t}$ decay but also the separation between the $t\tilde{t}^*$ decay and all other backgrounds. For this, a neural network is trained which separates the $t\bar{t}$ and the $W+jets$ background from the signal. These are the two dominant backgrounds. The neural network is evaluated for all other backgrounds after the training. The input variables for the neural network are checked for the agreement between measured data and the Standard Model predictions.

The output of the neural network is used to set limits on the different stop and neutralino masses. In addition, the performance of this method is compared to a simple cut based analysis.

In chapter 2 the Standard Model of particle physics is described. Afterwards the supersymmetric theory – especially the MSSM – is illustrated. In the end of that chapter the LHC and the ATLAS detector are shortly explained.

The motivation of this master thesis is explained in chapter 3.

Chapter 4 gives some general information about the used simulated data sets, the triggers and the general object definition and selection. Last, the general preselection for the following analysis is shown.

In chapter 5 the variables which separate signal and background are described. These variables are then used for the MVA training and optimisation, which is described in chapter 6. The agreement of measured data and the Standard Model prediction is analysed in chapter 7.

The limit setting procedure is described in chapter 8. The results are shown with and without using systematics. In addition, the resulting limits are compared to a simple cut method and the official ATLAS limits.

Chapter 9 shows the first preparation for Run II. In this chapter a neural network was optimised for samples with high stop masses. The events produced in these samples have a center of mass energy of 14 TeV.

In the end, chapter 10 gives a general conclusion and a short outlook on further improvements of this analysis and an application to the LHC Run II.

2. Theory of particle physics and the LHC

The Standard Model of particle physics is the best description of all known particles and interactions so far. There are still open questions and problems in the framework of the Standard Model. One possible solution to some problems is the theory of Supersymmetry. Both – the Standard Model and Supersymmetry – are described in the following.

The Standard Model and searches for new physics like Supersymmetry are analysed with high energy collisions. For this, a strong accelerator and powerful detectors are needed. In the last part of this chapter the LHC and the ATLAS detector which are both used for such analysis are described.

2.1. The Standard Model

The Standard Model (SM) describes all known elementary particles, the electromagnetic, the weak and the strong interaction. The gravitation is not included into the Standard Model, it can be neglected compared to the other forces [1, 2].

2.1.1. Elementary particles

Elementary particles are classified by their different properties as illustrated in figure 2.1. Particles with half-integer spin are called fermions and follow the Fermi Dirac statistic. Particles with integer spin are called bosons and follow the Bose Einstein statistic. The gauge bosons are the particles which transmit the different interactions. The Higgs boson gives masses to all particles.

For the fermions, one distinguishes between quarks and leptons. On the left side of figure 2.1 the six quarks – up, down, charm, strange, top and bottom – and the six leptons – electron, electron-neutrino, muon, muon-neutrino, tau and tau-neutrino – are shown. The fermions are categorised into three families. Therefore the fermions are depicted in three columns, where each column represents one family. The differences between a particle and its corresponding particle from a higher family is the higher mass. Therefore unstable particles from the second and the third family decay into particles from the first family. For every shown particle there exist also an antiparticle with exactly the same properties, but only reversed charge.

In the following, leptons are defined as only charged leptons which can be measured in the detector, i.e. e^\pm or μ^\pm .

2.1.2. Interactions

All known bosons are shown on the right side of figure 2.1. One of these bosons is the photon, which transmits the electromagnetic force. The photon has no charge and the range of this interaction is infinite because the photon is massless.

The W^\pm and the Z boson are the bosons of the weak interaction and they couple to the weak isospin. W^+ and W^- also transport electromagnetic charge. The weak force is the only force that violates the combined charge and parity conjugation (\mathcal{CP}) symmetry. This symmetry describes the exchange of particles and antiparticles. Due to the high mass of the three bosons the interaction is short-ranged.

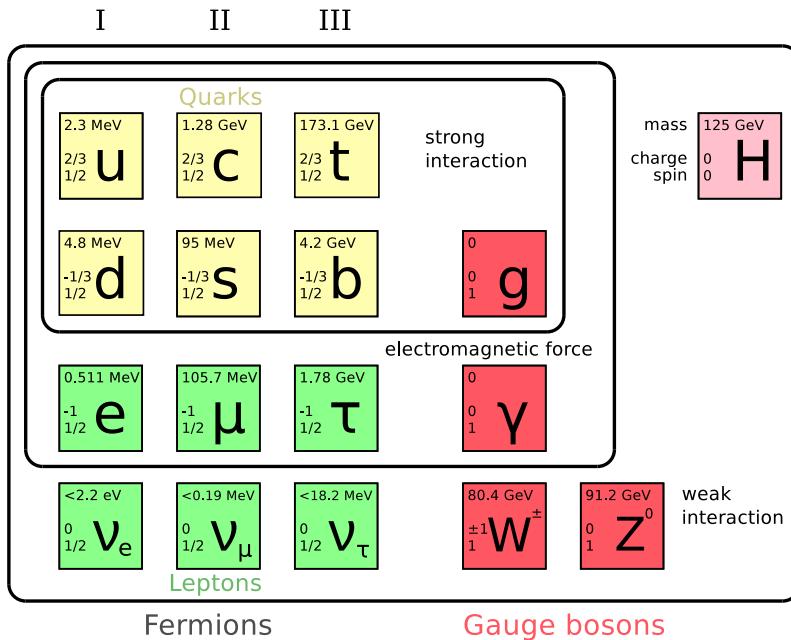


Figure 2.1: The known particles and interactions of the Standard Model. [3]

The strong interaction couples to the colour charge – red, blue or green – and is transmitted by gluons. Each gluon has two colour charges and therefore couples to itself. The effective force grows stronger with longer distance. When separating two particles with colour charge, new coloured particles are produced in order to neutralise the colour charges due to energetic reasons. Therefore it is not possible to see free colour charges. Each quark has exactly one colour charge. Therefore single quarks or gluons are not visible, but they form colour neutral combinations which are called hadrons. A combination of one quark and an antiquark is called meson, a combination of three quarks or three antiquarks is called baryon.

The Higgs boson was discovered in 2012 at the LHC [4]. This particle was motivated because the correct mathematical description of the Standard Model predicts only massless particles. But due to the masses of the W^\pm and Z bosons, the local gauge invariance is violated. These masses are measured and the Higgs mechanism can be used to solve the problem. The Higgs boson is an excitation of the Higgs field to which the fundamental particles couple. In addition, there are three free parameters which give the masses to the W^\pm and the Z boson. The interaction of a fermion with this field is interpreted as the mass of the fermions. The masses of the fermions differ due to different couplings to this field.

2.1.3. Open questions of the Standard Model

As already said, the Standard Model is a good description for all known particles and the interactions between these particles. On the other hand there are some unsolved issues which can be solved with an extension of the Standard Model.

The most known problem of the Standard Model is the so-called hierarchy problem. The mass of the Higgs boson is calculated at tree level. For higher precision, radiative correc-

tions through loop diagrams need to be applied to the tree level value. These corrections depend quadratically on the momenta of the virtual particles. As the upper limit for these corrections a theory cutoff Λ is chosen. If the Standard Model is valid to the Planck scale, Λ would be 10^{19} GeV and the radiative correction is in the order $m_H \propto \Lambda^2$. In order to obtain the measured mass as the remaining Higgs mass, the Higgs bare mass and the radiative corrections have to cancel out each other to about 40 digits. This fine tuning mechanism seems to be quite unnatural.

This problem can be solved in an elegant way by using Supersymmetry. The correction from a fermion (boson) has positive (negative) sign (compare figure 2.2). Using Supersymmetry, an additional symmetry between fermions and bosons is introduced (which is explained in detail in the following chapter), therefore the contributions cancel out.

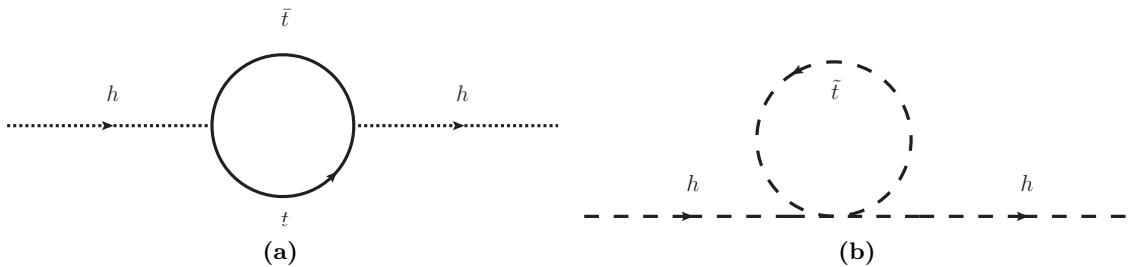


Figure 2.2: Loop corrections for the Higgs mass term. The contribution from a fermion (a) has a positive sign, the contribution from a boson (b) has negative sign and therefore cancel each other.

The Standard Model as described above only describes the ordinary matter. Due to cosmological measurements it is known that the ordinary matter makes up only 4.9% of the total mass of the Universe. The remaining part consists of 26.8% dark matter and 68.3% dark energy [5]. The dark matter does not interact electromagnetically, but it is visible through gravitational effects.

With the lightest supersymmetric particle (LSP), Supersymmetry provides a good candidate for dark matter. The LSP should be neutral and only weakly interacting, as it is stable but not yet observed.

Comparing the running of coupling constants – electromagnetic, weak, and strong interaction – in the Standard Model shows, that these three couplings do not meet at the same point. Using Supersymmetry with a grand unification of all the forces, the couplings meet at the same point for high energies.

In the Standard Model only left handed neutrinos are visible, this implies a mass equal to zero for neutrinos. Experimental observations have proved that this assumption is wrong.

As mentioned above, gravity is not included in the Standard Model. The strength of this force is comparable to the other forces at the Planck scale ($\Lambda \simeq 10^{19}$ GeV). Therefore the Standard Model is only valid to this cutoff.

2.2. Supersymmetry

Supersymmetry introduces an additional symmetry between fermions and bosons [2, 6, 7]. An operator Q , called supercharge, is defined as a generator for the symmetry transformation

$$Q|\text{boson}\rangle = |\text{fermion}\rangle, \quad (1)$$

$$Q|\text{fermion}\rangle = |\text{boson}\rangle. \quad (2)$$

This operator is a spinor operator which carries a spin angular momentum of $1/2$. Using this new symmetry for every Standard Model particle, a supersymmetric partner is defined. This partner differs by $\pm 1/2$ unit spin from the Standard Model particle, all other quantities are unchanged. For the naming scheme a $\tilde{\cdot}$ is superimposed on the Standard Model symbol, therefore \tilde{t} is the supersymmetric partner of a top quark. Scalar partners of fermions are called sfermions, or sleptons, and squarks. The left and the right handed Standard Model fermion each have their own supersymmetric partner. In addition, the superpartners of the Standard Model bosons obtain the additional suffix -ino.

The generators must satisfy commutation and anticommutation relations

$$\{Q_\alpha, Q_\beta^\dagger\} = 2 \cdot \sigma_{\alpha\beta}^\mu \cdot P^\mu, \quad (3)$$

$$\{Q_\alpha, Q_\beta\} = 0 = \{Q_\alpha^\dagger, Q_\beta^\dagger\}, \quad (4)$$

$$[P^\mu, Q_\alpha] = 0 = [P^\mu, Q_\alpha^\dagger], \quad (5)$$

with P^μ as four-momentum generator of spacetime translations under Lorentz boosts and rotations. The commutation relation between Q and P^μ implies

$$[P^2, Q_\alpha] = 0, \quad (6)$$

where P^2 corresponds to the mass of the particle. Therefore Standard Model particles and their corresponding partners have the same mass. The generators Q and Q^\dagger also commute with the generators of the gauge transformations and the momentum generator. Therefore Standard Model particles and their corresponding partners have the same weak isospin, electric and colour charge.

The commutation relations are invariant under chiral rotation. Therefore it is possible to introduce the so called R parity

$$R = (-1)^{3 \cdot (B-L) + 2 \cdot S}, \quad (7)$$

where B is the baryon number, L the lepton number and S the spin. In the Minimal Supersymmetric Standard Model (which will be introduced in paragraph 2.2), this additional R parity eliminates the B or L violating terms. Such terms would introduce a possibility for the proton to decay, for which stringent limits exist. For Standard Model particles the value is $R = +1$, for supersymmetric particles it is $R = -1$.

In models where this R parity is conserved, supersymmetric particles can only be produced in pairs. The lightest supersymmetric particle has to be stable, neutral and only weakly interacting. In this case it becomes a candidate for dark matter.

The mathematical description of Supersymmetry relies on so called superfields, i.e. functions of the extended spacetime (x^μ, θ^α). These superfields combine the Standard Model fields and their superpartners. The θ^α are fermionic spin coordinates. A series expansion in these variables always terminates, as they anticommute. Therefore the superfields can be written as

$$\begin{aligned}\Phi(x, \theta) = & \mathcal{S} - i\sqrt{2}\bar{\theta}\gamma_5\psi - \frac{i}{2}(\bar{\theta}\gamma_5\theta)\mathcal{M} + \frac{1}{2}(\bar{\theta}\theta)\mathcal{N} + \frac{1}{2}(\bar{\theta}\gamma_5\gamma_\mu\theta)V^\mu \\ & + i(\bar{\theta}\gamma_5\theta)\left[\bar{\theta}\left(\lambda + \frac{i}{\sqrt{2}}\not{\partial}\psi\right)\right] - \frac{1}{4}(\bar{\theta}\gamma_5\theta)^2\left[\mathcal{D} - \frac{1}{2}\square\mathcal{S}\right].\end{aligned}\quad (8)$$

The coefficients in this equation are the sixteen component fields, where V^μ is a vector field, ψ and λ are spinor fields and \mathcal{S} , \mathcal{M} , \mathcal{N} and \mathcal{D} are scalar fields.

Global supersymmetric transformations can then be written as translations in the super-space

$$\delta\Phi = \left(-\bar{\alpha}\frac{\partial}{\partial\bar{\theta}} - i\bar{\alpha}\not{\partial}\theta\right)\Phi. \quad (9)$$

General Lagrangian densities can be written as products of superfields.

Until now, no supersymmetric particles have been observed, although they should have the same mass as their Standard Model partners. Therefore this symmetry must be broken.

To break the symmetry, an additional term is added to the Lagrangian. This term is only a soft supersymmetry breaking term in order to have a chance to still solve the hierarchy problem. With a hard symmetry breaking term the masses of the particles can change in such a way that the cancellation for the Higgs corrections does not work anymore.

This leads to a new Lagrangian \mathcal{L}

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{Soft}} \quad (10)$$

where $\mathcal{L}_{\text{SUSY}}$ is the supersymmetric Lagrangian with all the kinematic terms and the superpotential which contains the needed couplings. $\mathcal{L}_{\text{Soft}}$ is the soft symmetry breaking term. It is not invariant under supersymmetric transformations, therefore it is not invariant for supersymmetric particles but for Standard Model particles. In addition it has no quadratically divergent terms.

Due to the breaking, there are remaining terms like $m_{\text{SUSY}}^2 - m_{\text{SM}}^2$. In order to preserve naturalness there is the assumption that the masses of the supersymmetric particles should be smaller than 1 TeV. Using this breaking many new free parameters are introduced.

Minimal Supersymmetric Standard Model The Minimal Supersymmetric Standard Model (MSSM) is the minimal extension of the Standard Model concerning the number of new particles. It has the same interactions as the Standard Model. In this model every Standard Model particle obtains its superpartner, and in addition there are more

superpartners for the Higgs boson. In a supersymmetric scenario, two complex doublets are needed for the Higgs field, which leads to five Higgs bosons. Therefore also five Higgsinos are introduced, the superpartners of the Higgs bosons. Two of them are electrically charged, the other three are neutral.

In the MSSM the charged gauginos and the charged Higgsinos mix to the two charged mass eigenstate $\tilde{\chi}_{1,2}^{\pm}$, called charginos. In addition, the mixing of the neutral gauginos and Higgsinos produces four neutral mass eigenstates $\tilde{\chi}_{1,2,3,4}^0$, which are called neutralinos.

2.3. The Large Hadron Collider

The largest research center for particle physics is CERN (Conseil Européen pour la Recherche Nucléaire) which is located near Geneva. There the Large Hadron Collider (LHC) – the most complex machine ever built – is operated. In the LHC two beams of protons are collided with a center of mass energy of $\sqrt{s} = 8 \text{ TeV}$. The data taking until 2013 gave an integrated luminosity of around 20 fb^{-1} . In addition, heavy ion collisions are possible.

The design center of mass energy is $\sqrt{s} = 14 \text{ TeV}$, which should be achieved after the first long shutdown during 2013 and 2014.

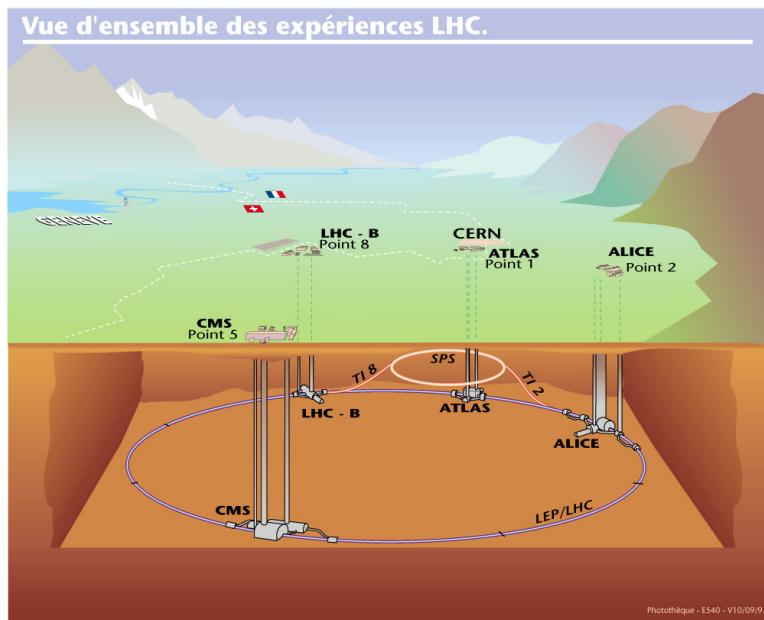


Figure 2.3: Schematic layout of the LHC accelerator ring. The two beams collide inside the four main detectors. [8]

A schematic layout of the LHC accelerator is shown in figure 2.3. In front of the LHC there are several pre-accelerators, which accelerate the protons to higher energies. The LHC receives protons with an initial energy of 450 GeV from the Super Proton Synchrotron (SPS).

As one can see in figure 2.3 the LHC is located 100 m underground. It is located in the old tunnel of the Large Electron Positron Collider (LEP) and has a circumference of 27 km.

There are two separate beam pipes in which two beams of protons are accelerated. In order to keep them on a circular track and focus the beam, dipole, quadrupole, sextupole, octupole and decapole magnets are used. They are cooled down to a temperature of $T = 1.9\text{ K}$ with superfluid helium in order to achieve superconductivity.

The protons are concentrated in 2808 bunches of about 10^{11} particles. At four points – for the four main detectors – they are focused to a diameter of $20\text{ }\mu\text{m}$ and a length of 8 cm and are brought to collision. This happens with a rate of up to 40 MHz. At these collision points detectors are installed. The four main experiments, which analyse the interaction after the collision, are ALICE, ATLAS, CMS and LHCb.

The LHCb experiment analyses collisions including bottom quarks and study the \mathcal{CP} -violation. The ALICE detector is specialised for heavy ion collisions. They analyse the produced quark-gluon plasma to study the conditions shortly after the big bang.

Both ATLAS and CMS are general purpose detectors. They are cylindric around their interaction point and designed to reconstruct all parts of a collision. This information is used to study the Standard Model further and also to search for new physics. The ATLAS detector is described in detail in the following chapter.

2.4. The ATLAS experiment

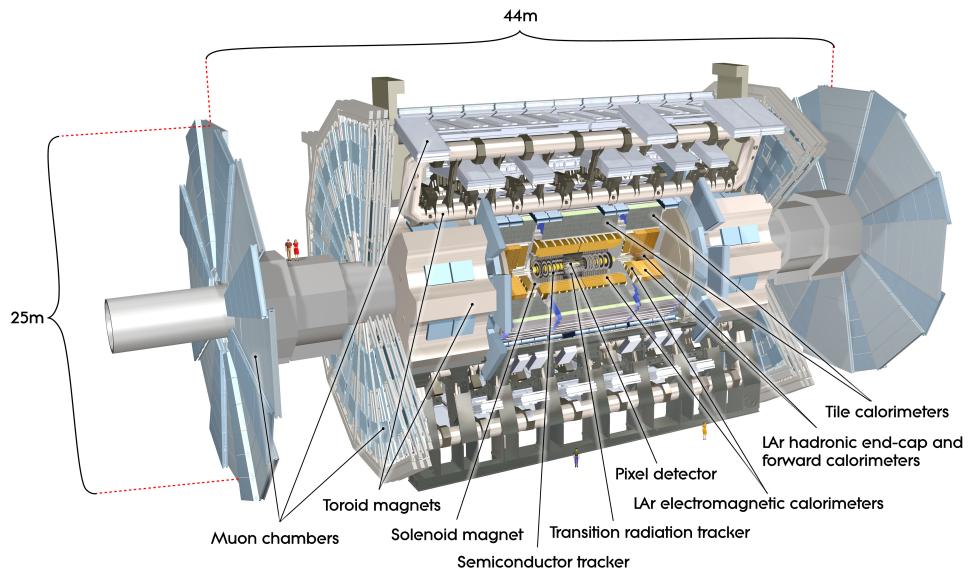


Figure 2.4: Schematic view of the ATLAS detector, which has an overall length of 44 m and a diameter of 25 m. [9]

The ATLAS detector is installed at the LHC ring at interaction point 1. As shown in figure 2.4 the detector is 44 m long and has a diameter of 25 m. It has a cylindric composition of different subdetectors around the beam pipe. Therefore it can detect particles over almost the complete solid angle around the collision point. The central – cylindric – part is called the barrel region. For both end regions there are additional end-caps.

All subdetectors are located cylindrically around the beam pipe. Therefore each particle goes through the same types of detectors and can be measured and classified with them.

Tracking system The innermost part of the ATLAS experiment is the tracking system with a pixel detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT) which are depicted in figure 2.5. These subdetectors are used to determine the track of charged particles. Using the curvature of the particles, caused by a magnetic field of $B = 2\text{ T}$ from the solenoid magnet, one can determine the momenta of charged particles, as well as their charge.

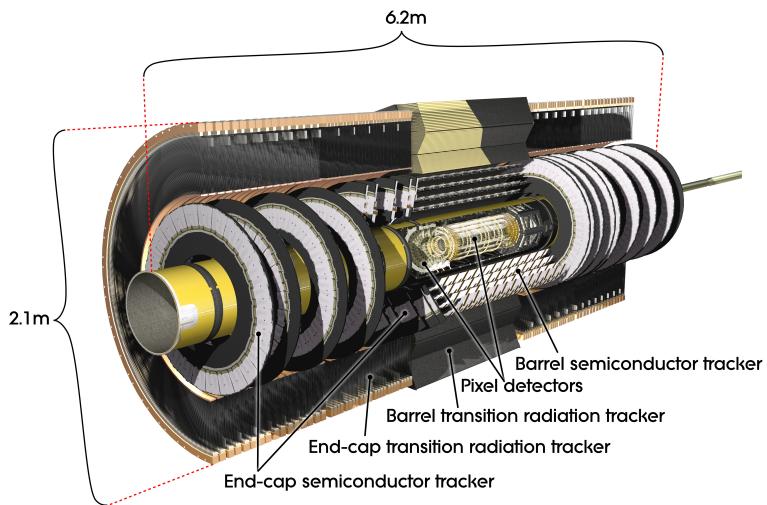


Figure 2.5: The inner detector of ATLAS. In the centre is the pixel detector, then the Semiconductor Tracker and the Transition Radiation Tracker. [10]

The pixel detector has three layers of silicon semiconductor pixels, which operate as diodes in reverse direction. Passing particles produce electron and hole pairs, which can be measured as an electric signal.

The pixel detector can determine the origin of charged particles with an accuracy of $11\mu\text{m}$. This is important for bottom tagging. Hadrons with bottom quarks have a long lifetime and decay several millimetres away from the primary vertex. Therefore one can look for a secondary vertex to identify events with a bottom quark involved.

The Semiconductor Tracker is composed of silicon semiconductor stripes and works similar to the pixel detector. For the Transition Radiation Tracker there are straw tubes, i.e. gas-filled drift chambers, with polypropylene as radiator material in between. Charged particles can ionise the gas atoms and the released electrons are accelerated with a high electric voltage. This leads to an avalanche of electrons, which can be detected as an electric signal. Additionally the polypropylene has a different refraction index from the gas in the drift tubes. Therefore the charged particles produce transition radiation which depends on the rest mass of the particle. This information is used to identify the types of the particles.

Calorimeter system Outside the solenoid magnet of the inner tracker, the calorimeter system is installed. It measures the kinetic energy of the decay products by depositing all energy. Two types of calorimeters are used. The electromagnetic calorimeter, which measures the energy of electrons and photons is the inner part. The energy of hadrons, which are not stopped in the electromagnetic calorimeter, is measured in the outer, hadronic calorimeter. The complete system is shown in figure 2.6.

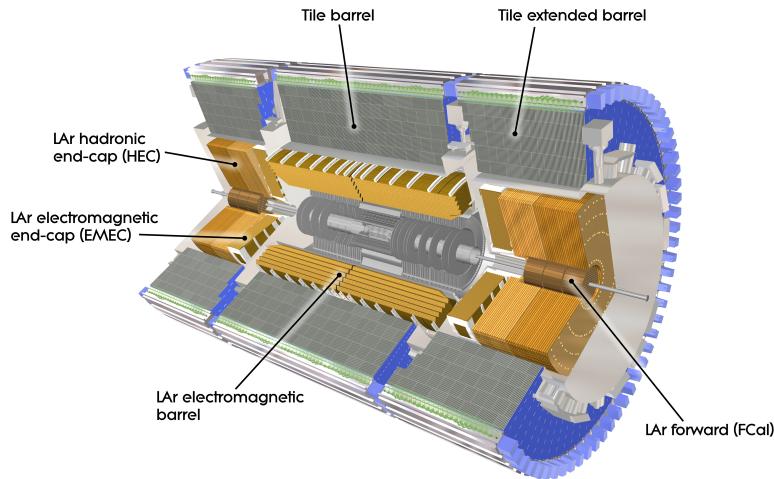


Figure 2.6: The calorimeter system of the ATLAS detector. The inner calorimeter is the electromagnetic calorimeter, surrounded by the hadronic calorimeter. [11]

Both calorimeters are so called sampling calorimeters. They consist of several layers of detector and absorber material. The absorber material produces particle showers which are detected with the detector layer. The absorber material is, depending on the calorimeter position, lead, copper, steel or tungsten. The detector part is either liquid argon or consists of scintillator tiles.

The energy resolution is much better for the electromagnetic calorimeter than for the hadronic one. The reason for this is the fact that electrons and photons radiate electromagnetic bremsstrahlung when scattering at electrons from the absorber or the detector material. The emitted photons produce more electrons and positrons; this process is called electromagnetic shower. Hadrons mainly interact by scattering at the nuclei of the absorber material. Due to the larger hadronic interaction length they are only stopped in the hadronic calorimeter for which denser material is used.

Muon spectrometer Around the calorimeters the muon spectrometer for the long-living muons is placed. The muons are not stopped in the tracker or the calorimeters. The muon spectrometer is enclosed by a toroidal magnetic field of $B = 0.5$ T in the barrel region and $B = 1$ T in the end-cap region in order to measure the muon momentum.

To detect a muon, charged capacitors are used. A crossing muon causes a discharge and this gives a electrical signal, which is measured.

Particle detection The cylindrical structure of the ATLAS detector with the different types of subdetectors is used to analyse collisions with different particles. Each particle crosses through the detector and can be identified by its characteristic interactions. This is illustrated in figure 2.7.

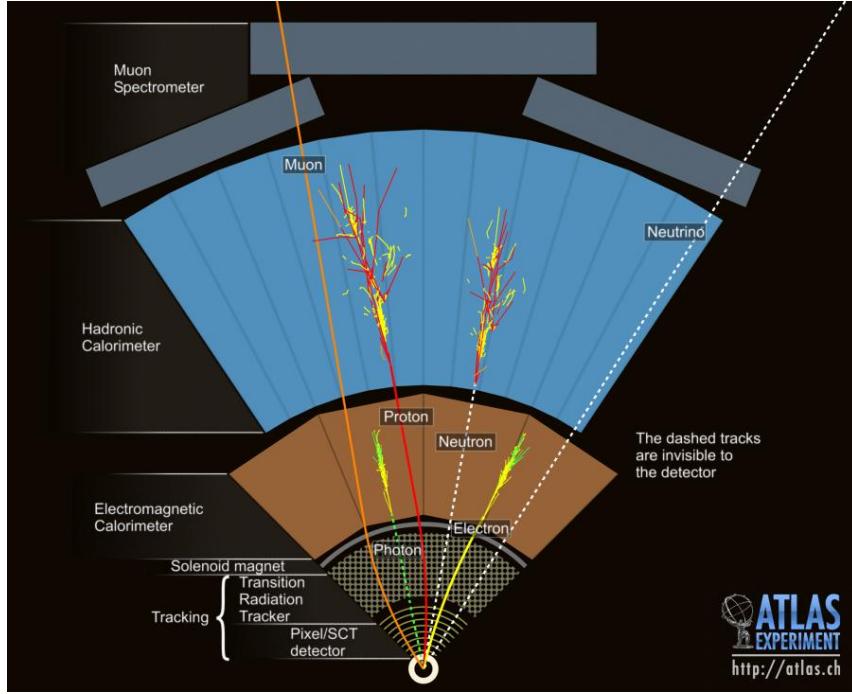


Figure 2.7: Illustration of the different particles in the ATLAS detector. [12]

As one can see, electrically charged particles – like electrons, muons or protons – have a visible track in the inner detector. Due to the magnetic bending one can determine their charge. The electrons loose their energy in the electromagnetic calorimeter. The muons are once again visible in the muon spectrometer. The protons leave a track in the electromagnetic calorimeter and a particle shower in the hadronic calorimeter.

Particles which are electrically neutral – as photons, neutrons or neutrinos – are not visible in the inner detector. The photons are visible in the electromagnetic calorimeter and one can distinguish between the electrons and the photons because of the missing track. The neutrons loose their energy in the hadronic calorimeter and are distinguishable from the protons because of the missing track in the inner detector.

The neutrinos – weakly interacting particles – are not visible in any subdetector. Using the momentum conservation one can determine the missing transverse energy. The transverse momentum of both colliding protons initially is zero. Therefore the outgoing particles must have a combined transverse momentum of zero. Is this combined transverse momentum not zero, at least one additional – not detected – particle was part of the collision. These particles could be neutrinos or also neutralinos for Supersymmetry.

3. Motivation

As described above, in this master thesis a supersymmetric decay is analysed. In the previous section the theory of Supersymmetry was described (chapter 2). The minimal form of this extension of the Standard Model, the MSSM, still has more than 100 free parameters and is therefore a very complex theory.

To study the decay of one supersymmetric particle, all possible production modes and decay chains need to be analysed in order to draw conclusions concerning the existence of this particle. Not only the complexity of such a study concerning the physical content is problematic, but also technical limitations become important. For a study, which should consider all free parameters, signal samples for each combination of parameters are needed. This is impossible for such a complex theory as the MSSM.

For the concrete analysis simplified models are studied. These models have fewer free parameters. Therefore both the production modes and the possible decay chains are known and can be analysed. In addition, all combinations of different parameters – as, for example, the particle masses – can be studied in the analysis.

While the MSSM has more than 100 free parameters, one can assume a simplified model with only two free parameters as a subset of the complex theory. In this simplified model only one squark is considered to be light in order to conserve the naturalness. All other squark masses are well above the production threshold. Therefore only this one squark can be produced and has to decay into the lightest supersymmetric particle (LSP). This scenario is analysed and limits on the masses of the squark and the LSP are obtained. Nevertheless the MSSM is still not excluded because it has far more free parameters than the simplified models, which are analysed at the moment.

In this master thesis the supersymmetric partner of the top quark – the stop quark \tilde{t} (with its antiparticle \tilde{t}^*) – is assumed to be the lightest squark and the neutralino is the LSP. All other superparticles are considered to be too heavy either to be produced directly or to be part of the decay chain. For the decay it is assumed that the stop squark decays into a top quark and a neutralino with a branching ratio of 100 %.

Due to the R-parity, stop quarks can only be produced in pairs. Therefore the decay $\tilde{t}\tilde{t}^* \rightarrow t\bar{t}\tilde{\chi}_0^1\tilde{\chi}_0^1$ (compare figure 3.1) is considered. The main difference to the Standard Model $t\bar{t}$ decay are the two neutralinos which contribute to the missing transverse energy. Therefore the Standard Model $t\bar{t}$ decay is the main background. Other backgrounds are $W+jets$, $Z+jets$, single top and diboson processes.

The production modes of a stop quark pair are similar to those of a SM $t\bar{t}$ process. For this analysis the semileptonic decay is analysed. This means that one top quark decays into a bottom quark and a hadronically decaying W boson, while the other top quark decays into a bottom quark and a leptonically decaying W boson. This means that the leptonically decaying W decays into a lepton and the corresponding neutrino.

This analysis considers the lepton, the four jets as well as the missing transverse energy due to the neutrino and in the supersymmetric decay also due to the neutralinos. This decay channel is chosen because it is easier to analyse than the full hadronic decay, as the lepton provides a clear trigger signal. The statistics is higher than for the full leptonic decay as the branching ratios of the W decay favour hadrons.

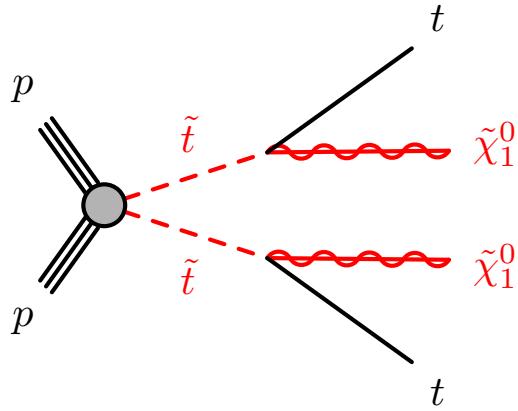


Figure 3.1: Feynman graph for a stop quark pair decaying into two top quarks and two neutralinos. [13]

For the analysis three different regions depending on the stop quark and the neutralino mass are used. An illustration for these three regions is shown in figure 3.2. The signals considered are all under the grey dashed line. This line shows the boundary of the decay $\tilde{t} \rightarrow t\tilde{\chi}_1^0$. Above this line the difference of the stop quark and the neutralino mass $m_{\tilde{t}} - m_{\tilde{\chi}}$ is bigger than the top quark mass and therefore the decay is not possible.

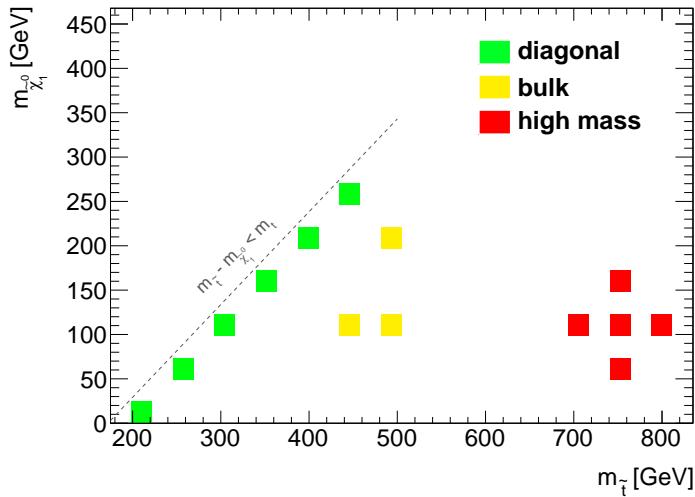


Figure 3.2: Illustration of the used mass range for the stop quark and the neutralino. The grey dashed line illustrates the boundary of the decay $\tilde{t} \rightarrow t\tilde{\chi}_1^0$. The three colours indicate where the three regions – diagonal, bulk and high mass – are located within the parameter space.

The coloured points indicate the position of different samples of the three analysis regions. The illustrated samples here are used for the training of the neural networks. More samples are used for the limit setting, which are listed in the next chapter. The green points show

the diagonal region which are samples near the boundary. The signal in this region is very similar to the Standard Model $t\bar{t}$ decay, because the additional missing transverse energy due to the neutralinos is very small. The yellow points illustrate the bulk region, which is the region in the middle. This bulk region is for stop masses around 500 GeV and neutralino masses up to 250 GeV. The high mass region is shown in red and stop quarks with high masses are considered here. The biggest problem in this region is the small number of expected events and therefore only few statistics for the analysis.

4. General Information

4.1. Simulated data samples

For the background and signal descriptions different Monte Carlo (MC) simulated samples are used. The samples which are used are the same as used in the official ATLAS analysis for the process $\tilde{t}\tilde{t}^* \rightarrow t\bar{t}\tilde{\chi}_0^1\tilde{\chi}_0^1$ with one isolated lepton in the final state [13, 14]. All samples either use a GEANT4 detector simulation [15, 16] or use a fast simulation of the ATLAS detector. For the fast simulation the electromagnetic and hadronic showers are simulated using a parametrised description [17], the remaining part of the detector is again simulated with GEANT4.

The used data samples were recorded by the ATLAS detector during the data taking periods A3-L3 (runs 200841 to 215643). The official standard ATLAS Good Run List was used to apply basic data quality requirements. Afterwards the needed triggers and the event selection were applied, which is described below.

Background samples All background samples are produced with a top quark mass of $m_t = 172.5$ GeV. The $t\bar{t}$ sample is generated using PowHEG [18] with the parton distribution function (PDF) CT10 [19]. The single top processes for the s - and the Wt -channel are also produced with PowHEG using the CT10 PDF. For the single top t -channel ACERMC [20] with the PDF CTEQ6L1 [21] is used. W , Z , WW and single top production with a Z boson are generated with MADGRAPH [22]. $W+\text{jets}$, $Z+\text{jets}$ and all diboson processes (WW , ZZ and WZ) are produced using SHERPA [23] with the CT10 PDF set. For PowHEG, ACERMC and MADGRAPH the fragmentation and hadronisation is done with PYTHIA [24], for the SHERPA production the built-in function is used.

Signal samples The signal samples for the decay $\tilde{t}_1 \rightarrow t\tilde{\chi}_1^0$ with a branching ratio of 100% are produced using HERWIG++ [25] with the CTEQ6L1 PDF set. The signals are produced for stop quark masses between 200 GeV and 800 GeV and for neutralino masses between 1 GeV and 620 GeV in order to cover the complete parameter space. The spacing between the different grid points is mostly 50 GeV. For the diagonal region – where $m_{\tilde{t}_1} \gtrsim m_t + m_{\tilde{\chi}_1^0}$ – more samples are produced in order to have sufficient samples near the boundary. The \tilde{t}_1 is mostly the partner to the right handed top quark (about 70%) and the neutralino is assumed to be almost pure bino.

A detailed list of the Monte Carlo samples used is given in the appendix A.1.

4.2. Triggers

For the trigger, a combination of single lepton triggers (electron or muon) and an \cancel{E}_T trigger was used. For events in the electron (muon) channel, the electron (muon) trigger is used. The \cancel{E}_T trigger is applied in both channels in order to enhance the overall trigger efficiency.

If an event activates the lepton, or both the lepton and the \cancel{E}_T trigger, it will be taken from the lepton trigger data set. Events which only activate the \cancel{E}_T trigger but not the lepton trigger are taken from the \cancel{E}_T trigger data set. With this arrangement most events

are selected from the lepton trigger where the efficiency correction factors are the most accurate.

The recorded data events are stored in the Egamma, Muons and JetTauEtmiss stream. These streams overlap, which means that one event can be stored in two streams. The selection ordering is the same as for the simulated samples, to avoid a double counting of events.

4.3. Object definition and selection

The following object definitions describe the particles which are used in the event selection. It follows the official ATLAS analysis for $\tilde{t} \rightarrow t\tilde{\chi}_1^0$ with one isolated lepton in the final state [13, 14].

Jets For the reconstruction of jets the anti- k_t jet algorithm [26, 27] with a distance parameter $R = 0.4$ is used. All jets with a $p_T > 20 \text{ GeV}$ and $|\eta| < 2.8$ are used for the object overlap removal which is described in paragraph 4.3.

After the overlap removal the jets have to pass tighter selections – $p_T > 25 \text{ GeV}$ and $|\eta| < 2.5$. If one jet fails one of these criteria the event is rejected.

Using the MV1 b-tagging algorithm at 70% efficiency operating point, jets are identified as b -jets. Using the b -jet scale factors provided by the FlavourTagging working group [28–30], a weight for each jet and for the event itself is calculated. Using this weight the tagging rate in the simulation is corrected in order to be similar to that in data.

Electrons Electron candidates are identified using the `loose++` definition. In addition, most of the energy deposit of the electron has to be in the electromagnetic calorimeter. Furthermore, $E_T = E_{\text{cluster}} / \cosh \eta > 10 \text{ GeV}$ is required. If the track of the electron contains at least 4 silicon hits η^{track} is used, otherwise $|\eta^{\text{cluster}}|$, which has to be smaller than 2.47. In addition, electrons are required to have a $p_T > 10 \text{ GeV}$. Electron candidates which fulfil all these criteria are called baseline electrons. These are used for the overlap removal and only after a tighter selection – $p_T > 25 \text{ GeV}$ and `tight++` selection – they are called signal electrons.

Muons For the reconstruction of muons, a track in the muon spectrometer is combined with the corresponding track in the inner detector. As first acceptance cuts muons have to fulfil: $p_T > 10 \text{ GeV}$ and $|\eta| < 2.4$.

The tracks for the muon reconstruction need to have at least one hit in any pixel layer and five SCT hits. In addition, the sum of the holes – that means the sum of the expected, but not measured hits – must be lower than three.

After these baseline selections the overlap removal is performed. Muons with a transverse (longitudinal) impact parameter larger than 0.2 (1) mm with respect to the primary vertex are rejected in order to remove events with a possible cosmic muon.

For the final selection the muons have to have a $p_T > 25 \text{ GeV}$ and $|\eta| < 2.4$. After this they are called signal muons.

Missing transverse momentum To calculate the missing transverse energy an object-based \cancel{E}_T algorithm based on MET_Egamma10NoTau is used. The vectorial sum of all transverse momenta of calibrated jets, electrons, muons and TopoClusters, which are not used for any reconstructed object, is calculated.

Overlap removal The signature of physic objects in the detector is ambiguous. Therefore an overlap removal is performed to remove single objects which are reconstructed in two categories. All objects which satisfy the baseline selection are used during the overlap removal process. All criteria are based on the distance parameter $\Delta R = \sqrt{\Delta\varphi^2 + \Delta\eta^2}$. For the overlap removal the following order is used:

1. If a baseline electron and a baseline jet are found within $\Delta R < 0.2$, the object is interpreted as an electron.
2. If a baseline electron and a baseline jet are found within $0.2 \leq \Delta R \leq 0.4$, the object is interpreted as a jet and the electron is ignored, because it is not isolated.
3. If a baseline muon and a baseline jet are found within $\Delta R < 0.4$, the object is interpreted as a jet and the muon is ignored, because it is not isolated.

After this overlap removal the objects have to fulfil the tighter selection as described above for each object individually.

4.4. General event preselection

For the general event preselection the objects which have been described before are used. Exactly one signal lepton – electron or muon – is required to pass the event selection. Additionally, at least four jets are required, of which one has to be b-tagged. The transverse momenta of these four jets have to be larger than 80, 60, 25, 25 GeV.

Cut	Value
Number of jets	≥ 4
Jet p_T	$> 80, 60, 25, 25$ GeV
Number of b-jets	≥ 1
Number of leptons	= 1 (electron or muon)
\cancel{E}_T	> 80 GeV
m_T	> 50 GeV

Table 4.1: General event preselection after the object selection.

In order to remove some of the main backgrounds, a cut on the missing transverse energy and the transverse mass is applied. All cuts are listed in table 4.1.

To calculate all variables, which are explained in chapter 5, both top quarks are reconstructed. First the hadronically decaying top quark is reconstructed. For this, two jets which have the minimal distance $\Delta R(j_1, j_2)$ and a common mass $m(j_1 + j_2)$ larger than 60 GeV

are combined. Then these two jets are combined with a third jet j_3 which has the minimal distance to both other jets $\Delta R(j_3, j_1 + j_2)$. In addition the common mass of all three jets $m(j_1 + j_2 + j_3)$ should be larger than 130 GeV. These three jets are added up to calculate the hadronically decaying top quark [13, 14].

$$t_{\text{had}} = j_1 + j_2 + j_3 \quad (11)$$

For the leptonically decaying top quark the jet j_4 with the highest transverse momentum is selected, which is not used in the reconstruction of the hadronically decaying top quark. In the end the signal lepton ℓ is combined with this jet to calculate the leptonically decaying top quark.

$$t_{\text{lep}} = j_4 + \ell \quad (12)$$

The missing transverse energy is not used for this reconstruction because of the additional contribution from the neutralinos [31]. A graphical illustration of the reconstruction of both top quarks is shown in figure 4.1.

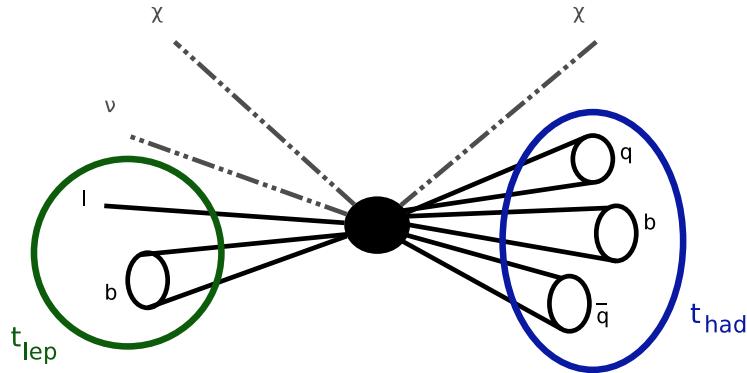


Figure 4.1: Illustration of the top quark reconstruction used in the following analysis. The green circle (left side) includes all particles which are used for the reconstruction of the leptonically decaying top quark, the blue circle (right side) shows all particles for the hadronically decaying top quark.

5. Distinction of signal and background

There are many variables which differ between the Standard Model decays, like $t\bar{t}$ and $W+\text{jets}$, and the $\tilde{t}\tilde{t}^*$ decay. These variables are explained in the following. For this, each variable is shown for the SM $t\bar{t}$ and $W+\text{jets}$ decays, which are the main backgrounds. Additionally, each distribution is shown for three different $\tilde{t}\tilde{t}^*$ samples which are all produced with a neutralino mass of 100 GeV. The sample for the diagonal region has a stop mass of 300 GeV ($\tilde{t}(300, 100)$), the sample for the bulk region has a stop mass of 500 GeV ($\tilde{t}(500, 100)$) and the sample for the high mass region has a stop mass of 750 GeV ($\tilde{t}(750, 100)$).

The shown distributions only apply to the electron channel, but the muon channel does not differ much. All distributions are produced using the SPyRoot framework.

5.1. Variables used in the official ATLAS analysis

In the first part, variables which are also used by the official ATLAS analysis for the decay $\tilde{t}\tilde{t}^* \rightarrow t\bar{t}\tilde{\chi}_1^0\tilde{\chi}_1^0$ with one isolated lepton in the final state [13, 14] are explained.

Missing transverse energy: One of the main differences between the Standard Model decays and the $\tilde{t}\tilde{t}^*$ decay is, as explained before, the existence of the two neutralinos. These neutralinos are not detected, so they contribute to the missing transverse energy. Therefore one would expect differences in the distribution of the missing transverse energy, which is shown in figure 5.1. Here, an increase in the missing transverse energy can be seen especially for higher stop masses. Therefore the separation between the Standard Model and $\tilde{t}\tilde{t}^*$ decay grows for higher stop masses.

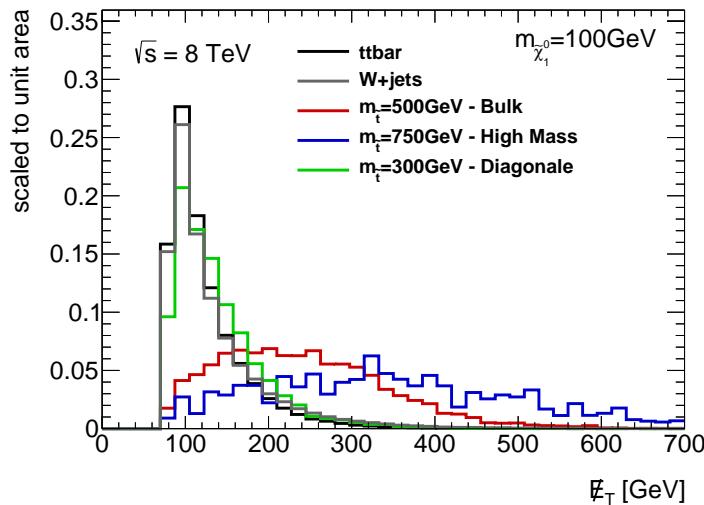


Figure 5.1: The missing transverse energy \cancel{E}_T for the Standard Model decays and for the $\tilde{t}\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

Transverse mass: Another variable is the so called transverse mass

$$m_T^2 = 2 \cdot p_T^\ell \cdot \cancel{E}_T \cdot (1 - \cos(\Delta\varphi)), \quad (13)$$

where $\Delta\varphi$ is the transverse angle between the lepton (ℓ) and the vector of the missing transverse momentum. The distribution is shown in figure 5.2.

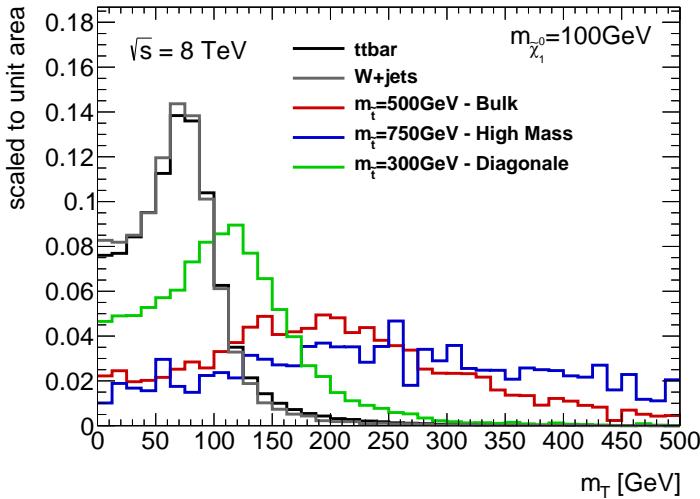


Figure 5.2: The transverse mass for the Standard Model decays and for the $t\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

This variable is often called m_T^W because it is equivalent to the transverse W mass for Standard Model samples with a W decay inside. Therefore the transverse mass is mostly below 80 GeV for Standard Model decays. For the different $t\tilde{t}^*$ decays the m_T does not match the transverse W mass. Due to the additional missing transverse energy caused by the neutralinos, the transverse mass also reaches higher values. This effect gets stronger for higher stop masses, which is shown in the plot.

\cancel{E}_T significance: The \cancel{E}_T significance is used in order to suppress mismeasured missing transverse energy. For this the missing transverse energy \cancel{E}_T is divided by $\sqrt{H_T}$, where H_T is defined as

$$H_T = \sum_{i=1}^4 p_T^{\text{jet } i}.$$

The differences between the Standard Model decays and the stop decay are expected to be similar to the missing transverse energy, due to the high correlation between both variables. The distribution is shown in figure 5.3.

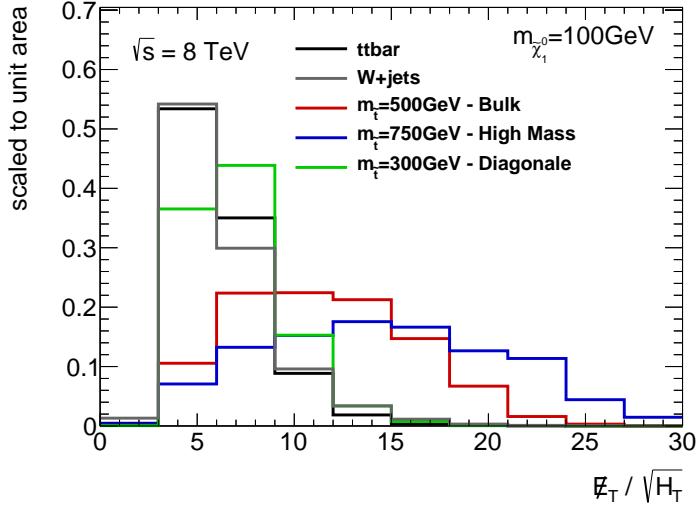


Figure 5.3: The \cancel{E}_T significance for the Standard Model decays and for the $\tilde{t}\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

Angle between \cancel{E}_T and one jet: Additionally, the transverse angles between the missing transverse energy \cancel{E}_T and both the leading and the second leading jet are calculated. The corresponding distributions are shown in figure 5.4. These angles are calculated in order to minimise the impact of detector effects. If a jet is mismeasured, the difference between the measurement and the real jet would be included into the calculation of the missing transverse energy. Therefore the missing transverse energy would be more orientated towards the jet direction. Cutting at these two angles reduces the number of mismeasured jets.

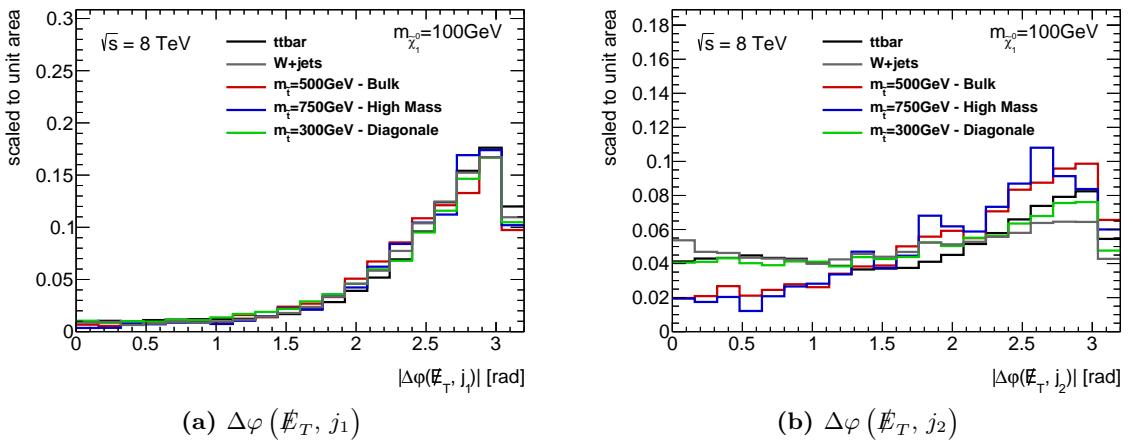


Figure 5.4: The transverse angle $\Delta\varphi$ between the \cancel{E}_T and one jet for the Standard Model decays and the $\tilde{t}\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

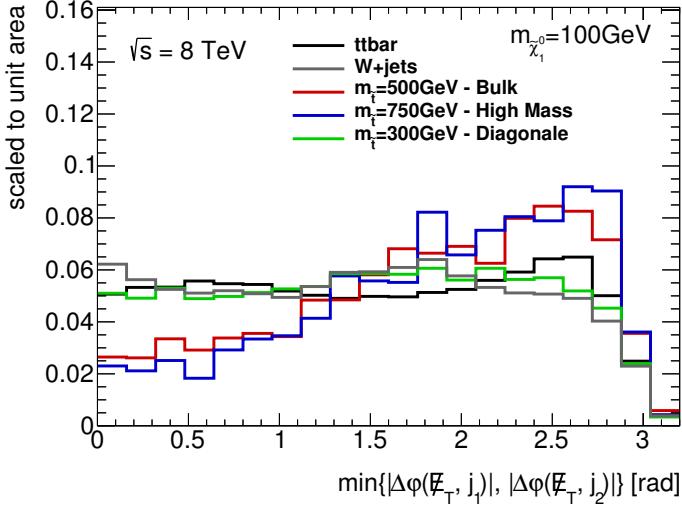


Figure 5.5: The minimum transverse angle $\Delta\varphi$ between the missing transverse energy \cancel{E}_T and one of the first two jets for the Standard Model decays and for the $\tilde{t}\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

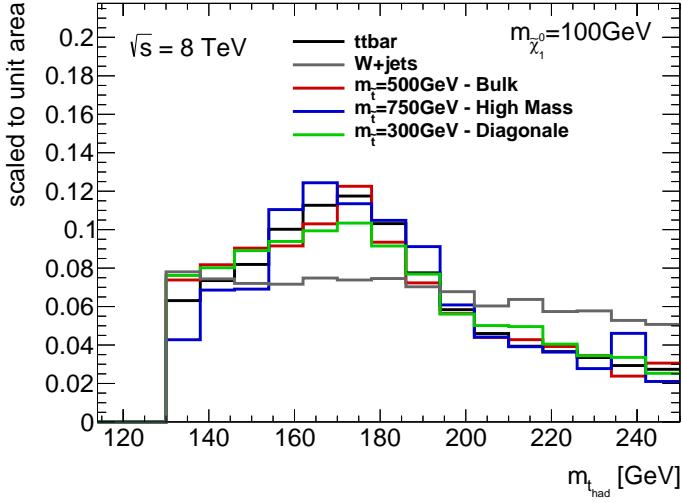


Figure 5.6: The mass of the hadronically decaying top quark for the Standard Model decays and for the $\tilde{t}\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

Using the leading jet, the distributions do not differ for the different decays. This means that the missing transverse energy is mostly orientated in the opposite direction of the leading jet. Using the second leading jet, the differences are more visible. For the SM decays and the diagonal sample – $\tilde{t}(300, 100)$ – the distribution is mostly flat. For the other $t\bar{t}\tilde{t}^*$ samples the missing transverse energy and the second leading jet are mostly orientated in the opposite direction.

To enhance the differences, the minimum of these two transverse angles can be used. The corresponding distribution is shown in figure 5.5. The minimum angle is strongly correlated to the other two angles.

Mass of the hadronically decaying top quark: In figure 5.6 the distribution of the mass of the hadronically decaying top quark is shown. This mass is similar for the stop and the $t\bar{t}$ sample and its maximum is around 170 GeV as expected. Only the $W+jets$ sample differs because no real top quark can be reconstructed in this sample. Therefore this variable helps mostly against background without any top quarks.

Topness: This variable is introduced in order to suppress dileptonic $t\bar{t}$ events where one lepton is not reconstructed [32]. It is defined as $\ln(\min S)$ where S is given as

$$\begin{aligned} S(p_{W,x}, p_{W,y}, p_{W,z}, p_{\nu,z}) = & \frac{(m_W^2 - (p_\ell + p_\nu)^2)^2}{a_W^4} + \frac{(m_t^2 - (p_{j1} + p_\ell + p_\nu)^2)^2}{a_t^4} \\ & + \frac{(m_t^2 - (p_{j2} + p_W)^2)^2}{a_t^4} + \frac{(4 \cdot m_t^2 - (\sum p)^2)^2}{a_{CM}^4}. \end{aligned} \quad (14)$$

The calculation of the topness variable is similar to a χ^2 -method, this implies that S is minimised for all possible momentum combinations. The first term in equation 14 constraints the reconstructed W mass from the missed lepton, the second term the reconstructed top mass from the identified lepton, the third term the reconstructed top mass from the missed lepton and the last term minimises the reconstructed center of mass energy.

For the minimisation, the neutrinos are assumed to be massless and the W boson with the missed lepton is assumed to be on-shell. For the jet momenta (p_{j1}, p_{j2}) all possible combinations of a b-tagged jet and the leading or subleading jet are evaluated. The given constants in the equation are used to model the detector resolution and are set to: $a_W = 5$ GeV, $a_t = 15$ GeV and $a_{CM} = 1$ TeV [32].

The distribution of this variable is shown in figure 5.7. The $t\bar{t}$ sample differs strongly from the stop samples with higher stop masses, but is similar to the diagonal sample. Both distributions have their maximum for negative values, while the stop sample with the highest stop mass tends to have its maximum for positive values. The distributions of the stop sample from the bulk region and the $W+jets$ sample have two similar maxima.

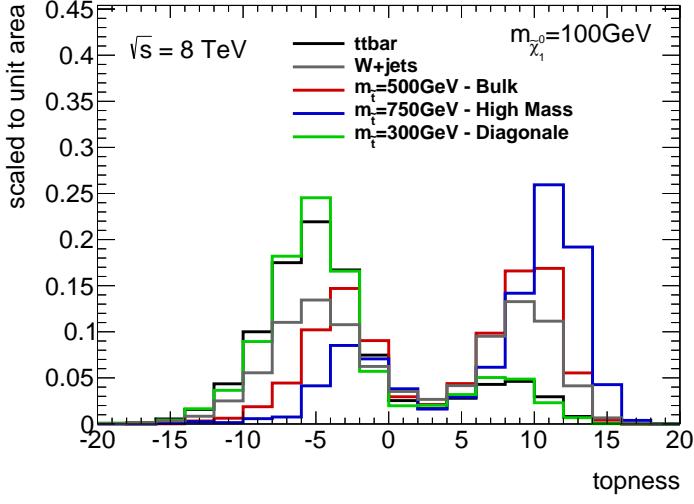


Figure 5.7: The topness distribution for the Standard Model decays and for the $t\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

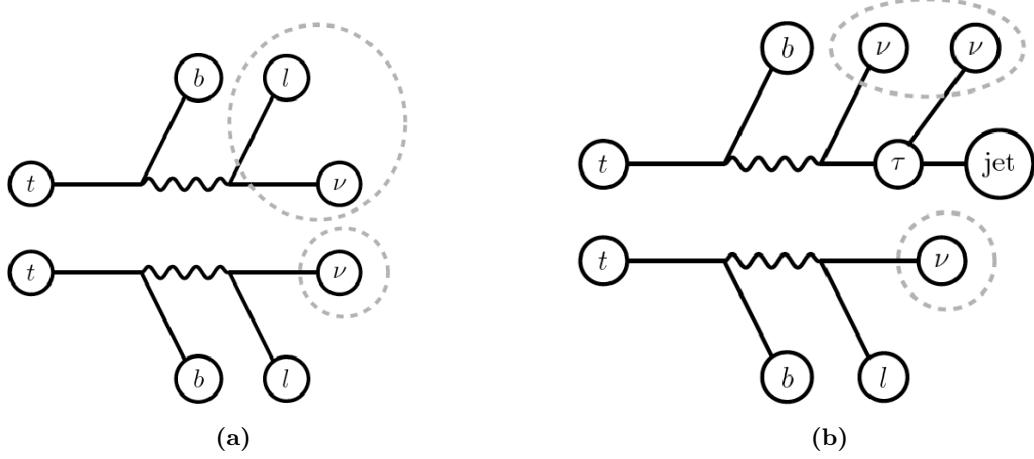


Figure 5.8: Illustration of the amT_2 (a) and m_{T2}^τ (b) variables. These variables are used to reduce dileptonic $t\bar{t}$ background where one lepton is either lost (a) or decays hadronically into a τ (b). The dashed lines indicate the “missing” objects which are used in the minimization [13, 14].

stransverse mass: The stransverse mass m_{T2} is used to bound masses from a pair of particles decaying into two visible and two invisible particles (here notated in a general form using a and b for the two particles). Therefore the stransverse mass is a function of these two visible particles and the missing transverse energy as given in equation 15, where m_{Ti} is the already defined transverse mass (cf. equation 13).

$$m_{T2} = \min_{\overrightarrow{p_{Ta}} + \overrightarrow{p_{Tb}} = \not{E}_T} \{ \max(m_{Ta}, m_{Tb}) \} \quad (15)$$

For the $t\tilde{t}^*$ analysis, there are two variables built upon this stransverse mass variable. Both variables are built mostly against the dileptonic $t\bar{t}$ decay where one lepton is misidentified. This can happen due to a wrong particle identification or the particle being outside the acceptance of the detector. The dileptonic $t\bar{t}$ background is the main background for the whole analysis. Especially the am_{T2} variable is sensitive to the second lepton, which is not reconstructed as illustrated in figure 5.8a. The corresponding distribution is shown in figure 5.9. The diagonal and the $t\bar{t}$ sample have a similar shape, as well as the bulk and the $W+jets$ sample. Only the high mass sample differs from the Standard Model decays due to the longer tail.

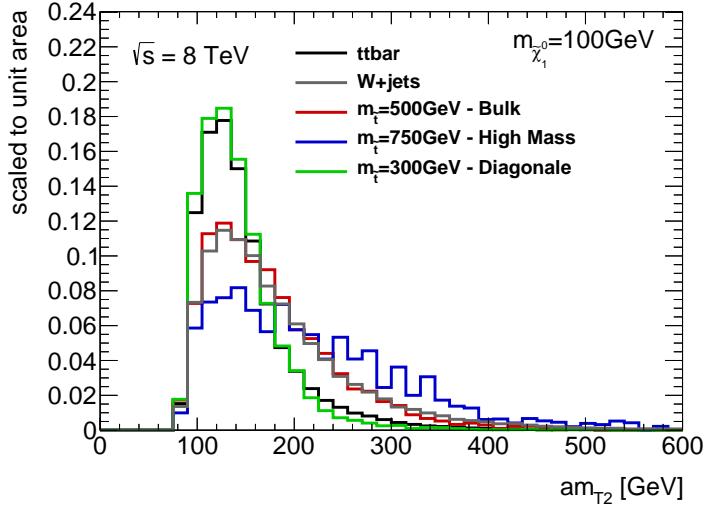


Figure 5.9: am_{T2} for the Standard Model decays and for the $t\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

The second variable similar to the stransverse mass is the m_{T2}^τ , which is designed to help against hadronically decaying tau leptons. This is illustrated in figure 5.8b. In figure 5.10 the distribution for the variable is plotted. Again the Standard Model decays and the diagonal sample look similar. The bulk and the high mass sample have longer tails and therefore more differences to the Standard Model decays.

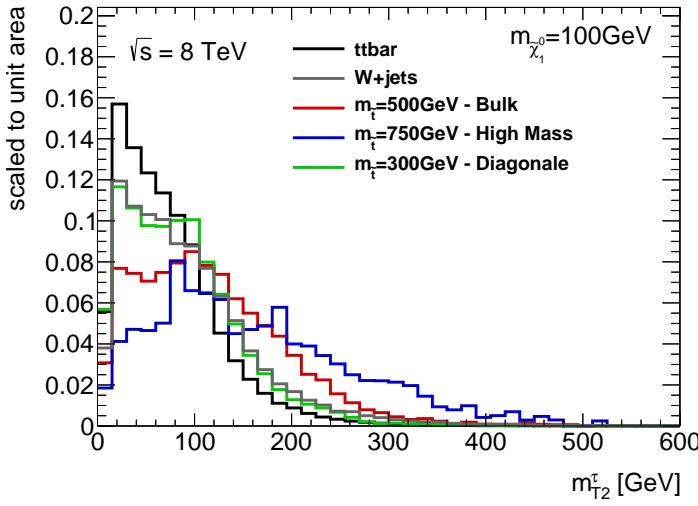


Figure 5.10: m_{T2}^{τ} for the Standard Model decays and for the $\tilde{t}\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

Track veto: As already mentioned, the main background in the signal-like regions is the dileptonic $t\bar{t}$ background. This background becomes dominant if one of the two leptons is not well reconstructed and therefore contributes to the missing transverse energy. These leptons may leave an isolated track in the Inner Tracker. This track can be used in order to veto such events. Using this veto, the dileptonic $t\bar{t}$ background is reduced.

In order to enhance the signal efficiency, the charge of the track is required to be the opposite of the signal lepton. Therefore less signal is removed without changing the background rejection.

Tau veto: Standard Model $t\bar{t}$ events including a tau lepton make up a large part of the background for this analysis. The additional neutrinos from the tau decay behave similar to the neutralinos from the signal process and provide a source for real missing transverse energy. Therefore the ATLAS tau identification is used to identify and veto events including a hadronically decaying tau lepton. The transverse momentum of these tau candidates has to be larger than 15 GeV. As for the track veto, the charge of the tau candidate has to be the opposite of the signal lepton.

5.2. Additional variables

In the second part, variables are shown which are not used in the official ATLAS analysis but can be used for a better separation between the signal and the background samples. This variables were analysed in the foregoing bachelor thesis [31].

Perpendicular \cancel{E}_T component to the leptonically decaying top quark: The perpendicular component of the missing transverse energy to the leptonically decaying top quark is determined by boosting the missing transverse energy and the leptonically decaying

top quark into the $t\bar{t}$ rest system. From the missing transverse energy only the x- and the y-component are known. Therefore only these two components for the leptonically decaying top quark are used as well. With this, the length of the perpendicular component of the \cancel{E}_T to the leptonically decaying top quark can be calculated. This is illustrated in figure 5.11. The corresponding distribution is shown in figure 5.12.

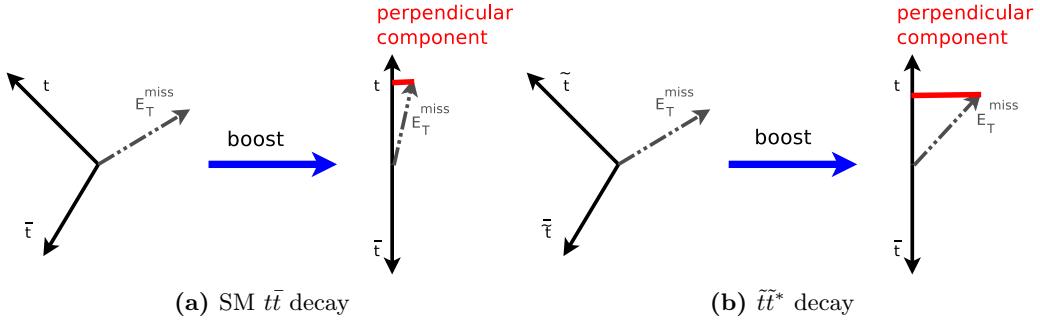


Figure 5.11: Illustration of the perpendicular \cancel{E}_T component to the leptonically decaying top quark. In the Standard Model decay, the neutrino is orientated in the same direction as the leptonically decaying top quark, therefore the component is small (a). In the $t\tilde{t}^*$ decay, the neutralinos contribute to the missing transverse energy and they are not collinear to the leptonically decaying top quark. Therefore the component is larger (b) than in the Standard Model decay.

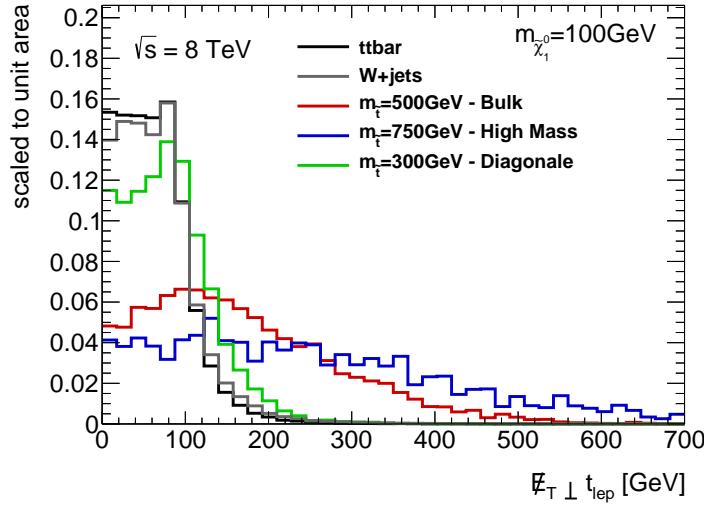


Figure 5.12: The perpendicular \cancel{E}_T component to the leptonically decaying top quark for the Standard Model decays and the $t\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

For the SM $t\bar{t}$ decay, small values are observed, because the neutrino is mostly collinear to the leptonically decaying top quark. In the $t\tilde{t}^*$ decay, the neutralinos contribute to the

missing transverse energy. Therefore the missing transverse energy is no longer collinear to the leptonically decaying top quark. This leads to higher values for the perpendicular component.

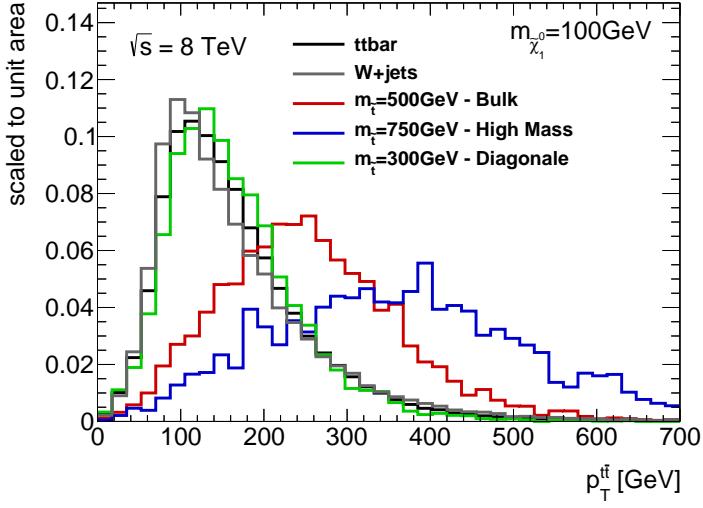


Figure 5.13: The transverse momentum of the $t\bar{t}$ pair for the Standard Model decays and the $\tilde{t}\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

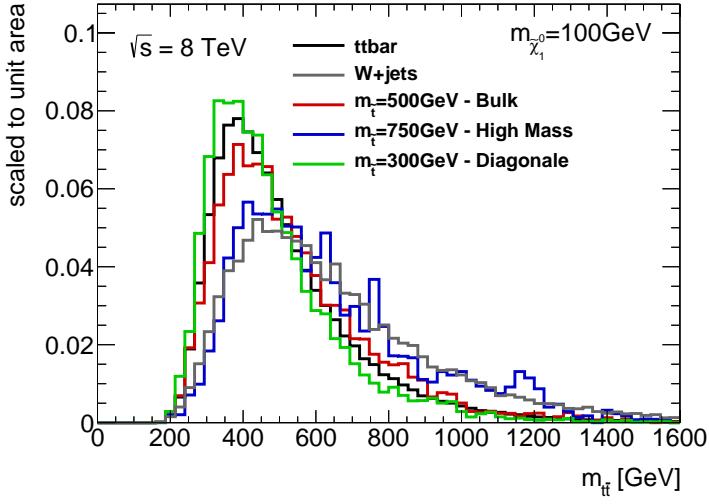


Figure 5.14: The mass of the $t\bar{t}$ pair for the Standard Model decays and the $\tilde{t}\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

Transverse momentum of the $t\bar{t}$ pair: The transverse momentum of the $t\bar{t}$ pair is relatively small for the Standard Model decay. For the $\tilde{t}\tilde{t}^*$ decay the transverse momentum

becomes larger for higher stop quark masses. This is due to the additional energy which the top quarks receive during the stop quark decay. The distribution is shown in figure 5.13, and one can see that there is good distinction between the Standard Model samples and $\tilde{t}\tilde{t}^*$ sample with higher stop quark masses.

Mass of the $t\bar{t}$ pair: In figure 5.14 the mass of the $t\bar{t}$ pair is shown. For the Standard Model $t\bar{t}$ decay the expected exponential drop off, which is starting at the production threshold around 350 GeV is visible. The diagonal sample ($\tilde{t}(300, 100)$) shows a similar behaviour as the $t\bar{t}$ sample. The distributions for the other samples differ more from the $t\bar{t}$ background because the maximum shifts to higher values and the tails become longer as also visible for the $W+jets$ sample.

Transverse angle between both top quarks: For the Standard Model $t\bar{t}$ decay both top quarks are predominantly situated back to back. Therefore the transverse angle between both top quarks is around π . In the $\tilde{t}\tilde{t}^*$ decay both stop quarks are also produced back to back. Due to the subsequent decay into a top quark and a neutralino, the top and stop quark are not completely collinear. This leads to smaller angles between both top quarks, which can be observed in figure 5.15.

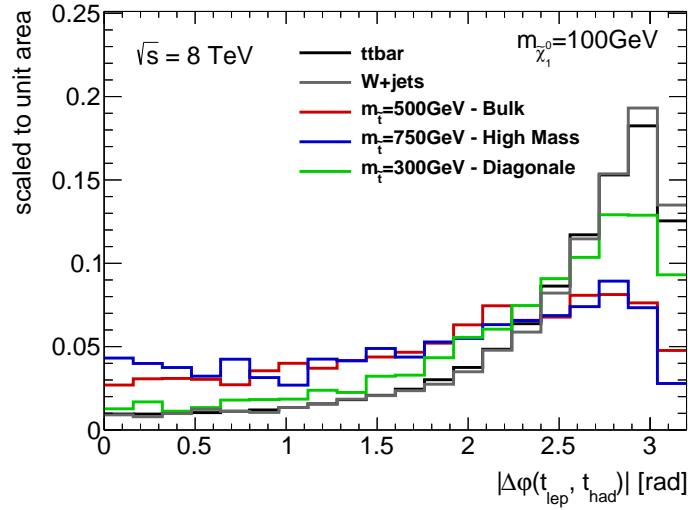


Figure 5.15: The transverse angle $\Delta\varphi$ between both top quarks for the Standard Model decays and the $\tilde{t}\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

Transverse angle between \cancel{E}_T and one top quark: In figure 5.16 one can see the transverse angle $\Delta\varphi$ between the missing transverse energy \cancel{E}_T and one top quark. For the leptonically decaying top quark, differences between the Standard Model $t\bar{t}$ decay and the $\tilde{t}\tilde{t}^*$ decay are visible. In the Standard Model decay only the neutrino contributes to the missing transverse energy. Therefore the angle between the top quark and the neutrino, i.e. the missing transverse energy, is rather small.

In the $\tilde{t}\tilde{t}^*$ decay the two neutralinos additionally contribute to the missing transverse energy. Therefore the angle between the top quark and the missing transverse energy differs compared to the Standard Model decay.

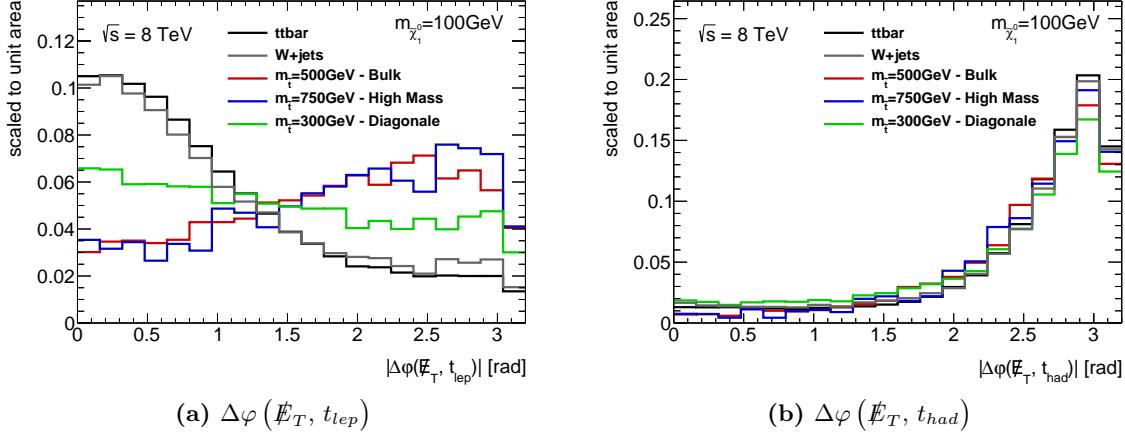


Figure 5.16: The transverse angle $\Delta\varphi$ between the missing transverse energy \not{E}_T and one top quark for the Standard Model decays and the $\tilde{t}\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

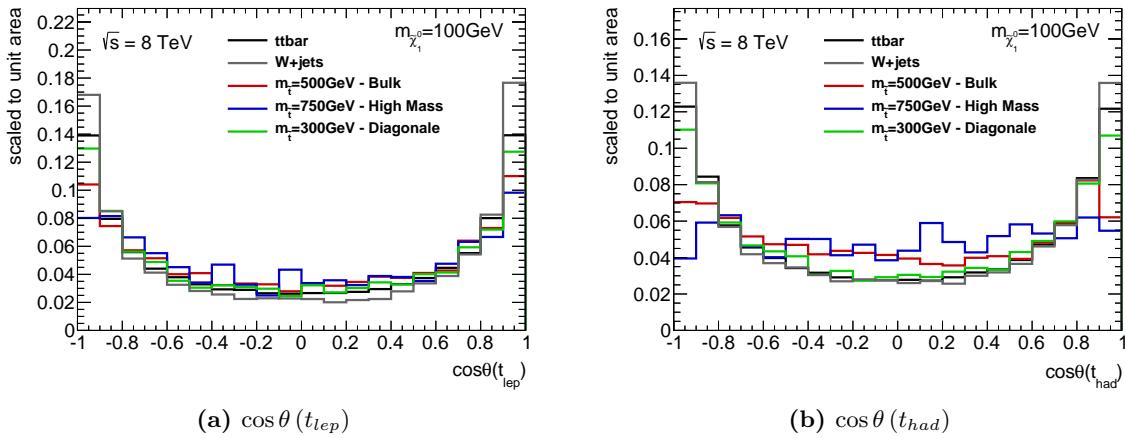


Figure 5.17: The polar angle $\cos\theta$ for each top quark for the Standard Model decays and the $\tilde{t}\tilde{t}^*$ decay is shown for the electron channel. The distributions are normalised to unit area in order to show the shapes.

The angle between the missing transverse energy and the hadronically decaying top quark matches the expectation for the Standard Model decay. As shown before, the leptonically and the hadronically decaying top quark are mostly back to back (see paragraph 5.2) and the leptonically decaying top quark and the missing transverse energy are nearly collinear.

Therefore the hadronically decaying top quark and the missing transverse energy are also back to back. For the $\tilde{t}\tilde{t}^*$ decay one would expect the angle between the hadronically decaying top quark and the missing transverse energy to look similar to the leptonically decaying top quark. As one can see in the distribution, this is not the case. Also, for the $\tilde{t}\tilde{t}^*$ decay the hadronically decaying top quark and the missing transverse energy are mostly produced back to back.

Angle of the single top quarks: Furthermore the polar angle for each top quark is shown in figure 5.17. In the Standard Model $t\bar{t}$ decay the top quarks are nearly produced in rest. Due to the longitudinal boost they are collinear to the beam axis, which translates to $\cos\theta = \pm 1$ as can be seen in the distribution.

In the $\tilde{t}\tilde{t}^*$ decay the stop quarks are produced in rest. During the decay, the top quarks and the neutralinos achieve more energy for higher stop quark masses. This leads to smaller absolute values for $\cos\theta$ in the $\tilde{t}\tilde{t}^*$ decay, especially when the stop squark is heavy. The behaviour for the leptonically and the hadronically decaying top quark is similar.

5.3. Conclusion

The explained variables are all used as starting point for the optimisation of neural networks. During the optimisation procedure which is described in chapter 6 some variables show good behaviour. This means that they either help to separate between the signal and the background processes or that they help to improve the training in general.

As already visible in the different distributions, there are variables which are powerful at least for one of the three analysis regions. For example, the separation power of kinematic variables is strong for the high mass region. But for the diagonal region the similarity between the signal and the background samples is too large. For the diagonal region especially angular distributions are good, because there signal and background differ more. All variables which are used for the following multivariate analysis have to be validated for their agreement of data and the Standard Model prediction. This study is explained in chapter 7.

6. Evaluation and optimisation of multivariate techniques

The major part of this master thesis deals with the optimisation of neural networks using the variables described in the previous section. This optimisation was done separately for all three analysis regions in order to include their individual requirements.

6.1. Multivariate methods

In order to separate between signal and background events, simple cuts at different variables can be used. If the shape of these distributions does not differ much between the signal and background sample, simple cuts are not the best solution. Using a multivariate method in order to separate between signal and background events is favourable, because it considers the shapes of the different distributions and also includes the correlations.

Using a multivariate method means to have a number of input variables which are trained with a special method – as, for example, a multilayer perceptron or a boosted decision tree in this analysis. A classification using such a method combines the input variables into one output variable which is sensitive to all the differences between the processes. In this analysis the TMVA framework is used [33].

6.1.1. Multilayer Perceptron

The first method described is the multilayer perceptron (MLP) [33], an implementation of a neural network.

Network composition A neural network is determined by the layout of the corresponding neurons, the weights for the inter-neuron connections and the response of the neurons to the input using the neuron response function ρ .

Any neural network includes several layers. The first layer is the input layer, the last one the output layer. The layers in between are called hidden layers. This is illustrated in figure 6.1. According to the theorem of Weierstrass, one single hidden layer can describe all correlations, as long as the number of neurons in this hidden layer is large enough. Using n_{var} input variables – $x_1, \dots, x_{n_{\text{var}}}$ – leads to an input layer with n_{var} neurons. The neuron in the output layer is the output variable, the neural net estimator y_{ANN} .

The neuron response function ρ maps the input neurons onto the output neurons by assigning a weight to each neuron, i.e. in the first layer to each input variable. This function is separable into a synapse function κ , which combines the input variables and their weights into a single quantity, and a neuron activation function α , which transforms the result of the synapse function, using $\rho = \alpha \circ \kappa$.

$$\kappa : (y, w) \rightarrow w_{0j} + \sum_{i=1}^n y_i \cdot w_{ij} \quad (16)$$

$$\alpha : x \rightarrow \begin{cases} x & \text{Linear} \\ \frac{1}{1+\exp(-kx)} & \text{Sigmoid} \\ \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)} & \text{Tanh} \\ \exp(-x^2/2) & \text{Radial} \end{cases} \quad (17)$$

Per default, the tanh activation function is used. During the following analysis, other activation functions were analysed in addition for a simple case using only two input variables.

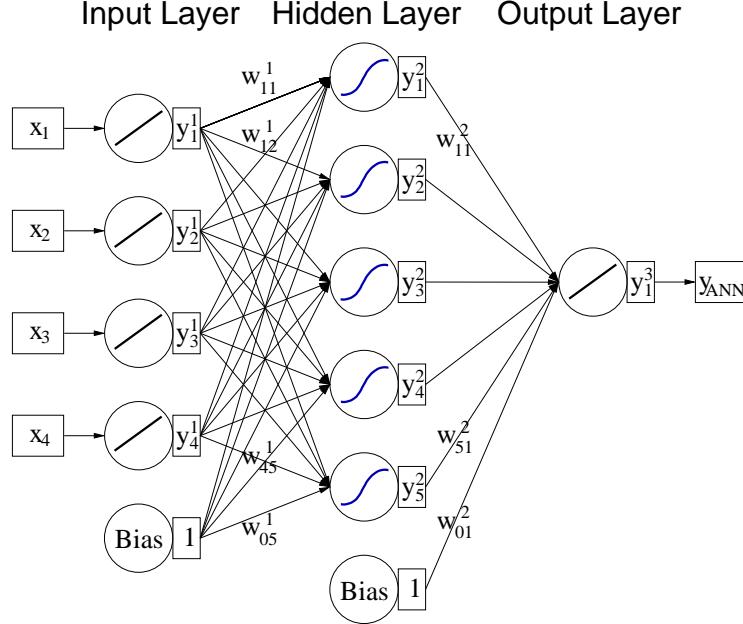


Figure 6.1: Illustration of a multilayer perceptron (MLP) with four input variables and one output variable. For the MLP one hidden layer with five nodes is used. For both the input and the hidden layer a bias node is used in addition. [33]

Back-propagation The back-propagation algorithm is the most commonly used algorithm for adapting the weights. This technique belongs to the supervised learning methods, this means that the desired output is known for every input event during the training. The output of a network is given by

$$y_{\text{ANN}} = \sum_{j=1}^{n_h} y_j \cdot w_{j1}, \quad (18)$$

where y_j is calculated using the neuron response function as described above. w_{j1} is the weight between the neuron j of the last hidden layer and the output neuron.

The training data set consists of N events. For each event the neural network output y_{ANN} is calculated and compared to the desired output \hat{y} . The desired value is 1 for signal and 0 for background events. In the end, the neural network output differs from these values, therefore the neural network distribution is normally between -0.1 and 1.2 . In cases with less separation power between signal and background, the distribution can be smaller. Using the information for the desired output value, the error function E is calculated

$$E(x_1, \dots, x_N | w_i) = \frac{1}{2} \cdot \sum_{i=1}^N (y_{\text{ANN}, i} - \hat{y}_i)^2. \quad (19)$$

The error function is minimised by changing the weights w_i . For this, one starts with a random set of weights and moves into the direction of $-\nabla_w E$ by a small distance in the weight-space where the error function decreases strongly

$$w^{(\rho+1)} = w^{(\rho)} - \eta \cdot \nabla_w E \quad (20)$$

with the learning rate η . The weights for the output layer change in the following way

$$\Delta w_{j1} = -\eta \cdot \sum_{i=1}^N (y_{\text{ANN},i} - \hat{y}_i) \cdot y_{j,i}, \quad (21)$$

and the weights connected with the hidden layer are updated by

$$\Delta w_{kj} = -\eta \cdot \sum_{i=1}^N (y_{\text{ANN},i} - \hat{y}_i) \cdot y_{j,i} \cdot (1 - y_{j,i}) \cdot w_{j1} \cdot x_{k,i}. \quad (22)$$

These new weights are used in the following training step.

Ranking For the MLP a ranking of the variables is given. The importance I_i of each input variable i ($i = 1, \dots, n_{\text{var}}$) is defined as

$$I_i = \bar{x}_i^2 \cdot \sum_{j=1}^{n_h} (w_{ij})^2 \quad (23)$$

where w_{ij} is the weight between the input variable and all neurons of the hidden layer. \bar{x}_i is the sample mean of the input variable i , therefore the importance is higher for variables with higher values in the distribution. This is the reason why the importance I_i is mostly only a hint for the relevance of each variable. A better solution for this would be to redefine the variables. For example one could redefine the variable in such a way that it is situated between zero and one in order to reduce the influence of the mean value for this ranking.

6.1.2. Boosted Decision Tree

A further classification method is the so called boosted decision tree (BDT) [33].

Composition A BDT is a binary tree structured classifier, using yes/no decisions on one single variable until some stop criterion is fulfilled. As illustrated in figure 6.2 there is one root node for the start of the tree. After this, a cut at the variable with the highest separation power is done and therefore the sample is splitted into two smaller sets and two nodes for the following analysis. Other cuts for all variables are applied afterwards until the events are separated into leaf nodes where most events belong to either the signal or the background.

In order to calculate the highest separation power, there are different possibilities. The most frequently used method, which is also used in the following analysis, is the so called Gini Index. For this, the purity p of each sample after the splitting is calculated and the variable for which $p \cdot (1 - p)$ is maximal is then used.

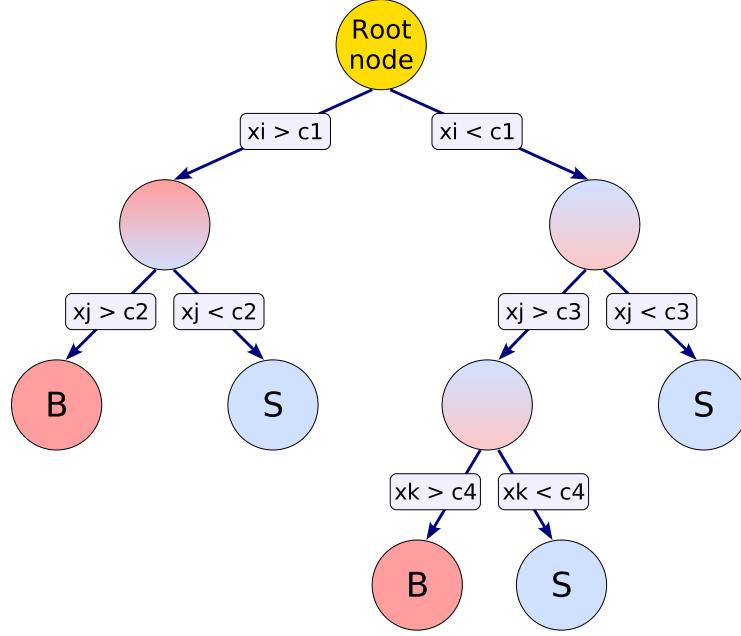


Figure 6.2: Illustration of a decision tree. One starts at the root node and makes binary decisions to obtain leaf nodes with mostly signal (S) or background (B) inside. [33]

To prevent the training on statistical fluctuations due to some cuts, not a single decision tree is used, but many trees are combined in order to construct a forest. All trees are constructed from the same training sample.

To avoid overtraining, i.e. training on statistical fluctuations of the data sets, the depth of the trees is limited. This can be configured by the maximal number of nodes in a tree, or the minimal number of events in a leaf node.

Boosting The boosting methods are used in order to enhance the classification performance and to increase the stability due to statistical fluctuations in the training sample. The multivariate method is applied to each training sample several times, but after each training the events are reweighted (boosted).

The most frequently used boosting method is the Ada Boost (adaptive boost). Higher event weights for the next training are assigned to misclassified events from a previous training. Starting with the original event weight for the first decision tree, the next tree is trained using modified event weights including a boost weight α . This weight is derived from the misclassification rate err

$$\alpha = \frac{1 - err}{err} \quad (24)$$

After this, the sample is renormalised to keep the sum of weights constant. The result of an individual decision tree is called $h(x)$ (x is a tuple of input variables). This function is

trained to be around 1 (-1) for signal (background) events and is needed for the boosted event classification

$$y_{\text{Boost}} = \frac{1}{N} \cdot \sum_{i=1}^N \ln(\alpha_i) \cdot h_i(x) \quad (25)$$

Small (large) values for y_{Boost} indicate a background (signal) like event.

Ranking The ranking of a variable in the BDT is proportional to the number of splitting decisions for which the variable is used. For this, each split is weighted by the separation gain and by the number of events in the node.

6.2. Optimising the neural networks

The optimisation of the neural networks is done separately for each region. For all regions the optimisation procedure was the same. As described before, both methods – MLP and BDT – are used in parallel in order to compare their performance. During the optimisation it turned out, that the BDT has more serious problems with the overtraining of samples with lower statistics, which is explained shortly. For this reason, the MLP was finally used as standard technique for the optimisation procedure.

There are different possibilities to check the performance of a neural network – like overtraining and convergence. The possible performance of a multivariate method is limited by the convergence of the training and by overtraining. Convergence means that a stable minimum of the error function has to be found for both the training and the testing sample. Large differences between the two samples can be a sign of overtraining. This means that the weights are adapted to statistical fluctuations of the training sample, which are not part of the real distribution. To detect overtraining, the output distributions for the training and testing sample can be compared in addition. An example for such a distribution is shown in figure 6.3a. There, the training and testing sample are displayed and a comparison between these two samples is possible. The Kolmogorov-Smirnov probability is given in order to decide the goodness of the overtraining test. The Kolmogorov-Smirnov test determines whether two distributions – the training and the testing distribution – differ significantly. For this, the test calculates the difference of the cumulative distributions of the test and the training sample and gives the probability for them being similar. If the given probability value is near to one (zero) the test and the training sample are similar (different).

Figure 6.3b shows an example for a convergence test of the MLP. On the x-axis the number of epochs is plotted, i.e. the number of repeated steps for the optimisation. Again, one compares the results for the training and testing sample which should converge to a similar value at the end of the optimisation.

When test and training sample do not converge or the overtraining test shows discrepancies, the optimisation has trained to specific features – statistical fluctuations – in the training sample. This limits the performance on other data, which do not have the same fluctuations. Therefore a setup is searched where the convergence and the overtraining test do not show any discrepancies.

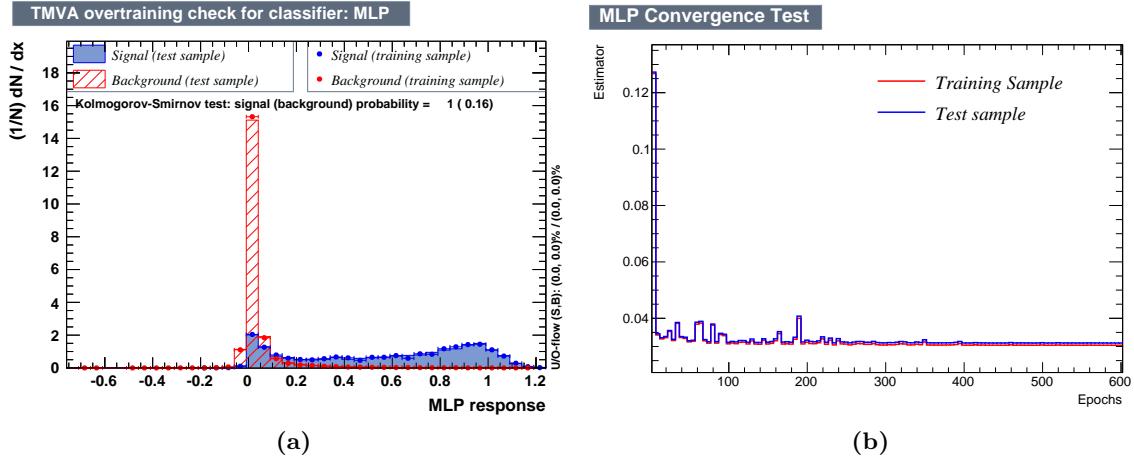


Figure 6.3: An example for an overtraining test distribution (a) and a convergence test (b) is plotted here for illustration.

Apart from these technical checks for the optimisation, the receiver operating characteristic (ROC) curve and the significance after a simple cut for the neural network output distribution are also analysed.

The ROC curve, as illustrated in figure 6.4, shows the signal efficiency versus the background rejection rate. In the best case the signal efficiency and the background rejection are nearly at one, this means the curve should be as high as possible in the top right corner. In the illustration one can see the two different methods – MLP and BDT – which were used for the training. Both methods show a comparable performance.

During the evaluation of the two methods in the analysis, the MLP generally showed a more stable behaviour, regarding the overtraining and the convergence tests. Therefore this method was chosen as the multivariate technique for this analysis.

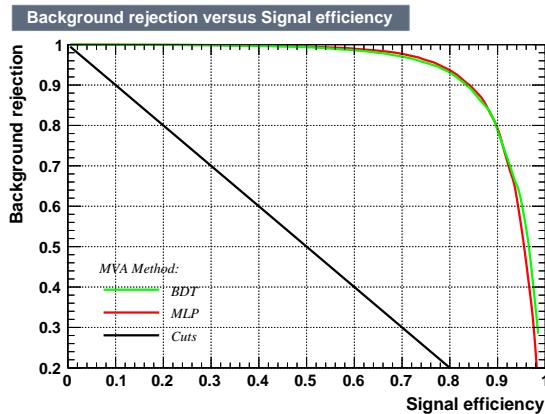


Figure 6.4: Illustration of a ROC curve as a comparison between the MLP, the BDT and a Cut method.

The neural network output distribution is used for the calculation of the significance Z after the training. For this, a cut is done at this distribution and the number of survived signal and background events is calculated. With the expected number of signal (N_S) and the expected number of background (N_B) events the significance Z is given by

$$Z = \frac{N_S}{\sqrt{N_S + N_B + \sigma_B^2}}. \quad (26)$$

The additional term σ_B is a simple consideration of a systematic error of the background. For the diagonal region $\sigma_B = 0.15 \cdot N_B$ is used, for the high mass and the bulk region $\sigma_B = 0.25 \cdot N_B$ is assumed as done in [13, 34].

The significance is compared for different cut thresholds at the output distribution. This is shown in figure 6.5. There one can see the signal efficiency and the background rejection, as well as two significance definitions depending on the output value. One of the two significances is the one defined in formula 26 using the additional σ_{Bkg} term, the other significance assumes $\sigma_{Bkg} = 0$. Both distributions have their maximum at nearly the same output value and do not differ substantially in their shape. The maximum of the significances is used for the comparison with other configurations of the neural network.

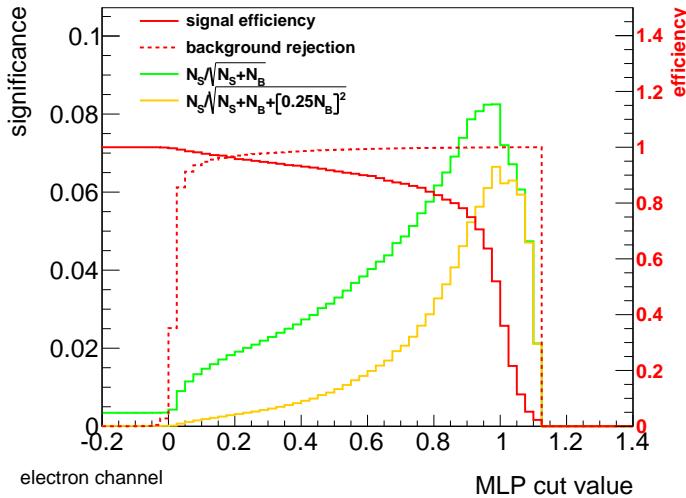


Figure 6.5: The signal efficiency (red) and the background rejection (red dashed) are shown. In addition, the significance Z (yellow) as defined in the text is plotted as well as the significance without the additional σ_{Bkg} term (green).

The optimal settings for the neural network, e.g. the choice of input variables, are determined by maximising the significance and reducing any overtraining.

Strongly correlated variables mostly do not give more information for the training, therefore it is possible to keep only one of the correlated variables. This leads to a better performance of the neural network without changing the separation between signal and background. In addition, variables can be decorrelated by taking their ratio. This helps for the BDT method which has more problems with correlations, but does not strongly affect the MLP method.

There are also some variables which are correlated for the signal sample, but not for the background sample, or vice versa. These variables can help a lot in the separation, even if the shape is not clearly different for signal and background.

Other items for the optimisation are changes in the setting of the neural network. There were some tests done for the MLP settings. During the optimisation, different numbers of training steps for the neural network are analysed and set to such a value that both the training and the testing sample converge.

As described in chapter 6.1.1, there are different activation functions available for the neural network. The performances of these activation functions were analysed. For this the optimisation was rerun for each activation function and the outputs were compared. A graphical comparison is only possible for two input variables, by a comparison of the output distribution, the convergence test and the boundary layers. The reasonability of the boundary layers can be determined for two variables by comparing the results to the correlation for signal and background (cf. figure 6.6).

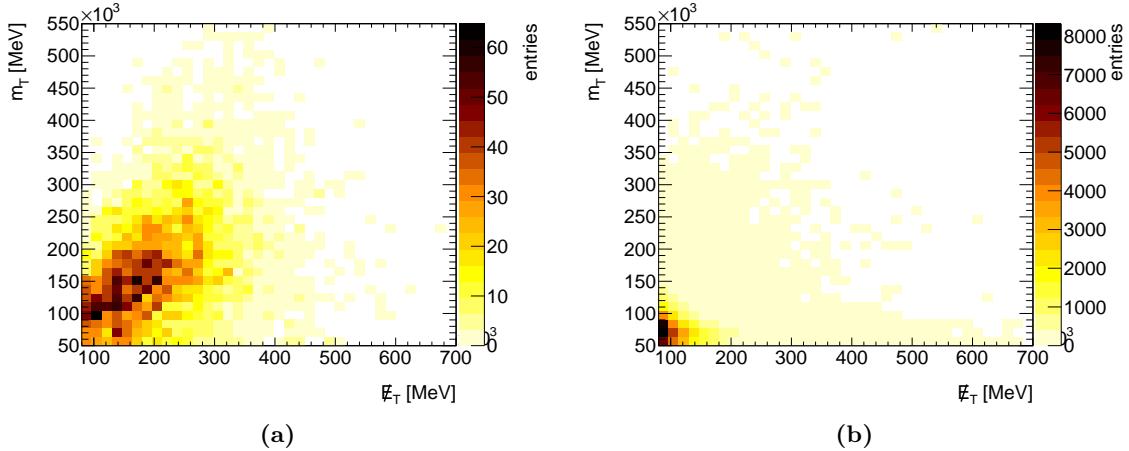


Figure 6.6: Distribution of the missing transverse energy \cancel{E}_T and the transverse mass m_T for a signal sample (a) in the bulk region and in comparison for $t\bar{t}$ and $W+jets$ background (b).

Boundary layers illustrate the cuts which are chosen by a neural network. Each boundary layer shows the decision of a neural network for a given signal efficiency and background rejection. In figure 6.7a the boundary layers for the linear activation function are shown. Due to the linear activation function only linear cuts are possible. Therefore the boundary layers in the plot are straight lines.

The boundary layers at $\cancel{E}_T \simeq 180$ GeV and $m_T = 120$ GeV show a good cut possibility to reject most background without loosing too much signal. Using the linear activation function there is no possibility to reject the background events for higher missing transverse energy values without loosing too much signal. Therefore this activation function is a poor choice, especially for higher dimensional problems as used in the following study.

In figure 6.7b the boundary layers for a radial activation function are shown. This configuration is more powerful because it is not bounded to simple lines. Therefore the separation

of the neural network is improved for a broader range of possible cut values. This gives a better signal efficiency and a higher background rejection.

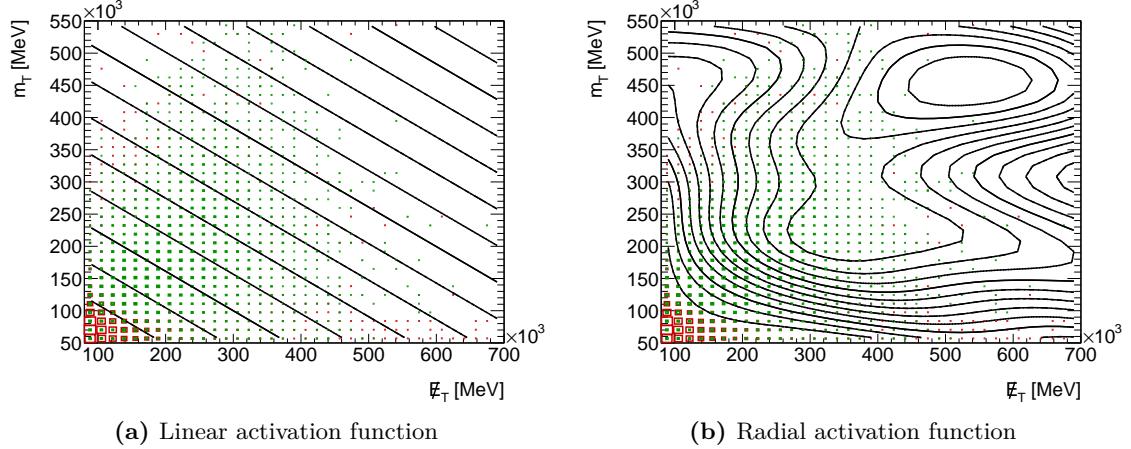


Figure 6.7: Graphical illustration of the boundary layers for a neural network with two input variables using different activation functions. As input variables the missing transverse energy \cancel{E}_T and the transverse mass m_T are used. The black lines show the boundary of every decision, the green (red) boxes show the correlation of the two variables for signal (background).

Similar behaviour of the sigmoid and the tanh activation functions is visible in figure 6.8a and figure 6.8b. Due to the similar boundary layers it is not important which activation function – radial, sigmoid or tanh – is used. For these three activation functions the neural network output distributions also look similar and the calculated significances do not differ significantly. For the linear activation function the neural network output distribution does not show a good separation between signal and background. The convergence test works fine for all activation functions.

The study of the different MLP activation functions shows that the linear activation function is the only one to differ significantly from the others and shows a decreased performance. The other activation functions give similar results. In the following study the tanh activation function is used.

Further extensions to the optimisation procedure are possible, which could improve the performance of the neural network. They are not taken into account for the following results, but require a further, dedicated analysis. In addition to the described procedure, one could require a minimal number of expected events after the cut on the neural network distribution. With this requirement, the optimisation does not concentrate on too few events and does not train on specific events. With this extension, the final neural network distribution will include slightly higher statistics in the signal-like region. This could help in the following analysis, i.e. the limit setting especially in the high mass region.

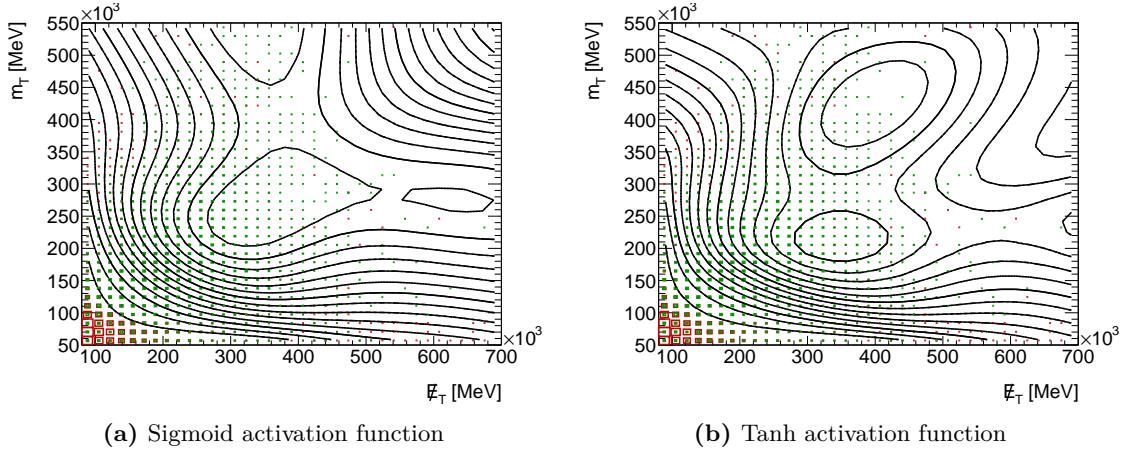


Figure 6.8: Graphical illustration of the boundary layers for a neural network with two input variables using different activation functions. As input variables the missing transverse energy \cancel{E}_T and the transverse mass m_T are used. The black lines show the boundary of every decision, the green (red) boxes show the correlation of the two variables for signal (background).

6.3. Results for the three regions

In the following, all results from the optimisation are shown for the three different signal regions. For this, the bulk region is explained in detail and for the diagonal and the high mass region only the particular details of the regions are shown in addition to the results. The optimisation is done for the electron and muon channel separately, as different identification and detection efficiencies for the lepton types can have an impact on the event kinematics. The shown results are only for the electron channel, as they are similar to the muon channel, which are only shown in the appendix A.4.

The training for the neural network is only done using the two main backgrounds – $t\bar{t}$ and $W+\text{jets}$. The neural network is also evaluated for the other backgrounds. They look similar to the main backgrounds, the corresponding distributions are shown in the appendix A.5.

6.3.1. Bulk region

For the optimisation of each region, a combination of various signal samples is used. Therefore the optimisation does not train to a specific mass point and the statistics for the signal is increased. For the bulk region the samples listed in table 6.1 are used. All significances are evaluated separately for each sample. Key indicators for the optimisation are the significances, while keeping the convergence and the overtraining stable. Changes which lead to a decrease in the significance for one signal are kept, if they perform better for the other signals. Due to slightly different behaviours of the different stop samples, sometimes the significances for a sample decreases when most other samples improve. This effect is visible especially for the samples with the lowest or highest stop masses.

Table 6.1: Signal samples which are used for the neural network optimisation for the bulk region with a center of mass energy of 8 TeV.

$m_{\tilde{t}} \text{ [GeV]}$	$m_{\tilde{\chi}_1^0} \text{ [GeV]}$
500	100
450	50
500	200

In table 6.2 the variables showing the best performance are ranked by the importance for the MLP. The ratio between $\Delta\varphi(t_{lep}, t_{had})$ and the missing transverse energy \cancel{E}_T leads to a slightly more stable performance of the neural network.

As expected, the transverse mass m_T and the missing transverse energy \cancel{E}_T are important variables in the neural network as they specify the main differences between signal and background – the additional neutralinos. In addition, the mass of the hadronic top quark and both transverse masses – am_{T2} and m_{T2}^τ – show a good separation power. Also the perpendicular \cancel{E}_T component to the leptonic top quark is used in the neural network. Other variables – like topness or some angular distributions – are used in the neural network in order to avoid overtraining. Their separation power is less strong compared to the other variables.

Table 6.2: List of variables showing the best performance in the neural network for the bulk region, ranked depending on their importance for the MLP.

Rank	Variable	Importance
1	m_T	41.84
2	$m_{t_{had}}$	29.57
3	am_{T2}	11.46
4	\cancel{E}_T	9.812
5	m_{T2}^τ	6.11
6	$\cancel{E}_T \perp t_{lep}$	4.515
7	$\cancel{E}_T / \sqrt{H_T}$	2.893
8	$\Delta\varphi(t_{lep}, t_{had}) / \cancel{E}_T$	0.1643
9	$\Delta\varphi(\cancel{E}_T, j_2)$	0.1287
10	topness	0.02153
11	$\min \left\{ \Delta\varphi(\cancel{E}_T, j_{1/2}) \right\}$	0.000229

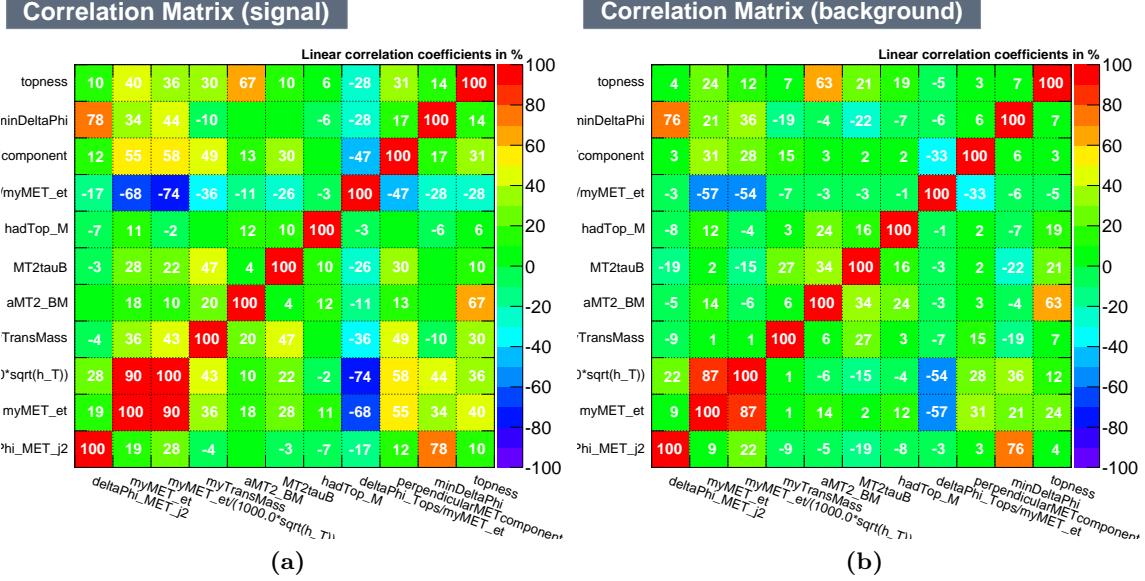


Figure 6.9: Linear correlation coefficients for the remaining variables after the optimisation. The linear correlation coefficients are shown for both signal (a) and background (b).

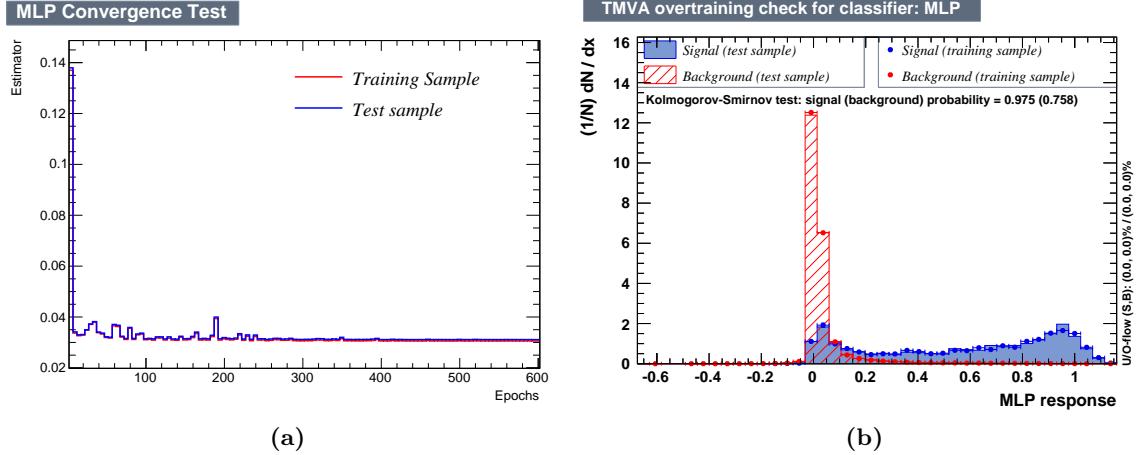


Figure 6.10: Convergence and overtraining check for the MLP configuration of the bulk region. The estimator for the test and the training sample are shown as a function of the number of epochs (a). The MLP distributions are shown for the test and the training sample (b).

The linear correlation coefficients of these variables are shown in figure 6.9 for signal and background separately. As one can see, there are strongly correlated variables – e.g. \cancel{E}_T and $\cancel{E}_T/\sqrt{H_T}$ – for signal and background. Some of these correlations are also similar for signal and background. Nonetheless, these variables improve the performance of the neural

network. A neural network without these variables was tested, but showed no differences in the significance. Considering the performance indicators of the neural network – like overtraining and convergence – these variables provide an increased stability.

While most variables are not correlated at all, there are some variables – e.g. m_T and $m_{T2}^{\tau\tau}$ – which are stronger correlated for the signal than for the background. Especially these variables improve the significances because they help to differ between signal and background due to the different correlations.

In figure 6.10a one can see the convergence of the training and the testing sample where both samples do not differ. The overtraining distribution is shown in figure 6.10b. One can see a good separation between signal and background. The Kolmogorov-Smirnov test gives a probability of 0.975 for the signal distributions and of 0.785 for the background distributions. Therefore a good convergence and no overtraining is seen for the neural network.

In figure 6.11, the signal efficiency and the background rejection for the central signal sample are shown. Additionally, both significances are included in the plot. As already shown before, the signal efficiency decreases for higher MLP values and the background rejection increases. Therefore it is important to find the optimal point of high background rejection without loosing too much signal. Good criteria for this are both significances. Their maxima are located nearly at the same point and therefore they would give nearly the same cut value for the MLP distribution.

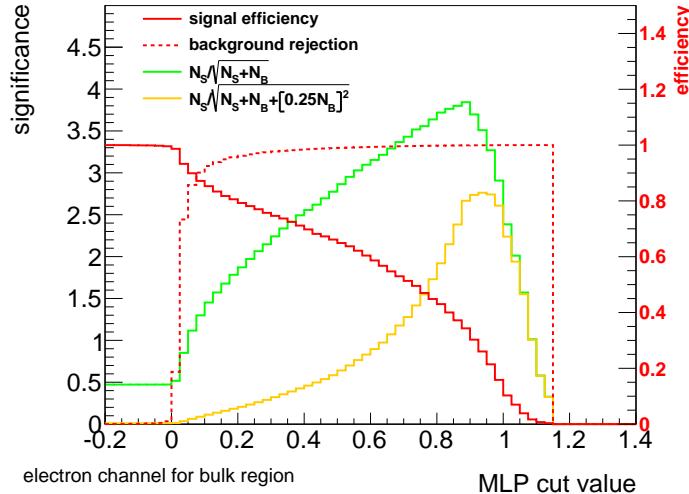


Figure 6.11: The signal efficiency (red) and the background rejection (red dashed) depending on the MLP value are shown for the central signal sample ($\tilde{t}(500, 100)$). In addition, the significance Z (yellow) as defined in the text and the significance without the additional σ_{Bkg} term (green) are plotted.

The results for the optimised neural network are compared to other methods. For this, the significance for simple cuts – as used in the official analysis [34] – is calculated for each signal point. The cuts for all three regions are listed in the appendix A.2. In table 6.3 the calculated significances for these cuts using the old variables at the optimised points

from the official analysis are listed in the column “cuts”. These values are compared to a neural network using only the old variables without any optimisation – column “MLP old variables”. For the calculation of the significance formula 26 with the additional σ_{Bkg} term is used. Using this setup it can happen that there is overtraining or a bad convergence, therefore it is a simplified comparison. Comparing the numbers from the first two columns, one can roughly see the improvement of using a neural network. For the bulk region, this improvement accounts for around 20%.

The last column – “optimised MLP” – consists of the calculated significances after the full optimisation and selection of variables. This gives an improvement by up to 50%.

Table 6.3: Significances for the old cuts (cf. appendix A.2), using all old variables and using the optimised neural network.

Sample $(m_{\tilde{t}}, m_{\tilde{\chi}_1^0})$	cuts	MLP old variables	optimised MLP
$\tilde{t}(500, 100)$	2.13	2.18	2.76
$\tilde{t}(450, 50)$	2.65	3.08	3.83
$\tilde{t}(500, 200)$	1.24	1.49	1.88

As shown in table 6.3, the usage of an optimised neural network gives a strong improvement for the bulk region. The differences between simple cuts at the old variables and cuts at the neural network distribution and the usage of the full distribution of the neural network are compared in detail for the limit setting in chapter 8.2.

6.3.2. Diagonal region

The main problem in the diagonal region is the similar kinematics of the signal and the different backgrounds. In this region, the mass difference between the stop and the neutralino is nearly the top mass. Due to this small mass difference, the outgoing particles have only low momentum and therefore the missing transverse energy is also smaller. This results in the signal looking very similar to the $t\bar{t}$ background.

In order to improve the separation, the samples for the optimisation were splitted into two orthogonal samples. One of these samples has an additional jet with $p_T > 150 \text{ GeV}$ (initial state radiation jet – ISR jet) and the other sample has no additional jet or an additional jet with low transverse momentum. This is sensitive to events in which the $\tilde{t}\tilde{t}^*$ system is boosted by the additional ISR jet. In this case, the missing transverse energy should be more pronounced, which is shown in figure 6.12. The differences in the shape for the missing transverse energy are clearly visible.

The distributions of other variables, like the transverse momentum and the mass of the $t\bar{t}$ pair or the stransverse masses, also change their shapes when the events include an additional jet. On the other hand most angular distributions are not affected. All plots with or without an additional ISR jet are shown in the appendix A.3.

In a first test, both regions were optimised independently. The low statistics of the sample with an additional jet caused problems with the convergence of the training and lead to strong overtraining of the neural networks. No gain in significance was achieved for the combination of both regions, as the sample including an ISR jet could not be optimised.

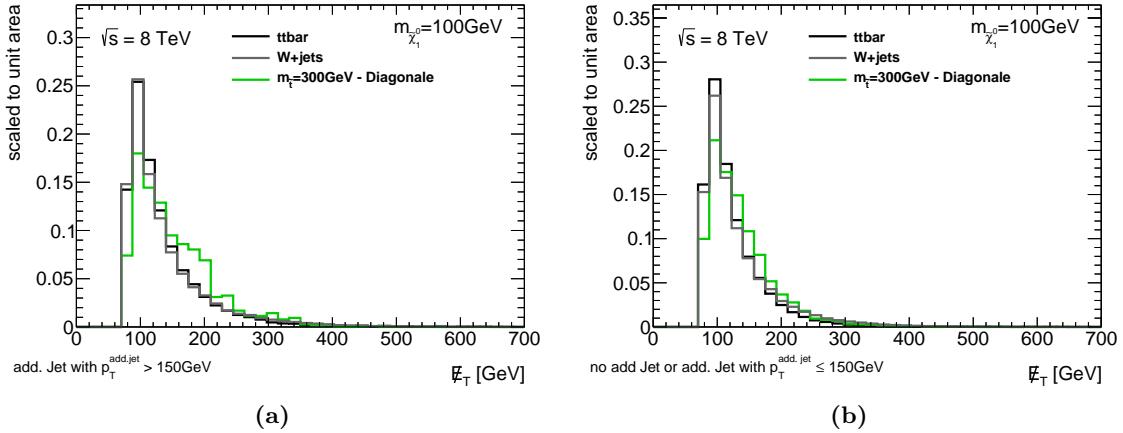


Figure 6.12: The missing transverse energy \cancel{E}_T for the diagonal region with an additional jet with $p_T > 150$ GeV (a) and without an additional jet (b).

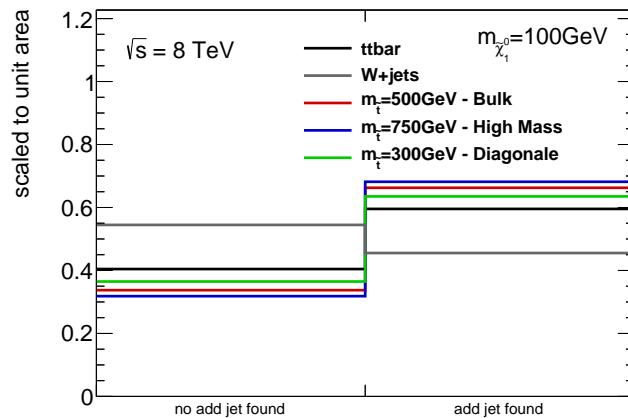


Figure 6.13: Fraction of events with and without an additional jet for the SM backgrounds and for the diagonal sample.

In the training of the neural network for the diagonal region, the information whether the event includes an additional ISR jet is used instead of using the individual regions. As shown in figure 6.13, the fraction of events in the two categories is different for the backgrounds, especially for the $W+jets$ sample. For the electron channel no clear improvement was found and the variable was excluded. For the muon channel the variable is used, as it helps to stabilise the performance of the neural network.

Results for the diagonal region In the following, the optimisation results for the diagonal region are shown. For this optimisation more samples are used in order to train for a wider range at the boundary of the $\tilde{t} \rightarrow t\tilde{\chi}_0^1$ decay. The used signal samples are listed in table 6.4.

Table 6.4: Signal samples which are used for the neural network optimisation for the diagonal region with a center of mass energy of 8 TeV.

$m_{\tilde{t}}$ [GeV]	$m_{\tilde{\chi}_1^0}$ [GeV]
300	100
250	50
350	150
200	1
400	200
450	250

The variables listed in table 6.5 show the best performance during the optimisation. Similar variables as for the bulk region are used for the diagonal region. Again, the best variable is the transverse mass m_T . Directly after this, the tau-veto shows the best separation power which is not obvious because the visible differences between signal and background are small, including also the correlation to other variables. After this, again the missing transverse energy and the corresponding significance improve the separation. From the new variables, the perpendicular component of the missing transverse energy \cancel{E}_T gives the best improvement in the significances. In addition, the mass of the hadronic top quark and the stransverse mass m_{T2}^τ is used. The variables with the lowest rank are again topness and several angular distributions.

Table 6.5: List of variables showing the best performance in the neural network for the diagonal region, ranked depending on their importance for the MLP.

Rank	Variable	Importance
1	m_T	20.94
2	tau-veto	10.31
3	$\cancel{E}_T / \sqrt{H_T}$	0.4609
4	\cancel{E}_T	0.2854
5	$\cancel{E}_T \perp t_{lep}$	0.0253
6	$m_{t_{had}}$	0.02002
7	m_{T2}^τ	0.0189
8	topness	0.0099
9	$\Delta\varphi(\cancel{E}_T, t_{had})$	0.008425
10	$\Delta\varphi(t_{lep}, t_{had})$	0.004186
11	$\min \left\{ \Delta\varphi(\cancel{E}_T, j_{1/2}) \right\}$	0.0002066

In figure 6.14 the linear correlation coefficients are shown for the variables given in table 6.5. As already shown for the bulk region there are strongly correlated variables – like \cancel{E}_T and $\cancel{E}_T / \sqrt{H_T}$ – which however help for the separation of signal and background. In

addition, there are variables – like m_T and $\Delta\varphi(\cancel{E}_T, t_{had})$ – which are stronger correlated for the signal and less correlated for the background. These variables help in the separation because they provide different information for signal and background.

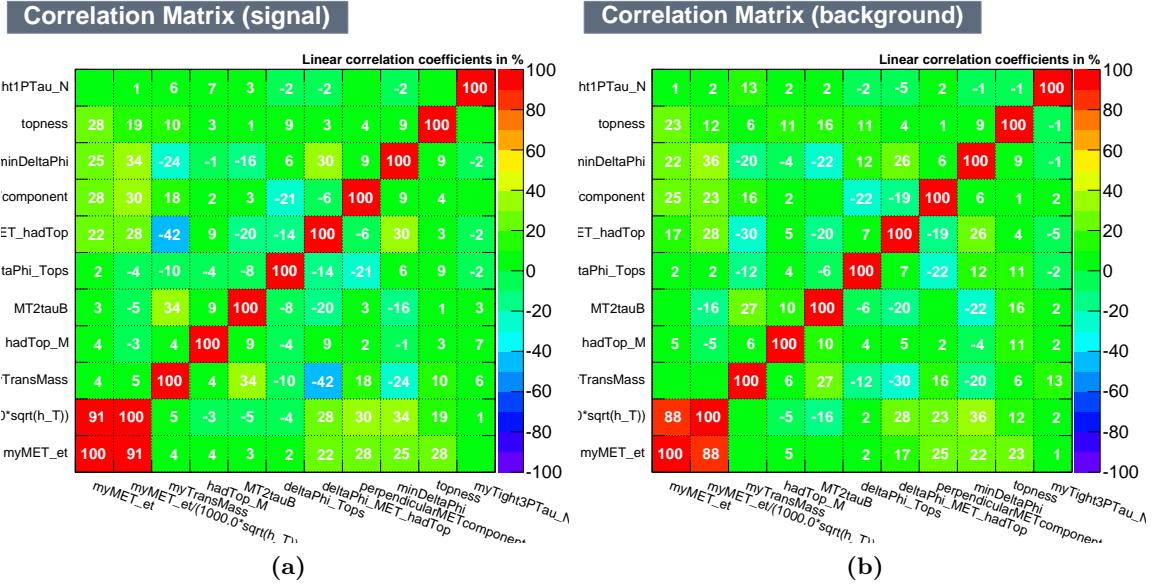


Figure 6.14: Linear correlation coefficients for the remaining variables after the optimisation. The linear correlation coefficients are shown for both signal (a) and background (b).

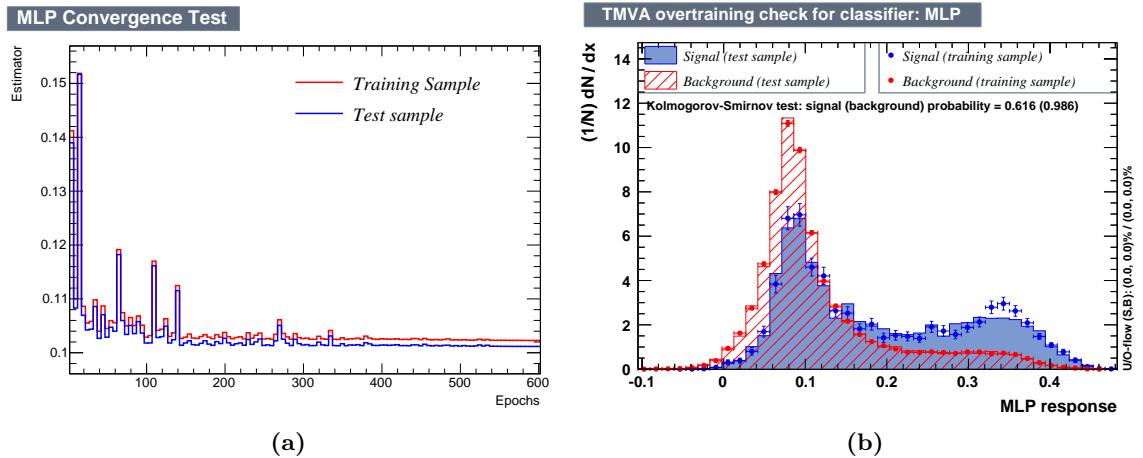


Figure 6.15: Convergence and overtraining check for the MLP configuration of the diagonal region. The estimator for the test and the training sample are shown as a function of the number of epochs (a). The MLP distributions are shown for the test and the training sample (b).

The convergence test, which is shown in figure 6.15a, shows a good convergence for the optimised neural network. In figure 6.15b the overtraining check for the diagonal region is shown. It is visible that the separation between signal and background is not perfect. A double peak structure can be observed for the signal, with a large background like fraction. This is due to the similarity between signal and background in the diagonal region. For the high neural network output values a good separation between signal and background can be seen.

The Kolmogorov-Smirnov probability for the signal distributions is 0.616. Minor differences can be observed in the peak region of the signal. The separation for the test sample is worse than for the training sample, but the overall difference is small. Due to the small differences between signal and background, the training was hard to stabilise in the diagonal region. For the background a good agreement can be seen, with a Kolmogorov-Smirnov probability of 0.986.

The signal efficiency and the background rejection are shown in figure 6.16. They both behave similar to the bulk region. The maximum of both significances are not at the same neural network value, but they show a similar behaviour for the cut selection. The effect of the additional σ_{Bkg} term is stronger in the diagonal region compared to the bulk region, because the ratio of the number of expected events for the background N_B is much higher than the number of expected events for the signal N_S .

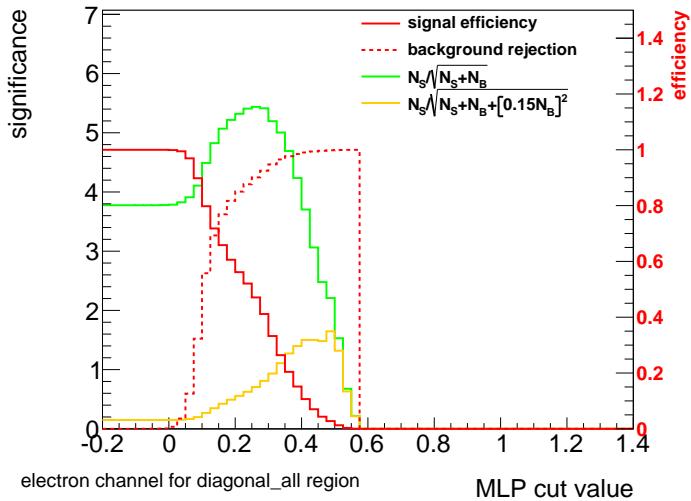


Figure 6.16: The signal efficiency (red) and the background rejection (red dashed) depending on the MLP value are shown for the central signal sample ($\tilde{t}(300, 100)$). In addition, the significance Z (yellow) as defined in the text as well as the significance without the additional σ_{Bkg} term (green) are plotted.

In table 6.6 the significances for the three different methods are shown as already explained for the bulk region. As in the bulk region, an improvement of the significances can be seen when using a neural network compared to the simple cuts. For the stop samples $\tilde{t}(200, 1)$ the significance value does not change. This sample has the lowest stop and neutralino masses, and is therefore most similar to the $t\bar{t}$ background. Due to the similar kinematic

behaviour, the separation between signal and background is difficult and the neural network, which is trained with all listed signal samples, favours other samples in the training. In addition, this sample has the lowest expected number of events of all signals. For the training of the neural network a combination of all samples is used and therefore the samples with the highest number of events are more important for the training.

Table 6.6: Significances for the old cuts (cf. appendix A.2), using all old variables and using the optimised neural network.

Sample $(m_{\tilde{t}}, m_{\tilde{\chi}_1^0})$	cuts	MLP old variables	optimised MLP
$\tilde{t}(300, 100)$	1.34	1.65	1.77
$\tilde{t}(250, 50)$	1.4	1.43	2.47
$\tilde{t}(200, 1)$	1.11	1.11	1.71
$\tilde{t}(350, 150)$	0.95	1.17	1.04
$\tilde{t}(400, 200)$	0.6	0.82	0.65
$\tilde{t}(450, 250)$	0.34	0.41	0.37

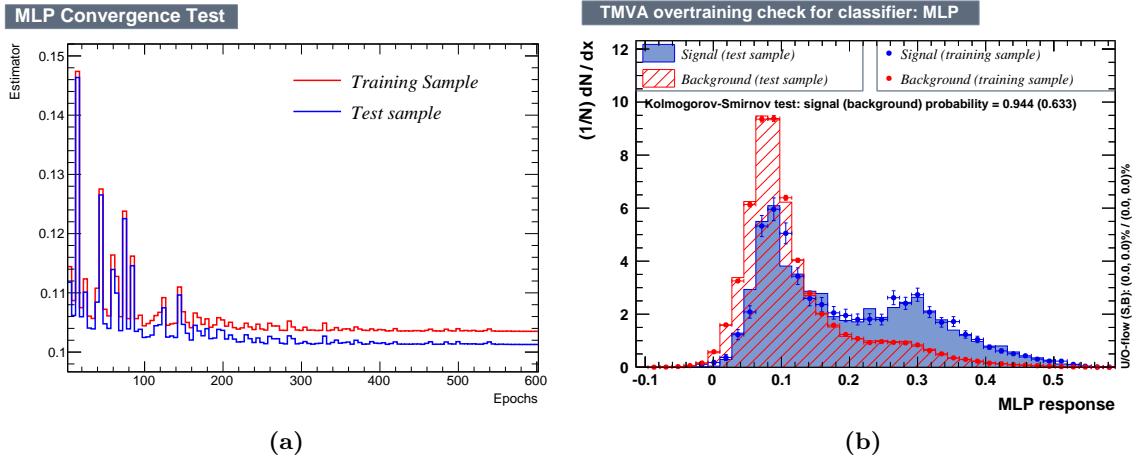


Figure 6.17: Convergence and overtraining check for the MLP configuration. The estimator for the test and the training sample is shown as a function of the number of epochs (a). The MLP distributions is shown for the test and the training sample (b).

Comparing the significances of the optimised neural network with the significances of the neural network using the old variables, a degradation can be observed for some samples. The neural network using the old variables exhibits overtraining which artificial boosts the significance. Figure 6.17a shows the convergence of the neural network using only the old variables. Here one can see a large offset between the test and the training sample, which is not visible for the optimised neural network (compare figure 6.15a). The Kolmogorov-Smirnov probability for the signal increases slightly compared to the optimised neural

network, but for the background the probability decreases (compare figure 6.15b and 6.17b). The distributions look similar for both networks.

The samples with the highest significance for the neural network with the old variables are the samples with the highest stop masses – $m_{\tilde{t}} = 350, 400, 450$. These samples are far away from the limits at the moment and the optimisation was concentrated on the first three samples with lower stop masses – $m_{\tilde{t}} = 200, 250, 300$. There the improvement of the significances is between 20% and 70%.

For the limit setting the whole distribution of the neural network output is used. The significances for simple cuts are therefore only a lower limit for the achievable separation.

6.3.3. High mass region

In the following, the results for the high mass region are shown. In table 6.7 the samples are listed which are used for the optimisation.

Table 6.7: Signal samples which are used for the neural network optimisation for the high mass region with a center of mass energy of 8 TeV.

$m_{\tilde{t}}$ [GeV]	$m_{\tilde{\chi}_1^0}$ [GeV]
750	100
750	150
750	50
700	100
800	100

Table 6.8: List of variables showing the best behaviour for the neural network for the high mass region ranked depending on their importance for the MLP.

Rank	Variable	Importance
1	tau-veto	20.2
2	m_T	15.05
3	$p_T^{t\bar{t}}/\cancel{E}_T$	12.73
4	\cancel{E}_T	11.77
5	$\cancel{E}_T/\sqrt{H_T}$	3.16
6	am_{T2}	3.074
7	$\cancel{E}_T \perp t_{lep}$	2.544
8	$m_{t\bar{t}}$	1.126
9	$\Delta\varphi(\cancel{E}_T, j_2)$	0.3691
10	topness	0.2938

Table 6.8 lists the best variables which are used after the optimisation. Looking at this table it is visible that most kinematic variables – like m_T , \cancel{E}_T , $m_{t\bar{t}}$ – are used for the high mass

region. In these variables the signal and the background differ most. Only $\Delta\varphi(\vec{E}_T, j_2)$ is used as angular distribution, which has a similar distribution for signal and background.

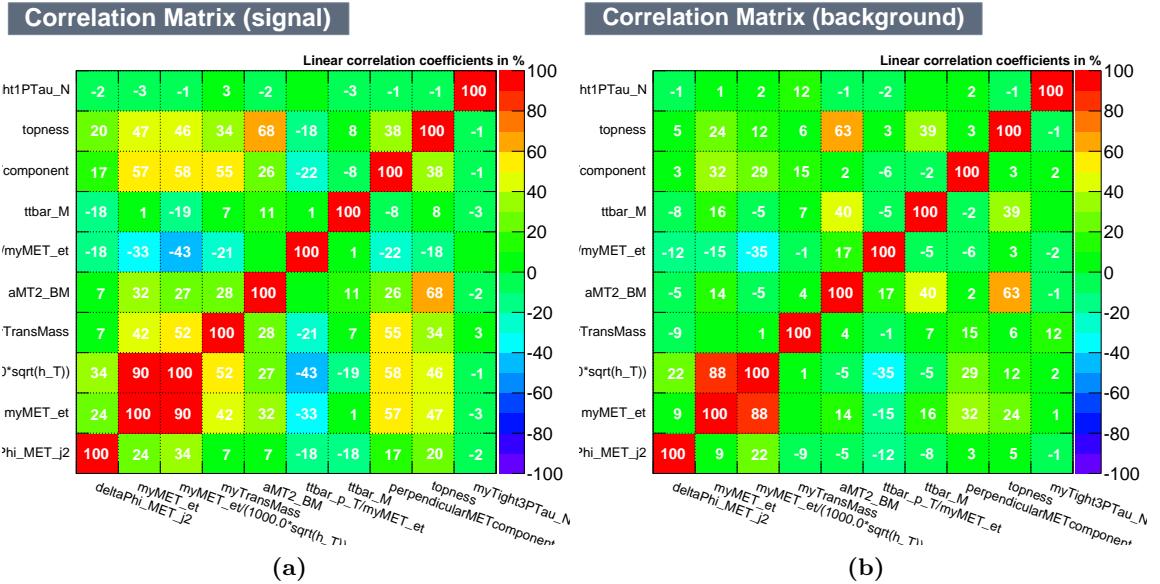


Figure 6.18: Linear correlation coefficients for the remaining variables after the optimisation. The linear correlation coefficients are shown for both signal (a) and background (b).

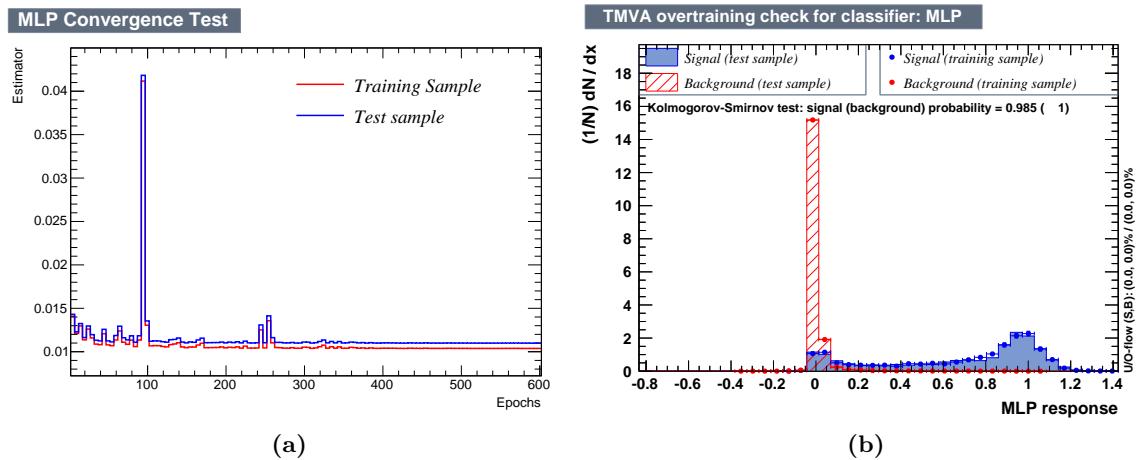


Figure 6.19: Convergence and overtraining check for the MLP configuration. The estimator for the test and the training sample is shown as a function of the number of epochs (a). The MLP distributions are shown for the test and the training sample (b).

The linear correlation coefficients for these variables are shown in figure 6.18. As already seen for the bulk and the diagonal region, the missing transverse energy and the corresponding significance are strongly correlated. Especially in the high mass region there are many variables which are stronger correlated for the signal and less correlated for the background or vice versa.

The overtraining check, which is shown in figure 6.19b, shows a clear separation between signal and background. The Kolmogorov-Smirnov probabilities show no sign of overtraining, as does the convergence test, shown in figure 6.19a.

In figure 6.20 the signal efficiency and the background rejection are shown as well as the significances. They show the same behaviour as for the bulk region, both significance maxima appear at the same neural network cut value.

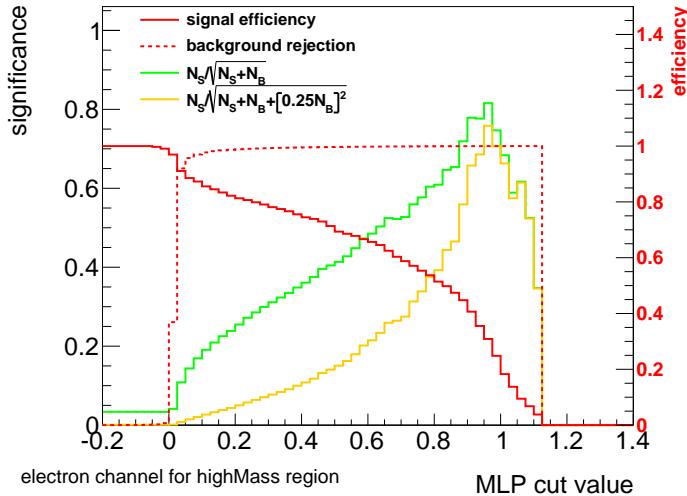


Figure 6.20: The signal efficiency (red) and the background rejection (red dashed) depending on the MLP value are shown for the central signal sample ($\tilde{t}(750, 100)$). In addition, the significance Z (yellow) as defined in the text as well as the significance without the additional σ_{Bkg} term (green) are plotted.

Table 6.9: Significances for the old cuts (cf. appendix A.2), using all old variables and using the optimised neural network.

Sample ($m_{\tilde{t}}, m_{\tilde{\chi}_1^0}$)	cuts	MLP old variables	optimised MLP
$\tilde{t}(750, 100)$	0.33	0.56	0.76
$\tilde{t}(750, 150)$	0.4	0.56	0.76
$\tilde{t}(750, 50)$	0.36	0.57	0.77
$\tilde{t}(700, 100)$	0.55	0.83	1.15
$\tilde{t}(800, 100)$	0.24	0.43	0.55

In table 6.9 the significances of the three different methods, as explicitly described for the bulk region, are listed. Comparing the calculated significances for simple cuts to the neural network using the old variables gives an improvement of around 60%. The improvement when using an optimised neural network is up to 120%. The improvement is compared again in the limit setting.

6.4. Conclusion

The shown results are only explained in detail for the electron channel, the results for the muon channel are shown in the appendix A.4. For both channels a slightly different variable selection is used for the final configuration. The differences of the two selections should be analysed in more detail, in order to further check the differences. Different detection and identification efficiencies for electrons and muons may have an impact on the event kinematics, which can be further used in neural networks.

With more Monte Carlo statistics for the diagonal region, the splitting into one region with and without an ISR jet could be repeated. The training and the performance of a neural network for the sample with an ISR jet should work more stable when using more statistics.

In this chapter, a procedure for the evaluation and optimisation of multivariate techniques is presented. Using the optimisation procedure, a neural network is trained for each analysis region respectively. The results of the optimisation for all three regions show a substantial improvement concerning the significances. The significances for a cut on the optimised neural network show an increase of 20 to 120% compared to the cuts used before.

The neural network output is used in the following analysis where the simulation is validated. In addition, the neural network distribution is used to set limits on the different stop and neutralino masses. During the limit setting, the cut procedure is again compared to the usage of a neural network (compare 8.2).

7. Validation of the simulation

In the following, the input variables which are used for the neural network optimisation are checked for the agreement between data and the Standard Model predictions. For this study, the electron and muon channels are considered separately.

7.1. First comparison of data and simulation

In figure 7.1a the comparison plot of the ATLAS data from 2012 with an integrated luminosity of 20 fb^{-1} and the Standard Model prediction for the missing transverse energy is shown.

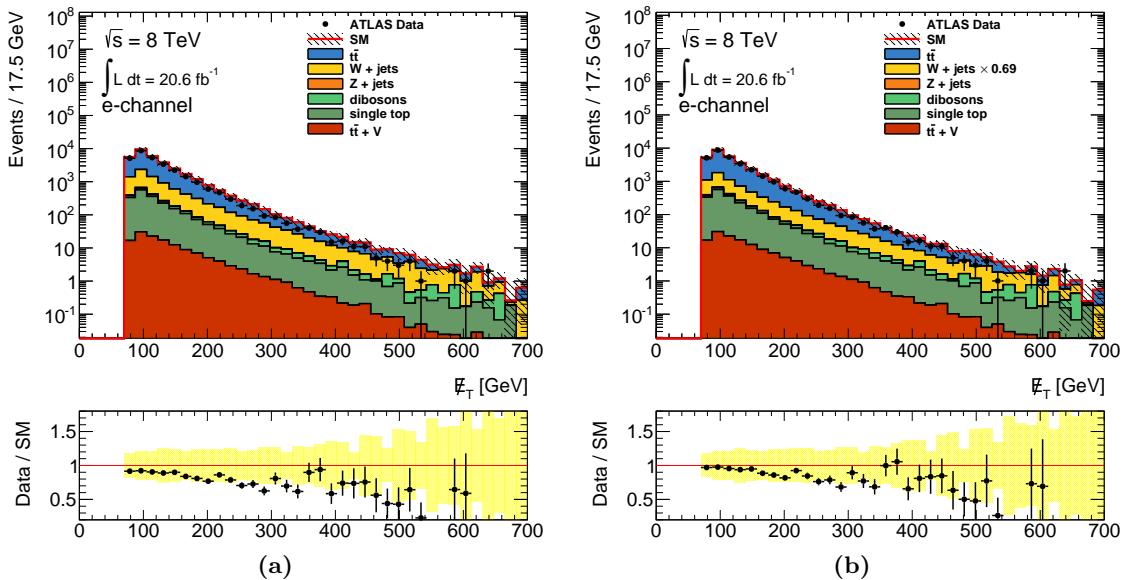


Figure 7.1: The missing transverse energy \cancel{E}_T distribution is shown for data with an integrated luminosity of 20.6 fb^{-1} and all Standard Model backgrounds stacked and scaled to this luminosity. The error bands show the statistical error and the systematic error due to the JES and JER uncertainties. The lower part of each plot shows the ratio of the data to the Standard Model prediction. The distribution is shown without any normalisation factor (a) and the normalisation factor for $W+\text{jets}$ of 0.69 provided by the official ATLAS analysis (b).

For these types of plots all backgrounds are combined into a stack and compared with the observed data. For the comparison, the ratio of the number of events for the data and for the simulation is plotted as well. Considering the ratio plot for this figure, differences between the measured data and the Standard Model prediction are clearly visible. This behaviour is compatible with the findings of the ATLAS analysis [13, 14] and is caused by a mismodelling of the normalisation of the $t\bar{t}$ and the $W+\text{jets}$ processes. To correct this, normalisation factors for these two processes were determined. The differences between the data and the Standard Model prediction vanish mostly when using a normalisation factor

for $W+\text{jets}$ of 0.69. This is the calculated value in the official ATLAS analysis [13, 14]. The better agreement between the data and the Standard Model prediction using the normalisation factor for $W+\text{jets}$ is shown in figure 7.1b.

The plot with the normalisation factor of 0.69 for $W+\text{jets}$ only illustrates that the disagreement of the data and the Standard Model prediction decreases when using such a normalisation factor. It is needed to determine the correct normalisation factors for $t\bar{t}$ and $W+\text{jets}$ for the used preselection in this analysis in order to reduce the discrepancies.

7.2. Determination of correct normalisation factors

The normalisation factors for the two major backgrounds can be determined by a fit to the neural network distribution, exploiting the different distributions for the processes. Technically this is done during the limit setting with `HistFitter`, which is described in chapter 8.1.

As a first test, the determination of a $W+\text{jets}$ normalisation factor was analysed. In a second step, the parallel determination of a $W+\text{jets}$ and a $t\bar{t}$ normalisation factor was investigated.

$W+\text{jets}$ normalisation factor The determination of the $W+\text{jets}$ normalisation factor is checked using all Standard Model backgrounds in order to produce pseudo data distributions with different $W+\text{jets}$ normalisations. The pseudo data distribution is constructed as

$$n^{\text{p.d.}}(i) = n_{t\bar{t}}(i) + \mu_W \cdot n_W(i) + n_{\text{others}}(i) \quad \text{for each bin } i \quad (27)$$

where the normalisation factor for the $W+\text{jets}$ sample is μ_W . The test is done for $\mu_W = 0.7$, 1.0 and 1.3. The resulting numbers are presented in table 7.1 for the bulk region. One can directly compare the given μ_W with the calculated μ_W using `HistFitter`. Both values match perfectly. The study was also done for the other regions and gives the same results which are shown in the appendix A.6.

Table 7.1: `HistFitter` output for the determination of the $W+\text{jets}$ normalisation factor. The value of the μ_W should be the same as used for producing the pseudo data (equation 27). Only statistical uncertainties are considered.

given μ_W	fitted μ_W
0.7	0.7 ± 0.1
1.0	1.0 ± 0.1
1.3	1.3 ± 0.1

In addition, a signal contribution was added to the pseudo data and the $W+\text{jets}$ normalisation factors are checked for three different normalisations (cf. equation 28). This checks the correlation of μ_W and μ_{Signal} . The outcoming numbers are listed in table 7.2.

$$n^{\text{p.d.}}(i) = n_{t\bar{t}}(i) + \mu_W \cdot n_W(i) + n_{\text{others}}(i) + n_{\text{Signal}}(i) \quad \text{for each bin } i \quad (28)$$

The normalisation factors for the $W+\text{jets}$ are again the same as the given values for the pseudo data. The calculated numbers for the other regions match also well and are given in the appendix A.6.

Table 7.2: `HistFitter` output for the determination of the $W+\text{jets}$ normalisation factor with an additional signal contribution in the pseudo data. The value of the μ_W should be the same as used for producing the pseudo data (equation 28). Only statistical uncertainties are considered.

given μ_W	fitted μ_W
0.7	0.7 ± 0.1
1.0	1.0 ± 0.1
1.3	1.3 ± 0.1

The $W+\text{jets}$ normalisation factor is not correlated to the μ_{Signal} and a signal contribution does not change the normalisation factor. This means that the $W+\text{jets}$ normalisation factors can be determined with this setup.

$W+\text{jets}$ and $t\bar{t}$ normalisation factor In the second study the $t\bar{t}$ sample is also added with different normalisations. For the factors $\mu_{t\bar{t}}$ and μ_W every combination of 0.7, 1.0 and 1.3 is analysed.

$$n^{\text{p.d.}}(i) = \mu_{t\bar{t}} \cdot n_{t\bar{t}}(i) + \mu_W \cdot n_W(i) + n_{\text{others}}(i) \quad \text{for each bin } i \quad (29)$$

The resulting numbers are given in table 7.3. The normalisation factors for $t\bar{t} - \mu_{t\bar{t}}$ – and $W+\text{jets} - \mu_W$ – should be in agreement with the used values for the generation of the pseudo data. Considering the calculated numbers for all nine configurations given in table 7.3, one can see that both normalisation factors match the value which is used for the generation of the pseudo data. Therefore one can determine the normalisation factors for the $t\bar{t}$ and the $W+\text{jets}$ background and there is no strong correlation between the normalisations for these two backgrounds.

To determine the correlation to a possible signal, the pseudo data are again generated with every combination for the normalisation factors for $t\bar{t}$ and $W+\text{jets}$ and an additional signal contribution (equation 30).

$$n^{\text{p.d.}}(i) = \mu_{t\bar{t}} \cdot n_{t\bar{t}}(i) + \mu_W \cdot n_W(i) + n_{\text{others}}(i) + n_{\text{Signal}}(i) \quad \text{for each bin } i \quad (30)$$

Signal contributions from all regions are checked. The resulting numbers for a signal from the bulk region are shown in table 7.4. The comparison of the used normalisation factors for $t\bar{t}$ and $W+\text{jets}$ with the outcome from `HistFitter` shows a good agreement for these values.

The normalisation of the two main backgrounds, $t\bar{t}$ and $W+\text{jets}$ processes, can be determined in a data-driven way. For this, the respective normalisation factors are calculated in a fit to the distribution of the neural network output. The determination is robust against any signal contributions.

Table 7.3: HistFitter output for the determination of the W +jets and the $t\bar{t}$ normalisation factors. The values of the μ_W and the $\mu_{t\bar{t}}$ should be the same as used for producing the pseudo data (equation 29). Only statistical uncertainties are considered.

given μ_W	given $\mu_{t\bar{t}}$	fitted μ_W	fitted $\mu_{t\bar{t}}$
0.7	0.7	0.70 ± 0.13	0.70 ± 0.03
	1.0	0.70 ± 0.15	1.00 ± 0.04
	1.3	0.70 ± 0.16	1.30 ± 0.04
1.0	0.7	1.00 ± 0.14	0.70 ± 0.03
	1.0	1.00 ± 0.15	1.00 ± 0.04
	1.3	1.00 ± 0.17	1.30 ± 0.04
1.3	0.7	1.30 ± 0.14	0.70 ± 0.03
	1.0	1.30 ± 0.16	1.00 ± 0.04
	1.3	1.30 ± 0.17	1.30 ± 0.04

Table 7.4: HistFitter output for the determination of the W +jets and the $t\bar{t}$ normalisation factors with an additional signal contribution in the pseudo data. The values of the μ_W and the $\mu_{t\bar{t}}$ should be the same as used for producing the pseudo data (equation 30). Only statistical uncertainties are considered.

given μ_W	given $\mu_{t\bar{t}}$	fitted μ_W	fitted $\mu_{t\bar{t}}$
0.7	0.7	0.70 ± 0.13	0.70 ± 0.03
	1.0	0.70 ± 0.15	1.00 ± 0.04
	1.3	0.70 ± 0.16	1.30 ± 0.04
1.0	0.7	1.00 ± 0.14	0.7 ± 0.03
	1.0	1.00 ± 0.14	1.00 ± 0.04
	1.3	1.00 ± 0.17	1.30 ± 0.04
1.3	0.7	1.30 ± 0.14	0.70 ± 0.04
	1.0	1.29 ± 0.16	1.00 ± 0.04
	1.3	1.30 ± 0.17	1.30 ± 0.04

7.3. Resulting comparison of data and simulation

In the following, different plots are shown in order to compare the Standard Model prediction with the measured data. For this, the different distributions are shown for all Standard Model backgrounds stacked. On top of the simulated samples the measured data are shown. In addition, the ratio between the data and the combined Standard Model prediction is shown in the lower part of the plot. This ratio should be around 1.

For the error band not only the statistical error but also the dominant systematic uncertainties due to the jet energy scale (JES) and the jet energy resolution (JER) are used. For each bin the statistical error σ_s is determined and both the error due to jet energy scale σ_{JES} and the error due to jet energy resolution σ_{JER} are calculated using

$$\sigma_{\text{JES}} = \frac{1}{2} \cdot (n(\text{JES up}) - n(\text{JES down})) \quad (31)$$

$$\sigma_{\text{JER}} = |n(\text{nominal}) - n(\text{JER})|. \quad (32)$$

Using these uncertainties, the total uncertainty for each bin is calculated by

$$\sigma_{\text{all}} = \sqrt{\sigma_s^2 + \sigma_{\text{JES}}^2 + \sigma_{\text{JER}}^2}. \quad (33)$$

This value is used in the plots in order to illustrate the possible variation of the Standard Model prediction.

The resulting normalisation factors determined using by the `HistFitter` framework (described in chapter 8) are given in table 7.5. The normalisation factors are not identical for all three analysis regions, although the same preselection is used. The different normalisation factors are influenced by a mismodelling in the simulated data samples. In a dedicated study, this should be analysed further.

The mean of the normalisation factors in each region is used in the validation plots, as it is done in the ATLAS analysis [13, 14]. For the limit setting procedure the normalisation factors determined for the individual region are used. The differences using either the individual normalisation factors for each region or the combined value, as given in table 7.5, are negligible. In the appendix A.7 this comparison is shown for the missing transverse energy \cancel{E}_T .

Table 7.5: The normalisation factors for the three analysis regions and the mean value as combination for all three regions.

	bulk	diagonal	high mass	combined
μ_W	0.918	0.916	0.757	0.856
$\mu_{t\bar{t}}$	0.831	0.93	0.959	0.91

The validation of the Standard Model prediction is shown for the missing transverse energy \cancel{E}_T , the transverse mass m_T , the perpendicular component of the missing transverse energy to the leptonically decaying top quark, and the transverse angle between both top quarks in figure 7.2. For all these plots the combined normalisation factors, given in table 7.5, are used.

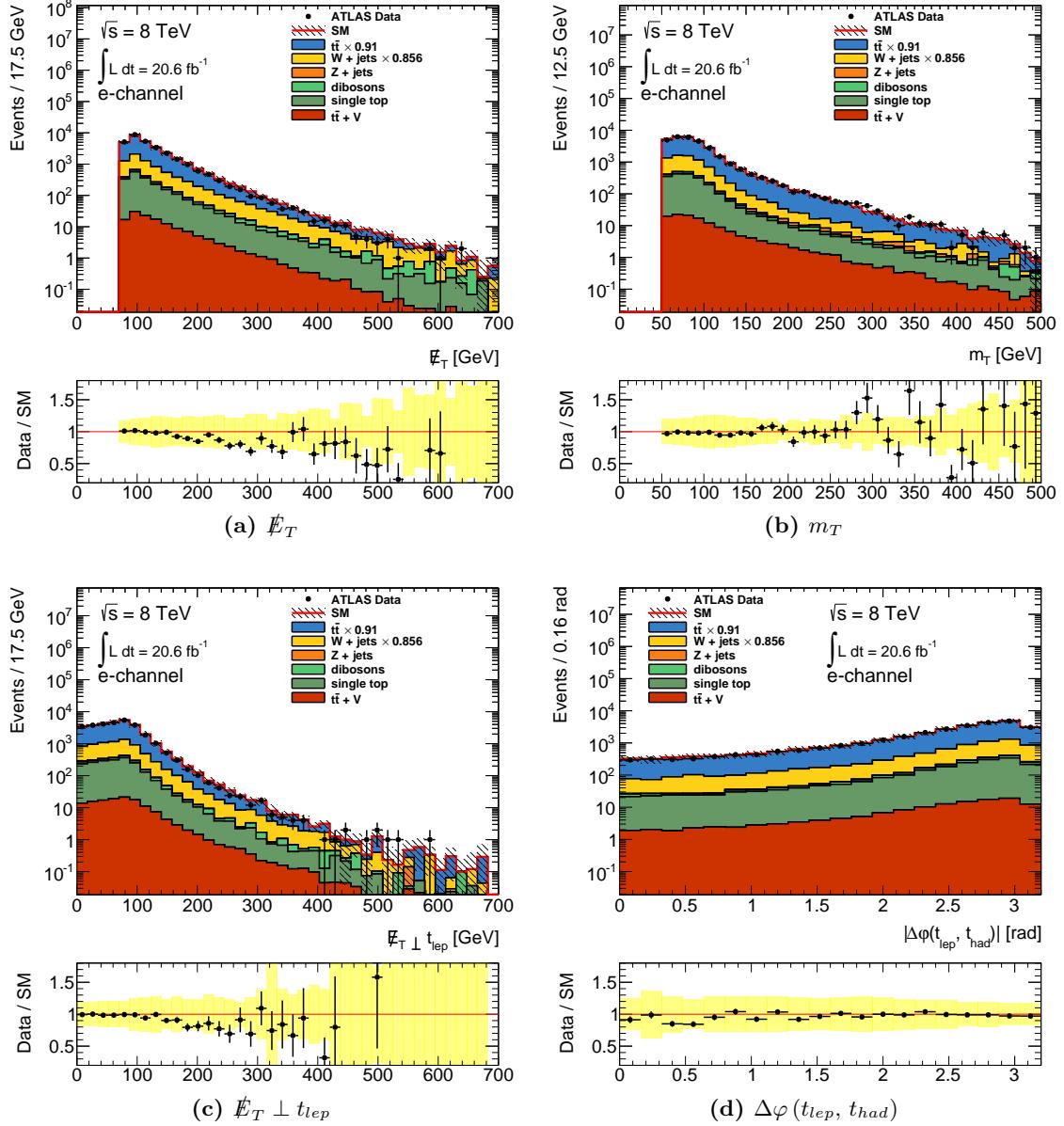


Figure 7.2: The distributions are shown for data with an integrated luminosity of 20.6 fb^{-1} and all Standard Model backgrounds stacked and scaled to this luminosity. The error bands show the statistical error and the systematic error due to the JES and JER uncertainties. The lower part of each plot shows the ratio of the data to the Standard Model prediction.

The data and the Standard Model prediction are compatible within the uncertainties. In the ratio between the data and the given Standard Model prediction, a slope is visible in some distributions, e.g. the missing transverse energy \cancel{E}_T . The influence of these slopes is analysed further in the following chapter.

7.4. Shape dependence of the neural network

In the comparison of the data and the Standard Model prediction some distributions show a disagreement in the shape. Such a slope in the ratio of data to simulation can, for example, be seen in the distribution of the missing transverse energy \cancel{E}_T (figure 7.2). In the following, the missing transverse energy \cancel{E}_T and the transverse mass m_T are analysed because both variables have a high impact in the training of all neural networks.

Both variables are used in order to change the event weights in the simulation separately. The new event weight w_{new} is calculated as

$$w_{\text{new}} = w_{\text{nom}} \cdot (1.0 + c \cdot \cancel{E}_T) \quad (34)$$

where w_{nom} is the nominal event weight and c is a number which is chosen in such a way, that the shape is similar to the shape seen in the ratio of the data to the Standard Model prediction. This event weight is calculated for each event in order to preserve the correlations of all variables depending on the missing transverse energy \cancel{E}_T . In addition, a larger variation is probed by using $c' = 2 \cdot c$ in order to illustrate a shape difference which is twice as large as the shape seen in the ratio plot.

As the transverse mass m_T is used in all three regions, possible variations in the distribution are checked in the same way.

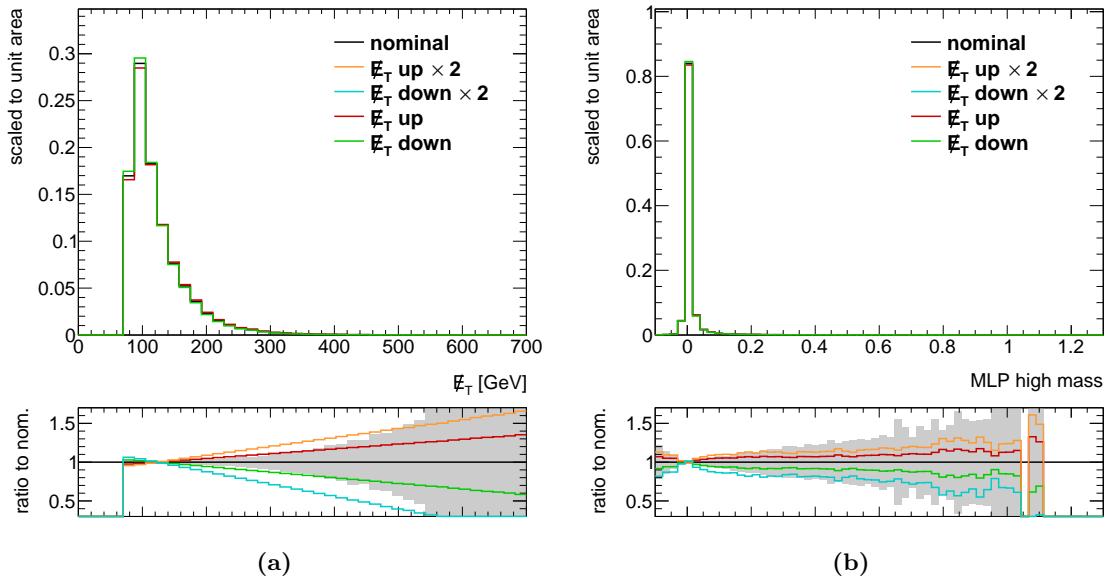


Figure 7.3: Illustration of the shape changes for the missing transverse energy \cancel{E}_T distribution (a) and the resulting changes in the neural network distribution (b) for the $t\bar{t}$ sample .

In figure 7.3 on the left side, the missing transverse energy \cancel{E}_T distribution is depicted. The nominal distribution is shown in black, the different shape changes are plotted using different colours. The resulting neural network distributions are presented on the right side of figure 7.3 and show a clear peak for low neural network values, as it is expected for the

$t\bar{t}$ sample. The small changes in the missing transverse energy are within the statistical error which is plotted as a grey band. The larger changes are slightly outside the statistical uncertainty. The changes in the neural network distribution due to the different missing transverse energy distributions are damped and mostly within the statistical error.

The changes due to a shape difference in the transverse mass m_T distribution are depicted in figure 7.4. The same colours are used as for the missing transverse energy and a similar behaviour is visible for the shape differences. The larger variation for the transverse mass are slightly outside the statistical error, but the changes of the resulting neural network distribution are within the statistical error. This means that also for the transverse mass the shape differences become damped in the neural network distribution.

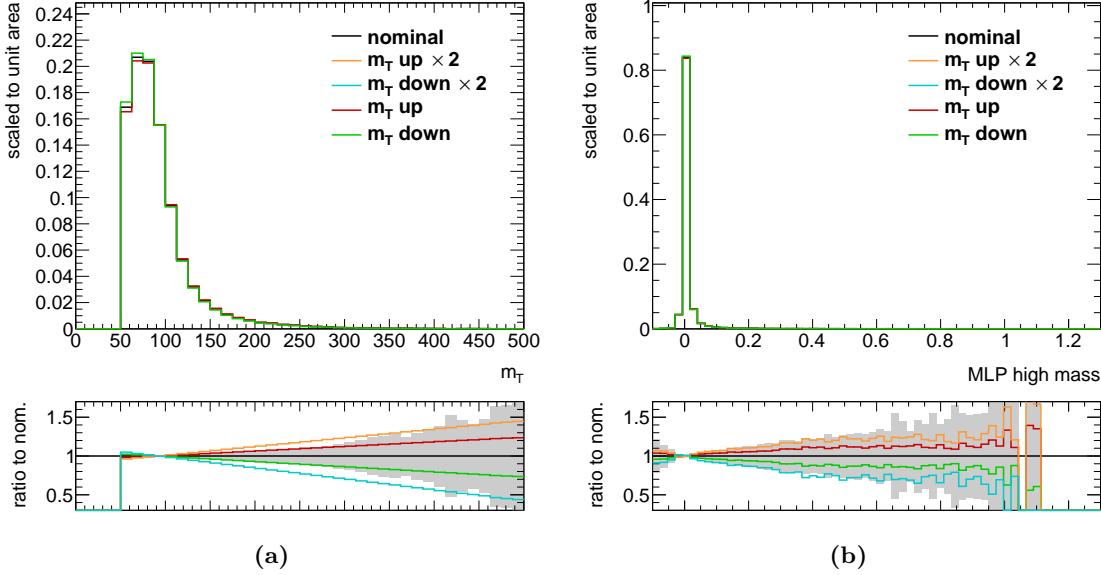


Figure 7.4: Illustration of the shape change for the transverse mass m_T distribution (a) and the resulting changes in the neural network distribution (b) for the $t\bar{t}$ sample.

This study shows that a shape difference in one of the input variables for the neural network is damped in the neural network itself. A mismodelling of one input variable has a lower effect on the overall results than for a simple cut method due to the smaller effect in the neural network.

7.5. Resulting neural network output distributions

In figures 7.5, 7.6a and 7.6b the distribution of the neural network output is shown for the bulk, the diagonal, and the high mass region respectively. For all three distributions the binning is changed in such a way that most bins contain similar numbers of events. This is done in order to improve the fit stability for the limit setting (cf. chapter 8.1).

The measured data and the Standard Model prediction agree well for all three regions. No slope is visible and especially the agreement for higher neural network values is good. For the bulk and the high mass region stronger fluctuations can be observed in the tail region, which is nonetheless still compatible within the uncertainty.

Especially for the bulk region a strong sensitivity can be observed in the neural network output. In the last bin, the same amount of events is expected for both signal and background.

All distributions show a good agreement between the observed data and the Standard Model prediction. No sign of physics beyond the Standard Model can be observed in any of the three regions.

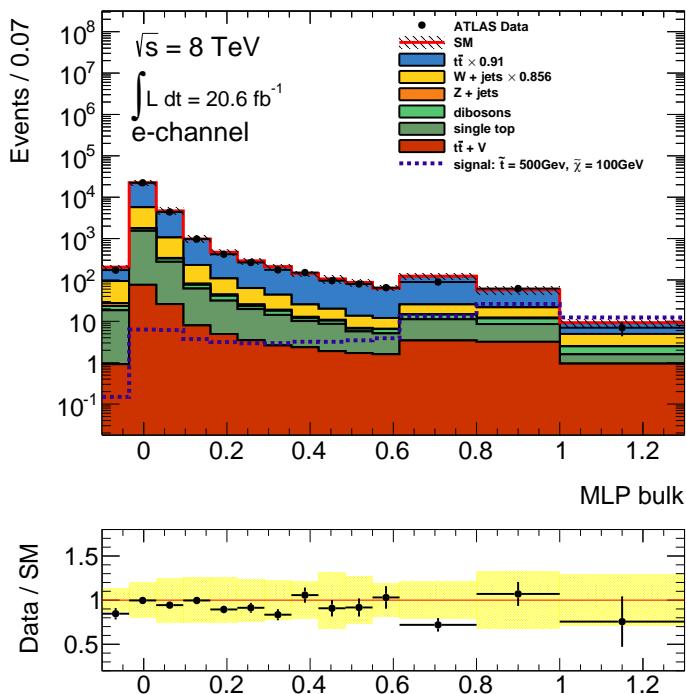


Figure 7.5: The neural network distribution for the bulk region is shown for data with an integrated luminosity of 20.6 fb^{-1} and all Standard Model backgrounds stacked and scaled to this luminosity. The error band shows the statistical error and the systematic error due to the JES and JER uncertainties. The lower part of the plot shows the ratio of the data to the Standard Model prediction.

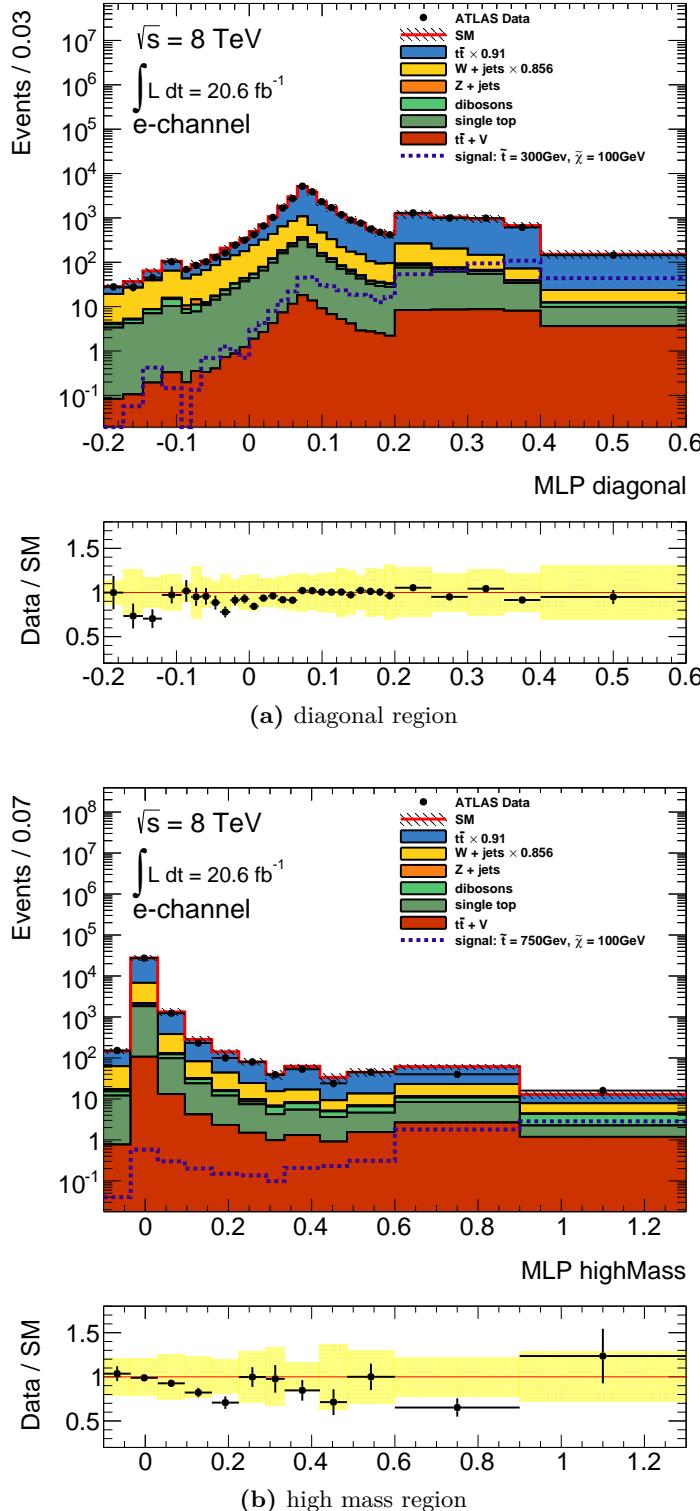


Figure 7.6: The neural network distributions are shown for data with an integrated luminosity of 20.6 fb^{-1} and all Standard Model backgrounds stacked and scaled to this luminosity. The error bands show the statistical error and the systematic error due to the JES and JER uncertainties. The lower part of each plot shows the ratio of the data to the Standard Model prediction.

8. Limits on simplified models

The neural network distributions do not show any excess of the measured data over the Standard Model predictions. The neural network distributions are therefore used to set limits on the stop and neutralino masses. For this, the `HistFitter` framework is used which is based on `RooFit`, `RooStats` and `HistFactory` [35–37].

8.1. Theoretical procedure

In order to set limits on the stop and neutralino masses in the context of simplified models, hypothesis tests using the profile likelihood approach are used. One hypothesis is the Standard Model of particle physics H_0 , the other hypothesis H_1 is the new physics with the Standard Model as background. In this study it is the Standard Model background and Supersymmetry as new physics with one light stop quark and one neutralino as the lightest supersymmetric particle.

The likelihood values for the data with respect to both hypotheses are calculated. For a cut and count analysis the likelihood is defined as

$$\mathcal{L}(\text{data} | H_i) = \text{Poisson}(\#\text{n}_{\text{data}} | \#\text{n}_{\text{theory}}(H_i)), \quad (35)$$

and for a distribution with more than one bin the likelihood is given by

$$\mathcal{L}(\text{data} | H_i) = \prod_{\text{bin}=1}^{n_{\text{bin}}} \text{Poisson}(\#\text{n}_{\text{data}}(\text{bin}) | \#\text{n}_{\text{theory}}(H_i, \text{bin})). \quad (36)$$

Systematic uncertainties are included into the likelihood definition as nuisance parameters. The profile likelihood ratio is then defined as

$$t = \frac{\mathcal{L}(\text{data} | H_1)}{\mathcal{L}(\text{data} | H_0)}. \quad (37)$$

According to the Neyman-Pearson lemma, this ratio is the optimal test statistic to distinguish between both hypotheses [38].

The observed value t_{obs} is calculated for the data and is used in order to determine the p -value p_1 for the hypothesis H_1 . The p -value is the integral over the probability distribution of t for all values equal or worse than the observed value. Larger p -values show a better compatibility of the data to the hypothesis. The hypothesis can be excluded for p -values lower than 0.05, corresponding to a 95% confidence level.

In order to avoid an overestimated sensitivity, the CL_s value [39] can be used instead of the p -value. For the CL_s value, the p -value for signal and background hypothesis, here the Standard Model and Supersymmetry, is divided by the p -value for the background only hypothesis, here the Standard Model,

$$\text{CL}_s = \frac{p_1}{p_0}. \quad (38)$$

In general, the p -value for the background p_0 is expected to be near 1. In case of a statistical fluctuation leading to fewer events than expected for the background only hypothesis and a resulting lower p -value, the signal hypothesis also becomes smaller and therefore could

be excluded, even though the experiment has little sensitivity. In such cases the p -value of the background is lower than 1, which leads to larger CL_s values. The signal hypothesis is excluded at 95% confidence level when the CL_s value is lower than 0.05.

8.2. Limit setting with pseudo data

In a first step, the limit setting procedure is evaluated on pseudo data. For this, limits on simplified models are calculated with only statistical uncertainty and with including systematic uncertainties. For this, all backgrounds are considered. In order to improve the fit stability only $t\bar{t}$ and $W+\text{jets}$ are used separately, for the smaller backgrounds a combination is used.

For the limit setting, the full neural network distribution is used. As already explained in chapter 7, during the limit setting process the scaling factors for $t\bar{t}$ and $W+\text{jets}$ are calculated. In order to improve this calculation especially for the $W+\text{jets}$ background, the neural network distribution using a $W+\text{jets}$ control region is used in addition. For this control region the b -tag requirement of the preselection is changed to a b -jet veto.

An example for the full distribution which is used for the limit setting is shown in figure 8.1. The left part shows the first bins of the neural network distribution of the $W+\text{jets}$ control region (CR W), the right part is the full neural network distribution using the standard preselection. The binning for this distribution is chosen in such a way that the bins for higher neural network values include enough events in order to increase the fit stability. Both distributions – the W control region and the full neural network – are binned with equal bins into a range between 0 and 1 in order to produce input histograms in a `HistFitter` compatible form.

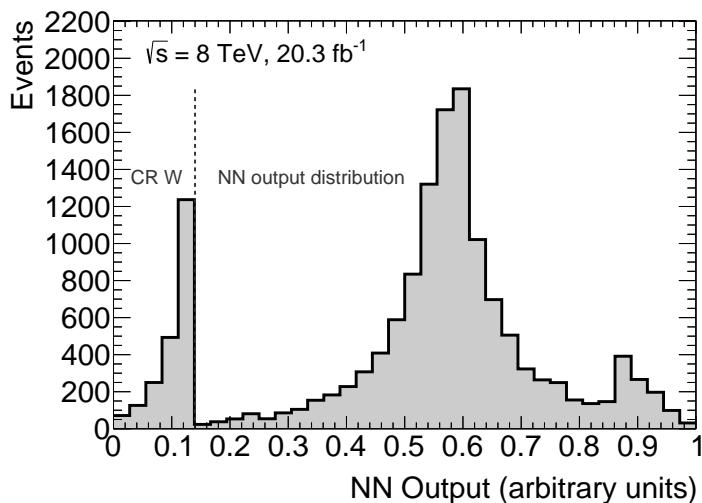


Figure 8.1: Example for the full distribution which is used for the limit setting procedure, shown for the $W+\text{jets}$ sample . The left part is selected from the $W+\text{jets}$ control region (CR W), the right part shows the standard neural network distribution.

Systematics All variations according to the systematic uncertainties are evaluated before applying the preselection. This is done in order to correctly treat changes on the object parameters in the selection, i.e. the jet p_T which can change due to the jet energy scale (JES). Therefore it is possible, that other objects, i.e. jets are selected as in the nominal sample. This includes the possibility of the expected number of events for each systematic variation differing from the expected number of events in the nominal sample.

The dominant uncertainty is the jet energy scale (JES) calibration. For this, the jet energy is varied up and down in a fully correlated way without a breakdown in the different JES components. For the jet energy resolution (JER), the transverse momentum of the jets is smeared. Using the JER systematic, a possible mishandling of the jet energy resolution in the simulated data samples is considered.

The same is done for the calibration of the missing transverse energy, i.e. the missing transverse energy scale (METs) and the missing transverse energy resolution (METR). The uncertainty of the b -tagging calibration is evaluated by changing the normalisation factors for each jet in the simulation in a range that reflects the systematic uncertainty on the measured tagging efficiency and the mistag rates. These normalisation factors are η -, p_T - and flavour-dependent. The variations are applied to B , C and light jets separately. This leads to three uncorrelated systematic uncertainties, called BTAG, CTag and LightTag. In order to determine the uncertainty of the pileup normalisation factor (PU), the average number of interactions per crossing μ in Monte Carlo is changed by $\pm 10\%$. Uncertainties on the lepton normalisation factors and the trigger efficiency are considered accordingly [40–42].

Due to the statistical uncertainty of the Monte Carlo sample, the evaluation of the different systematics leads to fluctuations in the distributions of the systematic uncertainties. The statistical error is already included in the neural network distribution without any systematics. In order to not double count the statistical error and to improve the stability of the fits for the limit setting, the distributions of the systematics are smoothed using a weighted average of three bins.

For the smoothing, first the ratio of the systematic variation and the nominal distribution is calculated for each bin i

$$R(i) = n^{\text{syst}}(i) / n^{\text{nom}}(i). \quad (39)$$

This ratio is used in order to calculate the weighted average

$$R'(i) = \frac{1}{w_{i-1} + w_i + w_{i+1}} \cdot (w_{i-1} \cdot R(i-1) + w_i \cdot R(i) + w_{i+1} \cdot R(i+1)), \quad (40)$$

where the weights are calculated in the following way

$$w_{i\pm 1} = 0.25 \cdot \left(\frac{\sigma(i \pm 1)}{n^{\text{nom}}(i \pm 1)} \right)^{-2} \quad (41)$$

$$w_i = 0.5 \cdot \left(\frac{\sigma(i)}{n^{\text{nom}}(i)} \right)^{-2}. \quad (42)$$

The weights of the adjacent bins are reduced by 50% compared to the central bin. When the ratio of the systematic variation and the nominal distribution is too high (> 5) or too

low (< 0.1) for one bin, the corresponding weight is set to zero in order to suppress the largest fluctuations. The outcome of smoothed variation is then

$$n^{\text{syst, smoothed}}(i) = n^{\text{nom}}(i) \cdot R'(i). \quad (43)$$

One illustration of this smoothed systematic is shown in figure 8.2. In the upper part of this plot, the nominal neural network distribution for the bulk region in the muon channel is shown for the $t\bar{t}$ sample. In addition, both the METR up and the METR down variations are shown; they do not differ much from the nominal distribution. In the lower part of the plot, the ratio of the variation to the nominal distribution is plotted. The grey band illustrates the statistical error of the distribution. The dashed lines are the ratio of the systematic variation without the smoothing procedure, the solid lines are the ratio of the systematic variation using the smoothing. As one can see, the smoothing procedure reduces the statistical fluctuations of the systematics.

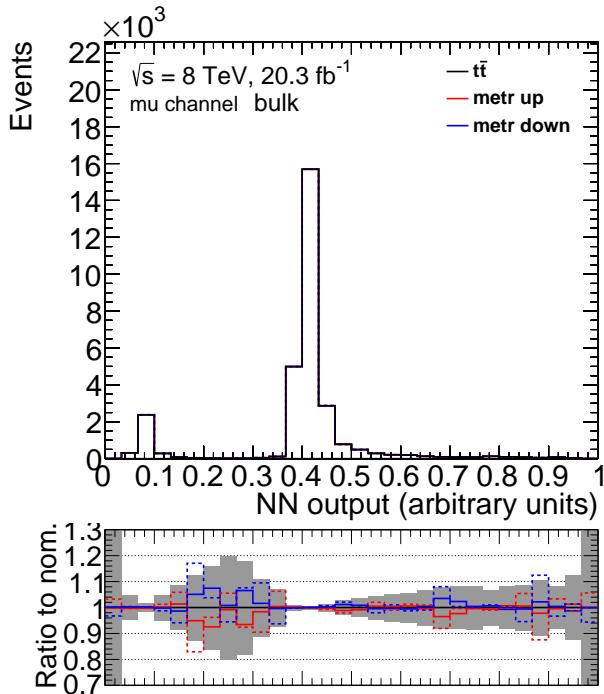


Figure 8.2: The nominal distribution for the $t\bar{t}$ sample for the bulk region in the muon channel. The up (red) and down (blue) variations for the METR are plotted too. In the ratio plot, the grey band illustrates the statistical error of the nominal distribution. The other lines show the ratio of the systematic variation to the nominal distribution using the smoothing method (solid) and without the smoothing (dashed).

This smoothing is applied to all systematics and the smoothed variation of each systematic is used for the following studies.

The systematics have different impact on the neural network output distribution for each region. When the changes due to a systematic are within the statistical error range, the systematic is not used for the limit setting in order to increase the overall fit stability. Therefore every systematic is investigated separately for each region. When the impact of a systematic for at least one sample, either a signal or a background sample, is larger than the statistical error, the systematic is selected for this region for the limit setting. In table 8.1 all used systematics are listed. The check marks show which systematic is used for which analysis region.

Table 8.1: The used systematics are listed for each region individually.

	JES	JER	METS	METR	BTag	LightTag
diagonal	✓	✓	✓	✓	✓	✓
bulk	✓	✓	✓	—	✓	✓
high mass	✓	✓	—	—	✓	✓

The systematics can be correlated to the normalisation factors for $t\bar{t}$ and $W+jets$. When the normalisation of one systematic uncertainty changes, also the normalisation factors can change. The systematic uncertainties JES, JER, METS, METR and BTag show strong correlations to the normalisation factors during the limit setting. Therefore the corresponding distributions are scaled to the same number of events as the nominal distribution in order to reduce the correlations and improve the calculation of the normalisation factors, as done in [13, 14].

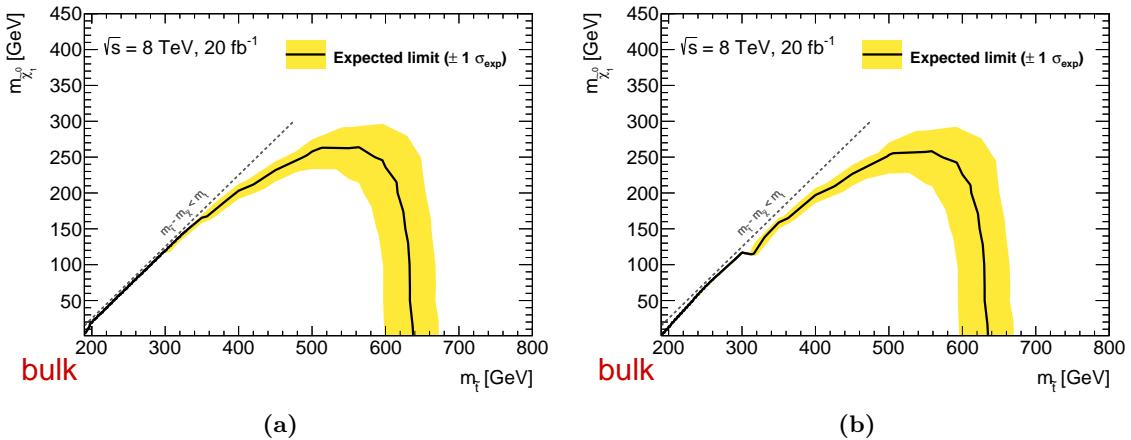


Figure 8.3: Limits on the stop and the neutralino masses using the neural network output of the bulk region without systematics (a) and including the significant systematics of this region (b) (cf. table 8.1).

Limits In the following, the results of the limit setting are shown for each region individually. For this, the expected limit is shown as black line and the $\pm 1\sigma_{\text{exp}}$ uncertainty is shown as yellow band. All stop and neutralino masses beneath the expected limit contour are excluded. In figure 8.3 the results for the bulk region are shown. In the left plot the limits for the bulk region without systematics are shown, on the right side the same limits are plotted including the significant systematics. The limits are compatible with each other, only in the diagonal region the limits worsen slightly using systematics. The stop masses between 200 GeV and 640 GeV are excluded for massless neutralinos. Neutralino masses up to 260 GeV can be excluded for stop masses around 500 GeV.

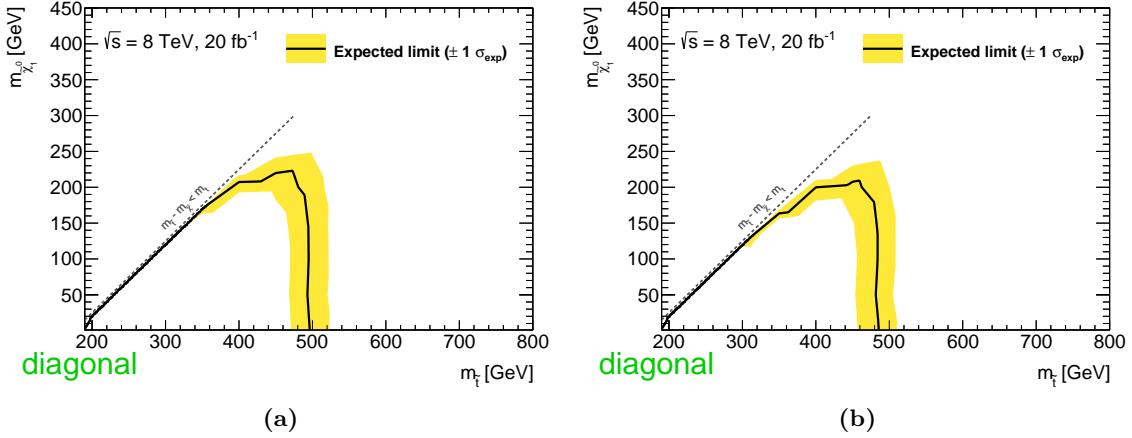


Figure 8.4: Limits on the stop and the neutralino masses using the neural network output of the diagonal region without systematics (a) and including the significant systematics of this region (b) (cf. table 8.1).

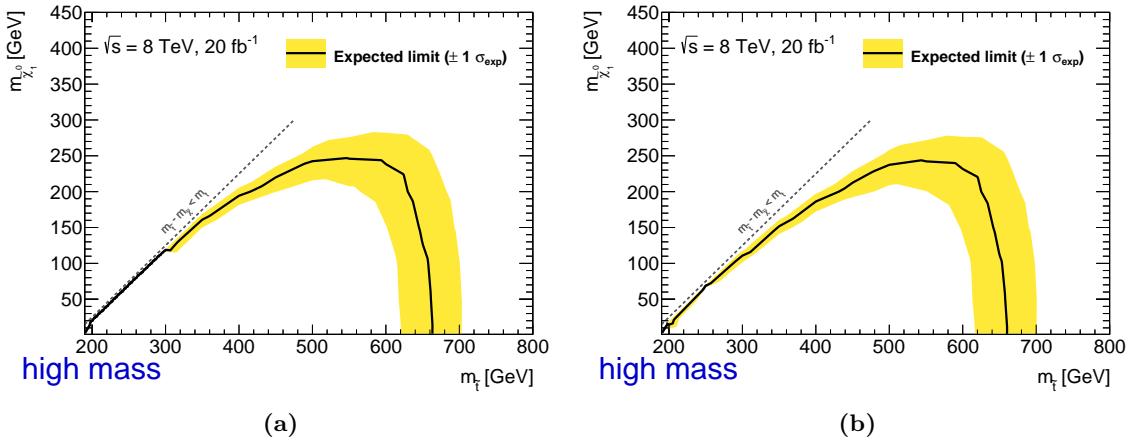


Figure 8.5: Limits on the stop and the neutralino masses using the neural network output of the high mass region without systematics (a) and including the significant systematics of this region (b) (cf. table 8.1).

The limits for the diagonal region are shown in figure 8.4 again both without systematics and when using the significant systematics for the diagonal region. Also these limits differ not much. Comparing these limits to the results in the bulk region shows that the limits of the diagonal region are closer to the diagonal, as expected. Near the diagonal boundary, the line of the expected limit gets up to a stop mass of 400 GeV, while it is only up to 350 GeV for the bulk region.

The limits for the high mass region are also shown without systematics and using the significant systematics. The corresponding plots are depicted in figure 8.5. In the diagonal part, the high mass region is much worse, which is expected as this region is especially trained for the high stop masses. For high stop masses both limits reach up to 660 GeV, which improves the range compared to the bulk region by ~ 20 GeV.

In order to determine the combined limit of these three regions, the region with the best, i.e. lowest expected CL_s value is chosen. Therefore for each sample only one region is chosen for the combined limit. The regions which are chosen for each sample point are shown in figure 8.6, again for both without and with systematics. The result is as expected: For the high mass stop samples the high mass region (red) is used, for the diagonal part, the diagonal region (cyan blue) is used in the majority of cases and for the samples in between the bulk region (green) is used. For some diagonal signals the neural network trained for the bulk region is chosen. This can happen due to failed fits for the limits using the neural network for the diagonal region.

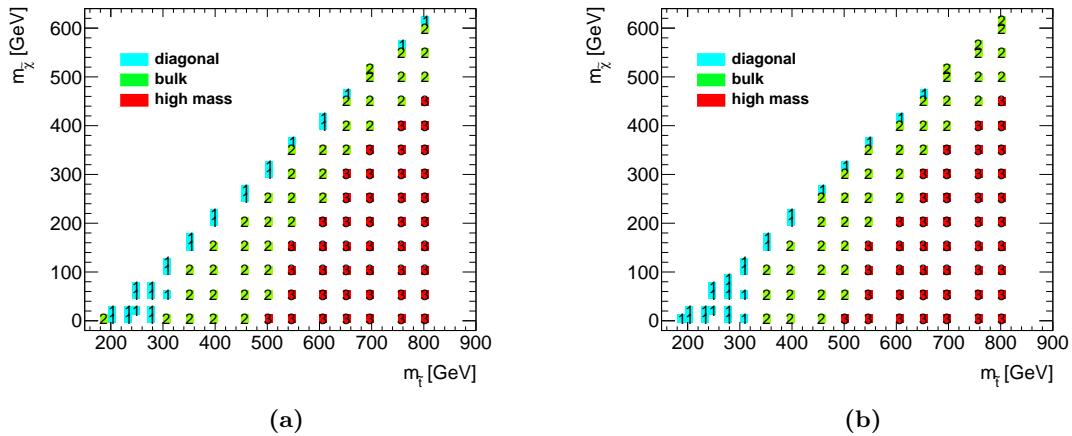


Figure 8.6: The region which is used to obtain the best expected limit of all three analysis regions without systematics (a) and with the significant systematics (b) (cf. table 8.1).

The resulting best expected limit, combining all three regions as explained before, is shown in figure 8.7. As one can see, both limits without systematics and with the most significant systematics do not change much. Using the pseudo data, stop masses up to 660 GeV are excluded. For the diagonal part, the limits are close to the diagonal up to 400 GeV. The limits are compared to the official ATLAS limits in chapter 8.4.

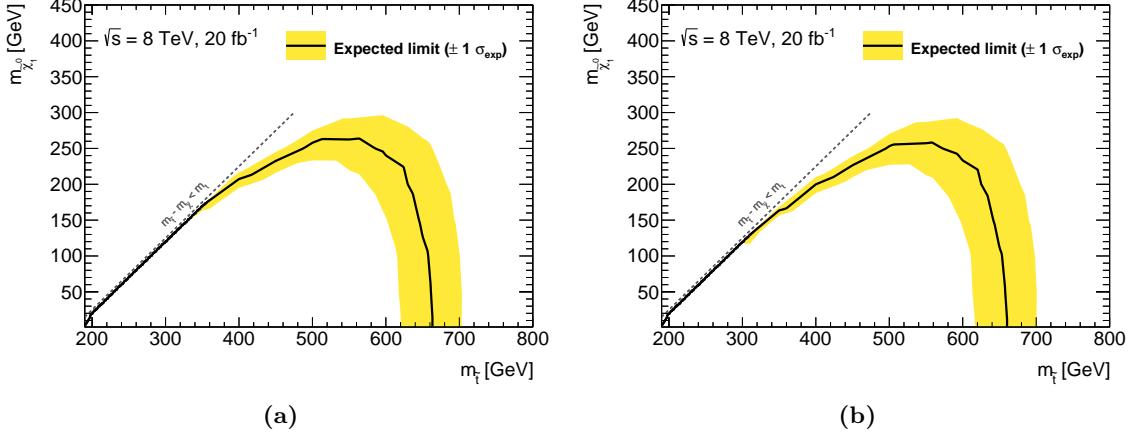


Figure 8.7: Best expected limits for the stop and the neutralino mass using the neural network output of all regions without systematics (a) and with the significant systematics (b) (cf. table 8.1).

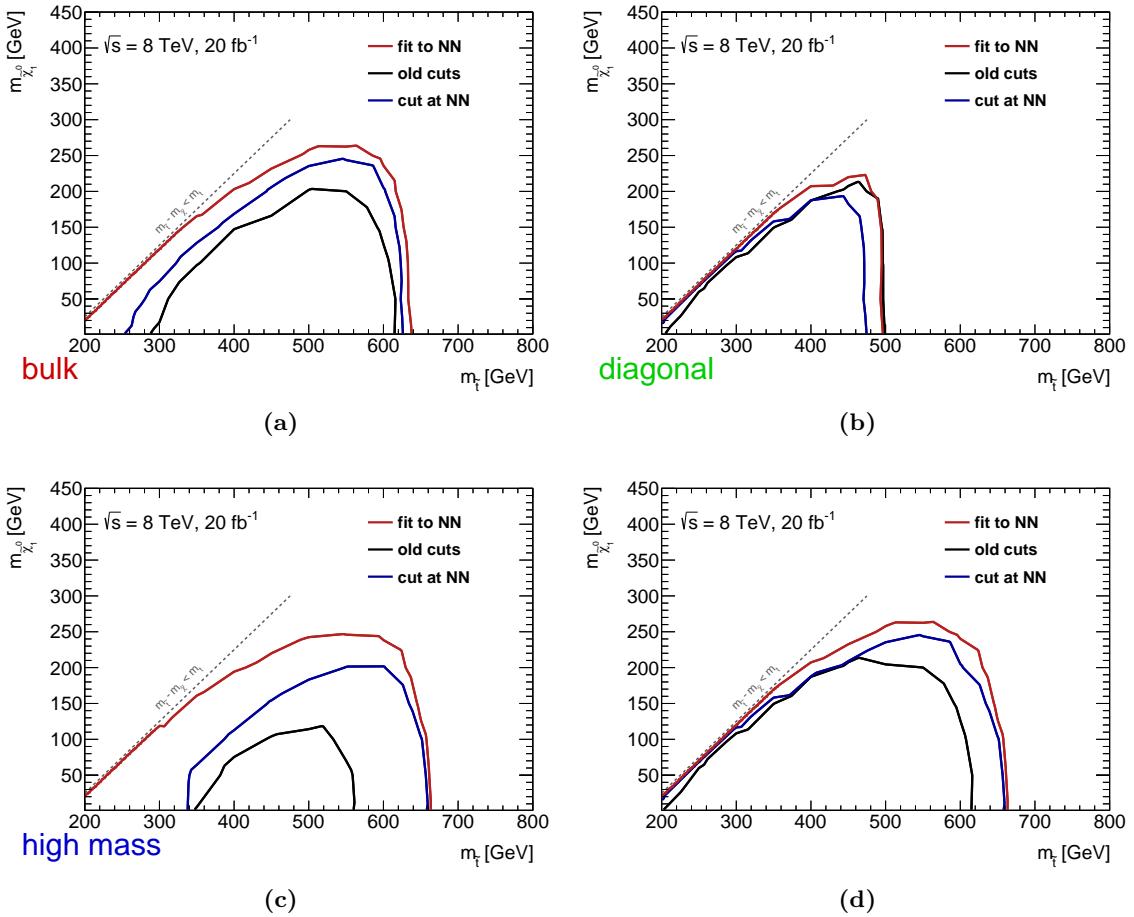


Figure 8.8: Limits for the stop and the neutralino mass using simple cuts (black), a cut on the neural network distribution (blue) and the full neural network distribution (red) for the bulk region (a), the diagonal region (b), the high mass region (c) and the best expected of all regions combined (d).

Limit comparison using different input methods The limits using no systematics are calculated for the old cuts (cf. table A.2) and a cut on the neural network distribution, in addition to the usage of the full neural network distribution. This cut is chosen for the best significance as explained in chapter 6.

The results using the cuts and also the full distribution of the neural network are compared in figure 8.8 for each region and the combined limit separately.

For all three regions a significant improvement is visible when the limits using the old cuts (black) and the limits using a cut at the neural network distribution (blue) are compared. In addition, an improvement is visible comparing the cut at the neural network distribution and using the full distribution (red). This improvement is slightly smaller for the high mass region. This can be caused by the lower statistics.

The observed improvement due to the neural network is consistent with the improvement of the significances after the optimisation (cf. chapter 6.2).

Limit comparison using the $t\bar{t}$ reweighting method The official ATLAS analysis uses a $t\bar{t}$ reweighting method [13, 14] in order to improve the agreement between the measured data and the Standard Model prediction. For this, the event weight is changed depending on the transverse momentum of the truth $t\bar{t}$ pair (cf. appendix A.8). The reweighting method is applied to the $t\bar{t}$ sample, and the neural network distribution is evaluated using this sample. The limits using both the reweighted sample and the nominal sample for the $t\bar{t}$ background are calculated and the comparison of the results is shown in figure 8.9.

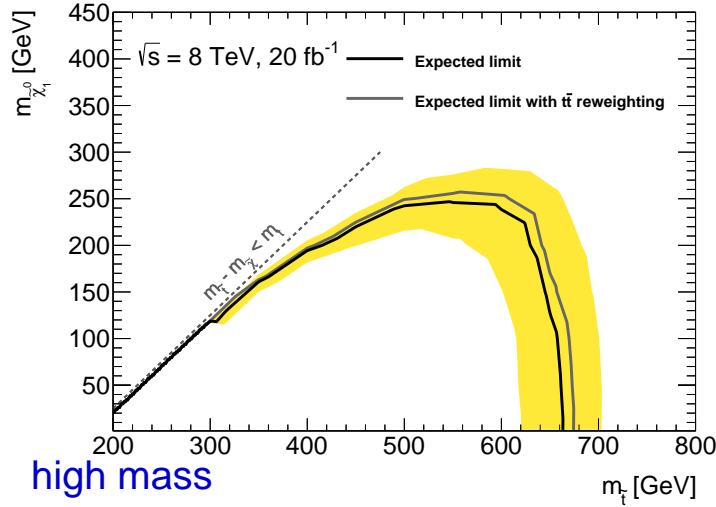


Figure 8.9: The expected limit using the nominal $t\bar{t}$ sample (black) with the uncertainty band (yellow) is compared to the expected limit using the reweighted $t\bar{t}$ sample (grey).

The $t\bar{t}$ reweighting method improves the expected limits by around 20 GeV. A reoptimisation of the neural network including the reweighted $t\bar{t}$ sample is expected to further improve the expected sensitivity. Using the reweighted sample includes the changes in the kinematic variables in the training.

8.3. Limit setting using data

In the following, the limits using data are calculated and explained. The expected limits change with respect to the pseudo data tests done before, as the results of the data-driven normalisation factors are used for the evaluation. The same systematics are used for each region as described before and documented in table 8.1.

The resulting limits are shown in the following. For the combination, all three signal regions are analysed separately for each mass point and the region with the best expected limit is used. The limit plot is depicted in figure 8.10. Appendix A.8 gives the limits for each separate region as well as the mapping of the used region for the combination.

The expected limit in figure 8.10 changes slightly compared to the expected limit shown before, using pseudo data only. Using data for the limit setting procedure implies the calculation of the normalisation factors for $t\bar{t}$ and $W+\text{jets}$, which changes the expected limit slightly to better values. The observed limit is generally stronger than the expected limit. As one can see in the comparison plots between the measured data and the Standard Model prediction (cf. figures 7.5, 7.6a and 7.6b), the number of events for data is always lower than the expected number of events using the Standard Model prediction for the high neural network values. Especially for the bulk and the high mass region, the differences are easy to see in the ratio plot. A possible signal would be most probably visible for the high neural network values. As fewer events are observed than expected from the Standard Model (cf. chapter 7.5), stronger limits can be set to additional sources like Supersymmetry.

Light stop quarks can be excluded for masses between 190 GeV and 680 GeV for massless neutralinos. For a stop mass of 640 GeV, neutralinos lighter than 250 GeV can be excluded.

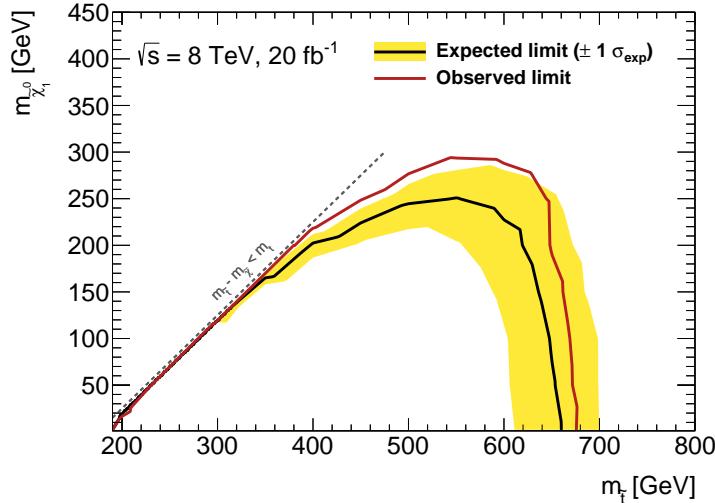


Figure 8.10: Limits for the stop and the neutralino mass using the neural network output with systematics and measured data (cf. table 8.1).

In figure 8.10 the stronger observed limit can be seen clearly for the bulk and the high mass region. For the diagonal region a transformation is done in order to produce a zoom into the diagonal part. As transformation, the y-axis is changed using:

$$y = m_{\tilde{t}} - m_{\tilde{\chi}_1^0} - m_t. \quad (44)$$

Using this transformation, the y -axis directly shows the distance to the diagonal boundary. The end point for $m_{\tilde{t}} = m_{\tilde{\chi}_1^0} + m_t$ is transformed to 0 GeV and is constant for all stop masses. Therefore the distance to the diagonal is directly visible. The displayed contour line is not continuous due to a visualisation problem as the signals at the edge of the parameter space can be excluded. The excluded region contains all signal points between the left and the right part of the contour line. The numbers on the plot give the observed CL_s values for all used samples.

Using this transformation, it is visible that the expected limit closes the gap to the diagonal up to 10 GeV for stop masses lower than 350 GeV. Also for the diagonal region the observed limit is stronger than the expected limit and excludes light stop quarks with a mass lower than 400 GeV for distances to the diagonal of 10 GeV. Very light stop quarks with $m_{\tilde{t}} = 200$ GeV are excluded for nearly massless neutralinos.

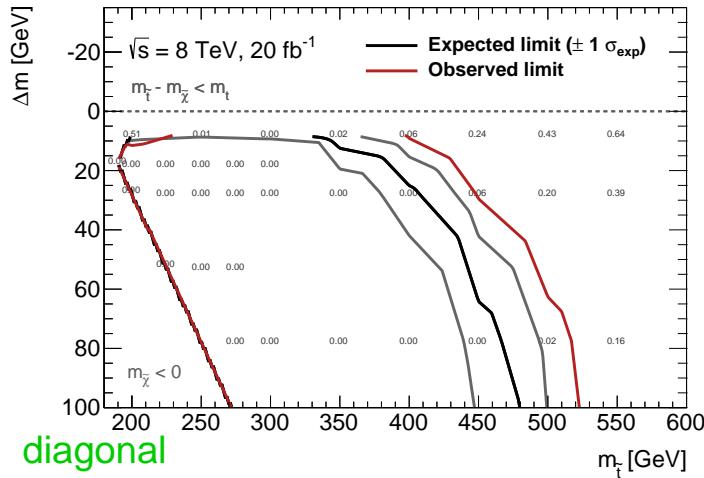


Figure 8.11: Limits for the stop and the neutralino mass using the neural network output of the diagonal region with systematics and measured data (cf. table 8.1) using a zoom into the interesting region. The transformation is used in such a way, that the diagonal boundary layer is a constant at 0 GeV for all stop masses using $\Delta m = m_{\tilde{t}} - m_{\tilde{\chi}_1^0} - m_t$ (displayed using an inverted y-axis). The given numbers on the plot are the observed CL_s values.

8.4. Comparison to official results

In the following, the results using the full neural network distribution for the limit setting procedure are compared to the newest official ATLAS results [13, 14]. The best combined limits for both analyses are shown in figure 8.12. Here the observed limit for the analysis using the full distribution is improved in the bulk region and similar in the high mass region compared to the official results. The observed limit starts at the same stop mass around 680 GeV for both analyses, but the exclusion using the full neural network distribution is stronger for increasing neutralino masses.

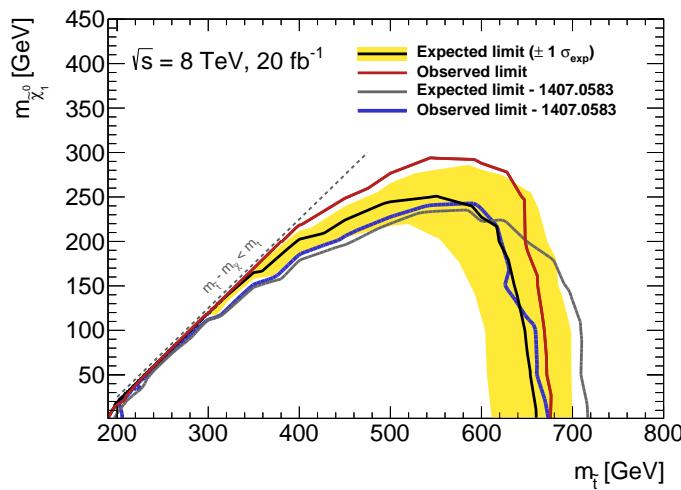


Figure 8.12: Best combined limits using the full neural network distribution (black with yellow band and red) and from the official ATLAS analysis (grey and blue) [13, 14].

The expected limit for the analysis using the full neural network distribution is much better in the bulk region, but worse in the high mass region. The better expected limit in the high mass region for the official ATLAS analysis is caused by the use of an additional region which is especially optimised for boosted events containing very collimated decay products and works better for high stop masses.

In order to compare similar results, only the so called tN3 region from the official analysis is compared to the limits using the full neural network distribution. This is the tightest cut and count region which does not use boosted topology variables for the selection and is compatible to the optimised high mass region from this study. The only difference is the $t\bar{t}$ reweighting method which was described in chapter 8.2 and is applied in the official ATLAS analysis. In figure 8.13 the results of the tN3 region are compared to the expected limits with and without applying the $t\bar{t}$ reweighting procedure explained before.

As one can see, the expected limit using the full neural network distribution, which is plotted in black, is around 20 GeV worse than the official results, which are shown in blue. Adding the $t\bar{t}$ reweighting method (shown in grey) the expected limits using the full neural network distribution improve around 20 GeV. The same difference between the expected limit from this analysis and the official ATLAS results was seen before. Therefore, the

expected limit using the $t\bar{t}$ reweighting method is similar to the official tN3 region for high stop masses. Starting from a neutralino mass of around 120 GeV, the expected limit using the full neural network distribution becomes much better than the expected limit from the official tN3 region.

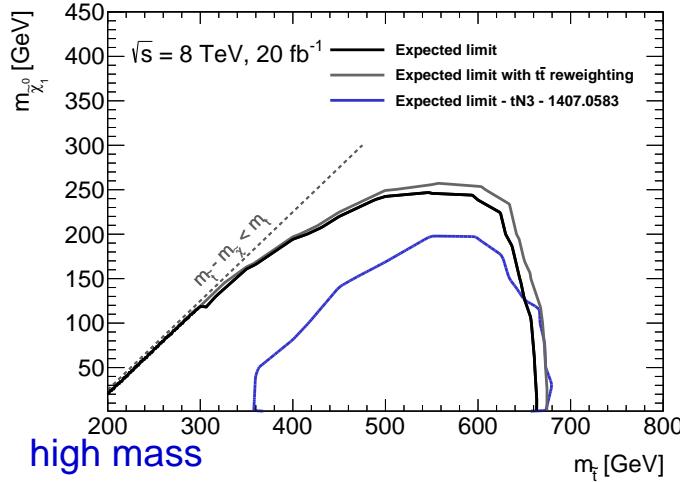


Figure 8.13: Expected limits for the high mass region using the full neural network distribution without (black) and with (grey) the $t\bar{t}$ reweighting method, and the limits from the official ATLAS analysis (blue) which uses the $t\bar{t}$ reweighting method [13, 14].

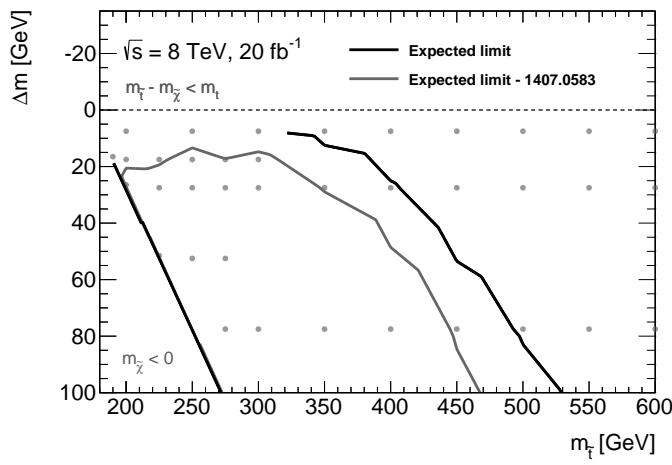


Figure 8.14: Best expected limits for the stop and the neutralino mass using the full neural network distribution (black) and for the official ATLAS analysis (grey) [13, 14] using a zoom into the interesting region. The transformation is used in such a way, that the diagonal boundary layer is a constant at 0 GeV for all stop masses using $\Delta m = m_{\tilde{t}} - m_{\tilde{\chi}_1^0} - m_t$ (displayed using an inverted y-axis). The grey dots illustrates the position of signal samples.

For the comparison of the official results and the results from this analysis in the diagonal region, the same transformation is applied as described for figure 8.11. The expected limits for both regions using this zoomed plot are depicted in figure 8.14. As one can see, the expected limit using the neural network distribution has a better sensitivity than the official study. The distance to the diagonal (dotted grey line in the plot) is lowered by at least 10 GeV for low stop masses and even more for higher stop masses. All signal samples near the diagonal for low stop masses can be excluded using the neural network to set limits. Therefore the contour line should continue somewhere above the grey dots in this figure. The correct continuation for the contour is not exactly defined as no more signal samples with a minimal distance lower than 10 GeV are available.

8.5. Conclusion

The limit setting procedure is applied to the full neural network distribution which was optimised before. The limits are calculated without and with the important systematics for the three signal regions.

The cut and count method, a simple cut using the neural network distribution, and the full neural network distribution are compared. The expected limits without any systematics for all three methods are calculated. A good improvement is visible for a cut at the neural network distribution. The best expected limit is achieved using the full neural network distribution for all three regions.

The $t\bar{t}$ reweighting method was applied and the neural network was evaluated using this new sample. Comparing the official results from the high mass region without boosted events – the tN3 region – with the results using the full neural network distribution, a slight decrease for the results of this analysis is visible for neutralino masses lower than 120 GeV. Including the effects of the $t\bar{t}$ reweighted sample, the limits are compatible.

The official expected limits are around 60 GeV higher (40 GeV compared to the limits with the $t\bar{t}$ reweighting method) but they use a special optimisation for boosted events. A different neural network could be optimised especially for these boosted events. In a dedicated study good variables for a separation should be found and a neural network should be trained using these new variables.

Comparing the expected limits for the diagonal and the bulk region between the official ATLAS results and the results from this analysis show an overall improvement. In addition, the observed limits show a similar sensitivity in the high mass region and a better one in the other two regions.

The expected limits reach stop masses up to 660 GeV for massless neutralinos. Neutralino masses lower than 250 GeV are excluded for stop masses around 550 GeV. The observed limit is stronger and excludes stop masses up to 680 GeV for massless neutralinos. For stop masses around 550 GeV, neutralino masses lower than 300 GeV are excluded.

9. Preparation for run II and reoptimisation

The optimisation of the neural network was redone for the high mass region in the electron channel, using Monte Carlo samples produced at a centre of mass energy of 14 TeV. For the optimisation the same procedure was used as described in chapter 6.2.

9.1. Variable distributions

In the following, the shape distributions of the input variables for the neural network are discussed. The variables are shown for a signal sample – $\tilde{t}(1500, 400)$ – and a $t\bar{t}$ sample, which is the major background. For the training all stop samples from table 9.1 are considered as signal in order to prevent a training specific to exactly one mass point. The Monte Carlo samples which are used for this study are only available on truth level without any detector simulation.

Table 9.1: Stop samples which are used for the neural network optimisation with a centre of mass energy of 14 TeV. The samples used exist only on truth level without any detector simulation.

$m_{\tilde{t}}$ [GeV]	$m_{\tilde{\chi}_1^0}$ [GeV]
1300	300
1400	200
1400	500
1500	400
1500	700

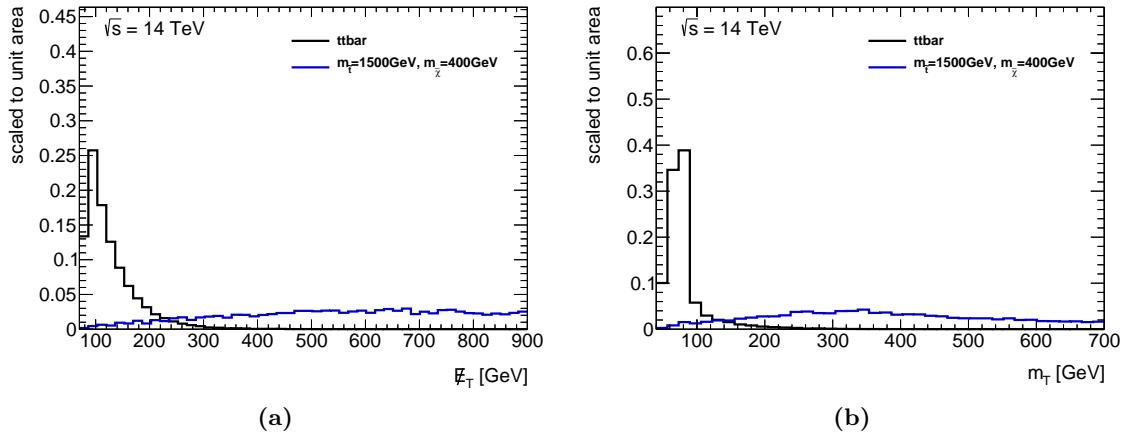


Figure 9.1: Distributions of the missing transverse energy \cancel{E}_T (a) and the transverse mass m_T (b) for a signal with $m_{\tilde{t}} = 1500$ GeV and $m_{\tilde{\chi}_1^0} = 400$ GeV and the $t\bar{t}$ background. The distributions are normalised to unit area in order to illustrate the shapes.

In figure 9.1a one can see the missing transverse energy, and figure 9.1b shows the transverse mass. Both distributions show a clear separation between the signal and the $t\bar{t}$ decay. For the $t\bar{t}$ sample both variables peak around 100 GeV, while the signal shows a broad distribution towards much larger values. A similar behaviour is visible for the perpendicular \cancel{E}_T component in figure 9.2a.

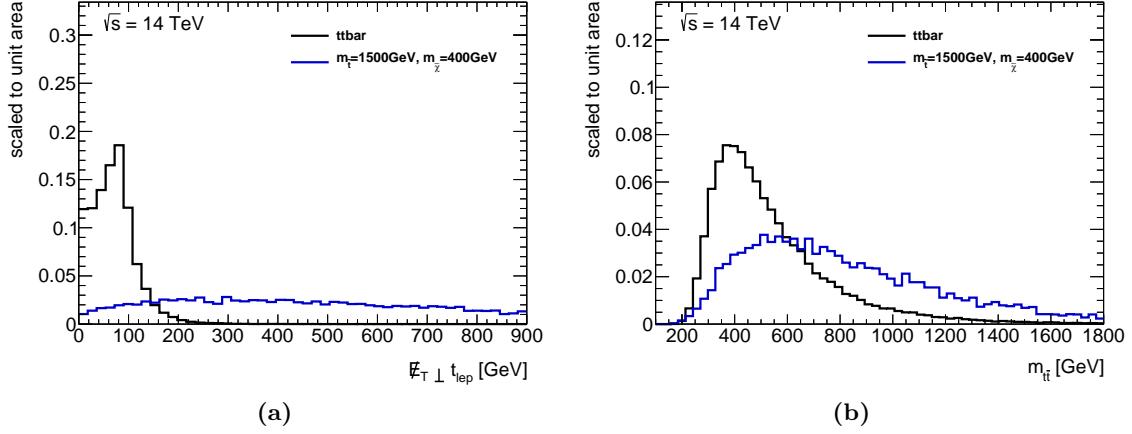


Figure 9.2: Distributions of the perpendicular \cancel{E}_T component $\cancel{E}_T \perp t_{lep}$ (a) and the mass of the $t\bar{t}$ pair $m_{t\bar{t}}$ (b) for a signal with $m_{\tilde{t}} = 1500$ GeV and $m_{\tilde{\chi}_1^0} = 400$ GeV and the $t\bar{t}$ background. The distributions are normalised to unit area in order to illustrate the shapes.

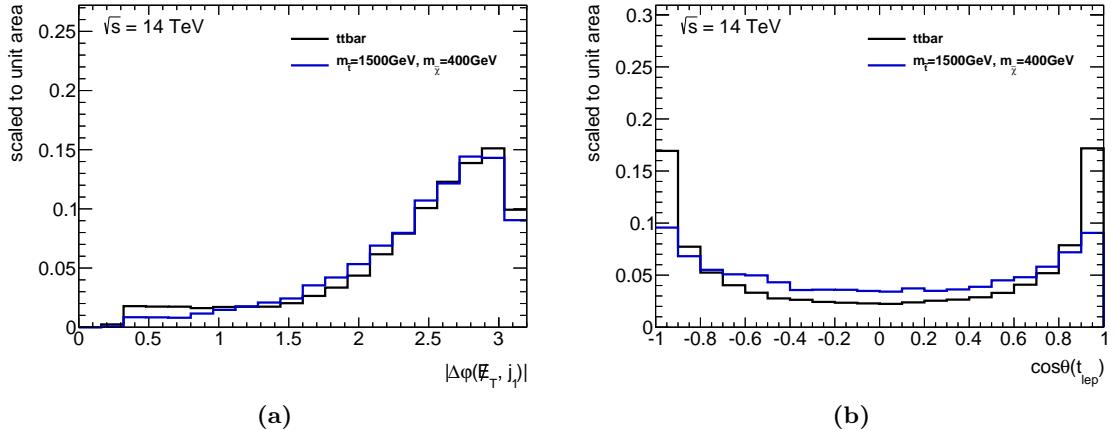


Figure 9.3: Distributions of the transverse angle between the \cancel{E}_T and the first jet $\Delta\varphi(\cancel{E}_T, j_1)$ (a) and the angle of the leptonically decaying top quark $\cos\theta(t_{lep})$ (b) for a signal with $m_{\tilde{t}} = 1500$ GeV and $m_{\tilde{\chi}_1^0} = 400$ GeV and the $t\bar{t}$ background. The distributions are normalised to unit area in order to illustrate the shapes.

In figure 9.2b the mass distribution of the $t\bar{t}$ pair is shown. The $t\bar{t}$ sample shows the expected exponential drop off starting at the production threshold around 350 GeV, while the stop sample has a maximum around 600 GeV with a long tail. The shapes have a good separation power also for this distribution. Less significant shape differences are visible in figure 9.3a. The shape of the transverse angle between the missing transverse energy and the first jet does not differ much between the stop and the $t\bar{t}$ sample. But due to correlations to other variables, e.g. the missing transverse energy, these variables also provide additional information in the neural network.

Figure 9.3b shows the angle of the leptonically decaying top quark. For the $t\bar{t}$ sample it peaks at ± 1 as expected, for stop samples it tends to have also lower values, but the shape differences are not as clear as for the \cancel{E}_T .

The behaviour of the other variables does not differ much from the variables for a centre of mass energy of 8 TeV, all distributions are shown in the appendix A.9.

9.2. Optimisation

For the optimisation the same preselection and the same optimisation procedure (cf. chapter 6.2) was used as for 8 TeV. During the optimisation the variables, which are listed in table 9.2, show the best behaviour for the high mass region with a centre of mass energy of 14 TeV.

Table 9.2: List of variables which show the best behaviour for the neural network. They are ranked depending on their importance for the MLP.

Rank	Variable	Importance
1	m_T	14.130
2	\cancel{E}_T	4.250
3	$m_{t_{had}}$	3.171
4	$\Delta\varphi(\cancel{E}_T, j_1)$	2.372
5	$\cancel{E}_T \perp t_{lep}$	1.628
6	$m_{t\bar{t}}$	1.235
7	$ \cos\theta(t_{lep}) $	0.556
8	$p_T^{t\bar{t}}$	0.534
9	$ \cos\theta(t_{had}) $	0.214
10	$\Delta\varphi(\cancel{E}_T, t_{lep})$	0.005

The linear correlation coefficients for these variables are shown in figure 9.4 for signal and background. As one can see, most variables are not strongly correlated to each other. These observables offer new information about the event type and help in the separation. Additionally, there are variables which are strongly correlated for signal but not for background, or vice versa. The perpendicular \cancel{E}_T component to the leptonically decaying top quark, for example, is more strongly correlated to the transverse mass for the signal than for the

background. Using variables with different correlations for signal and background improves the separation power.

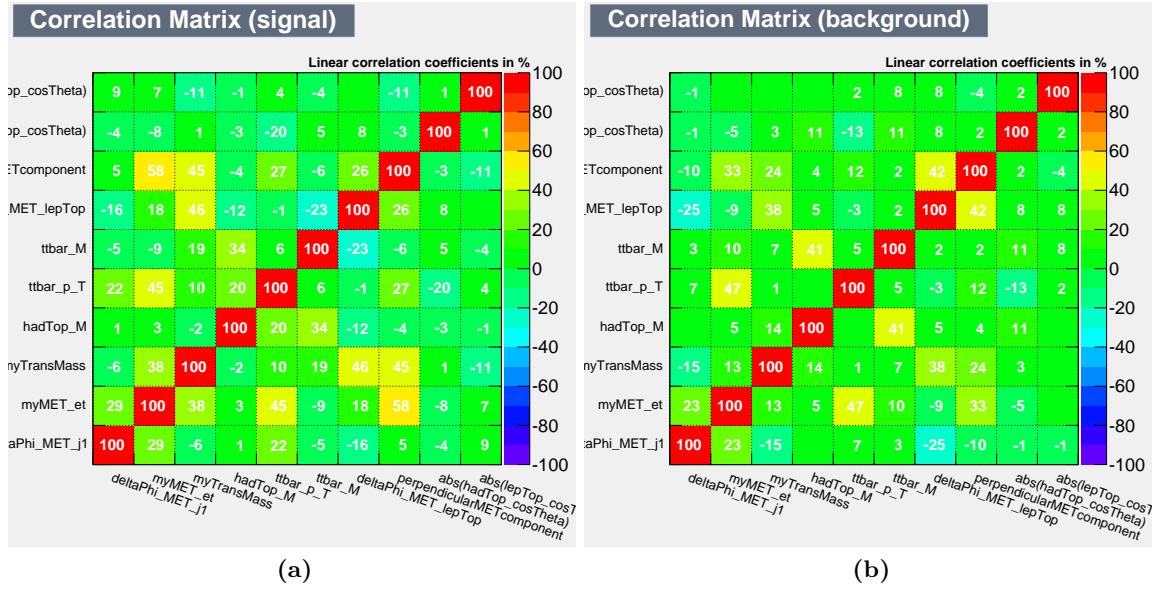


Figure 9.4: Linear correlation coefficients for the remaining variables after the optimisation. The linear correlation coefficients are shown for both signal (a) and background (b).

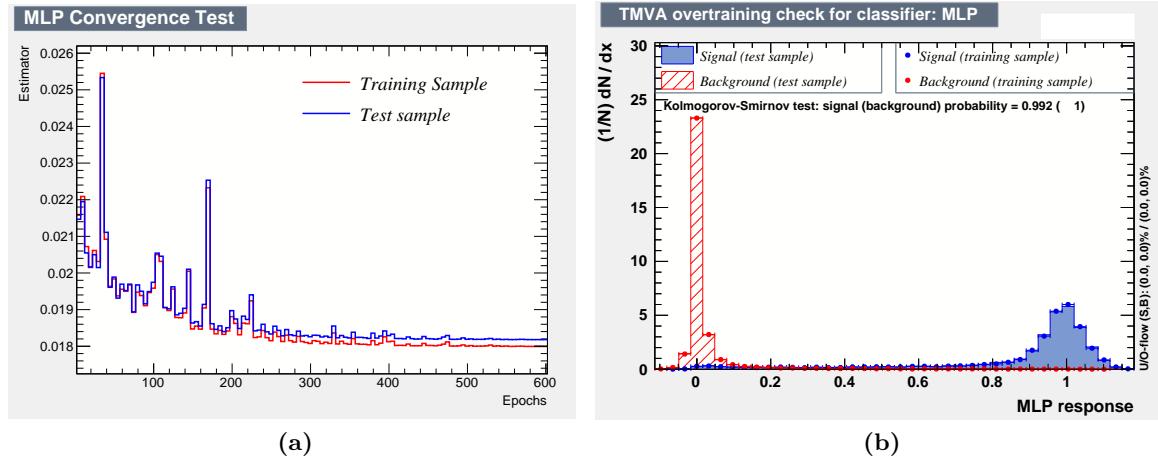


Figure 9.5: Convergence and overtraining check for the MLP configuration. The estimator for the test and the training sample is shown as a function of the number of epochs (a). The MLP distribution is shown for the test and the training sample (b).

In figure 9.5a one can see that, when using this configuration, the evaluation of the neural network converges for both test and training sample. The resulting neural network distribution is shown in figure 9.5b. No sign of overtraining can be seen, as both test and training sample look similar. Considering the Kolmogorov-Smirnov test one receives a probability of 0.992 for the signal and of 1 for the background of the agreement of test and training sample. These values are both nearly 1 and therefore this configuration shows no overtraining.

Figure 9.6 displays the background rejection, which is nearly at 1 for most MLP cut values, so one can reject most of all background events. The signal efficiency is at 1 for low MLP values and has a small slope between 0 and ~ 0.9 . From 0.9 on the signal efficiency decreases strongly. The significance values peak before the signal efficiency drops too strongly. The numbers of signal and background events – N_S and N_B – are calculated for different cuts at the MLP distribution and used to calculate the significance for an integrated luminosity of 20 fb^{-1} (cf. chapter 6.2). The significances shown in the figure correspond to the MLP cut values on the x-axis.

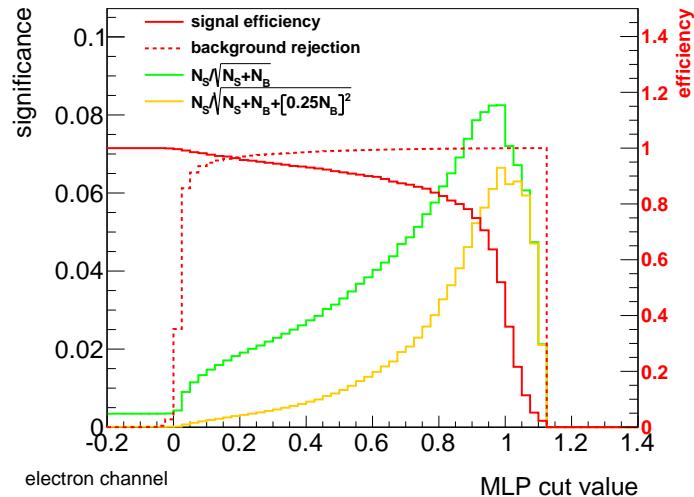


Figure 9.6: The signal efficiency (red) and background rejection (red dashed) depending on the MLP value are shown for the centred stop sample ($\tilde{t}(1500, 400)$). Additionally the significances for cuts on the MLP output $N_S/\sqrt{N_S+N_B}$ (green) and $N_S/\sqrt{N_S+N_B+\sigma_B^2}$ (yellow) are shown.

9.3. Optimisation with a different preselection

A strong separation in the shape distribution for the shown optimisation can be observed. In figure 9.8a the number of expected events for a signal and the $t\bar{t}$ sample is shown. The maximum N_S/N_B ratio is 0.01, which is too low to achieve any sensitivity. A possible reason for this bad N_S/N_B ratio could be that the neural network trains against many background events which strongly differ from the signal events.

Requiring an additional \cancel{E}_T cut for the preselection leads to fewer background events. The number of expected events for the signal sample is mostly not influenced. The improved N_S/N_B ratio due to the additional preselection cut – $\cancel{E}_T > 250 \text{ GeV}$ – can be seen when comparing both distributions in figure 9.7.

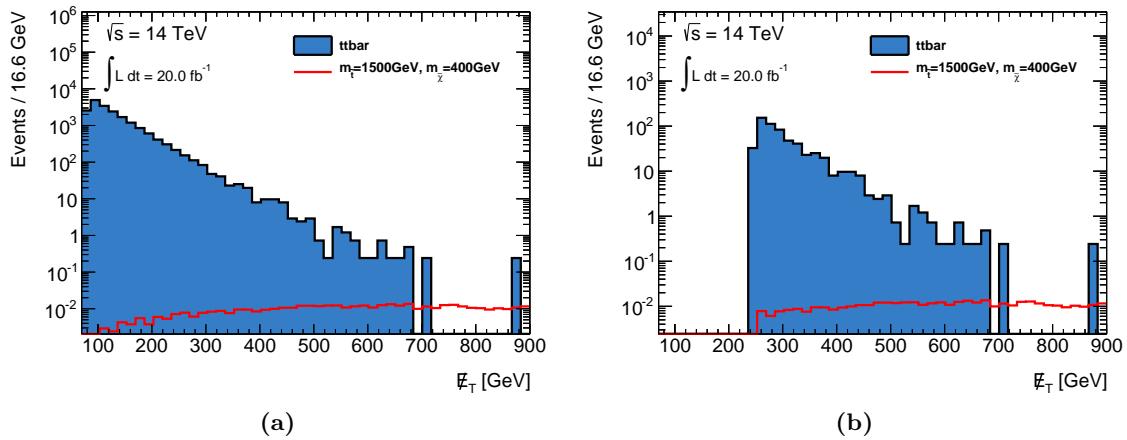


Figure 9.7: Expected number of events for an integrated luminosity of 20 fb^{-1} showed in the \cancel{E}_T distribution for 14 TeV . The number of expected events before (a) and after (b) the \cancel{E}_T cut.

Using the new preselection one can again optimise a neural network. This neural network is trained especially against those events which are similar between the signal and the $t\bar{t}$ background. The variables presented in table 9.3 are the optimal choice for the neural network. The linear correlation coefficients for these variables are shown in the appendix A.9. The convergence test, the overtraining check as well as the efficiencies and the significances can be found there, too. All these distributions are similar to the previous optimisation. In figure 9.8 the expected number of events for the MLP distributions for both optimisations with and without the additional \cancel{E}_T cut are displayed. The plot on the left shows the distribution for the standard preselection, which leads to low N_S/N_B ratios up to 0.01. Because of the small stop cross section only few signal events are expected even for $\sqrt{s} = 14 \text{ TeV}$. The N_S/N_B ratio using the additional \cancel{E}_T cut is around ~ 0.1 , which is an enhancement of a factor of 10 with respect to the other neural network. This enhancement is due to the stronger preselection and therefore a neural network which especially trains against the signal-like $t\bar{t}$ events.

Table 9.3: List of variables which show the best performance for the neural network. They are ranked depending on their importance for the MLP.

Rank	Variable	Importance
1	m_T	9.028
2	$\Delta\varphi(\cancel{E}_T, j_1)$	8.238
3	\cancel{E}_T	6.585
4	$m_{t\bar{t}had}$	4.389
5	$\Delta\varphi(\cancel{E}_T, t_{had})$	2.783
6	$\cancel{E}_T \perp t_{lep}$	2.020
7	$m_{t\bar{t}}$	1.479
8	$\Delta\varphi(\cancel{E}_T, j_2)$	0.909
9	$\cancel{E}_T/\sqrt{H_T}$	0.795
10	$\min \Delta\varphi(\cancel{E}_T, j_{1/2})$	0.426
11	$\Delta\varphi(\cancel{E}_T, t_{lep})$	0.079
12	$p_T^{t\bar{t}}$	0.059

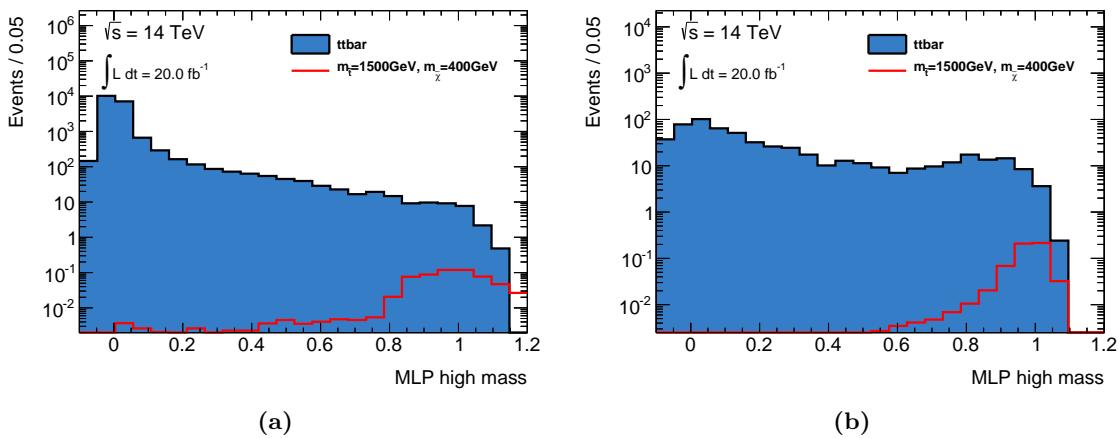


Figure 9.8: Expected number of events for an integrated luminosity of 20 fb^{-1} showed in the MLP distribution for the standard preselection (a) and the additional cut of $\cancel{E}_T > 250 \text{ GeV}$ (b).

9.4. Reoptimisation for the high mass region using $\sqrt{s} = 8 \text{ TeV}$

A good improvement in the N_S/N_B is visible for the neural network when using a stronger preselection for the Monte Carlo samples generated with a center of mass energy of 14 TeV. In order to check a similar reoptimisation for a center of mass energy of 8 TeV, a stronger preselection was applied. For the stronger preselection the missing transverse energy and the transverse mass cut are increased to

$$\cancel{E}_T > 150 \text{ GeV} \text{ and } m_T > 120 \text{ GeV}. \quad (45)$$

All events which pass the new, stronger preselection are used in order to train a new neural network. The variables which show the best behaviour for the neural network are listed in table 9.4. The linear correlation coefficients for these variables, the convergence test, the overtraining check, the efficiencies and the significances can be found in the appendix A.9.

Table 9.4: List of variables which show the best performance for the neural network. They are ranked depending on their importance for the MLP.

Rank	Variable	Importance
1	tau-veto	17.47
2	track-veto	17.18
3	$\Delta\varphi(\cancel{E}_T, j_1)$	8.899
4	$p_T^{t\bar{t}}/\cancel{E}_T$	8.672
5	\cancel{E}_T	8.325
6	m_T	3.811
7	$\cancel{E}_T \perp t_{lep}$	3.75
8	m_{thad}	3.00
9	$m_{t\bar{t}}$	2.65
10	$\Delta\varphi(\cancel{E}_T, t_{had})$	2.384
11	topness	0.3868

In table 9.5 the significances are given for the optimised neural network using both the standard preselection and the new preselection with the increased cuts for the missing transverse energy and the transverse mass. As one can see, all significances rise when using the new preselection around 20%.

In figure 9.9 the resulting neural network distributions for the standard and for the new preselection are shown for a signal sample and the $t\bar{t}$ and the $W+jets$ background. The distributions are scaled to an integrated luminosity of 21 fb^{-1} in order to compare the expected number of events. The signal distribution for the new preselection peaks strongly for high neural network values around 1. For the neural network with the standard preselection, the signal distribution is flatter with only a small peak for high neural network values. In addition, the distribution for both dominant backgrounds – $t\bar{t}$ and $W+jets$ – are flatter for the new preselection. The stronger peak for low neural network values for the standard

Table 9.5: Significances for the optimisation of a neural network using the standard preselection and using the preselection with additional cuts of $\cancel{E}_T > 150 \text{ GeV}$ and $m_T > 120 \text{ GeV}$.

Sample ($m_{\tilde{t}}, m_{\tilde{\chi}_1^0}$)	standard preselection	new preselection
$\tilde{t} (750, 100)$	0.76	0.91
$\tilde{t} (750, 150)$	0.76	0.91
$\tilde{t} (750, 50)$	0.77	0.93
$\tilde{t} (700, 100)$	1.15	1.26
$\tilde{t} (800, 100)$	0.55	0.72

preselection derives mostly from the events which are not very signal-like and get cut away using the tighter preselection.

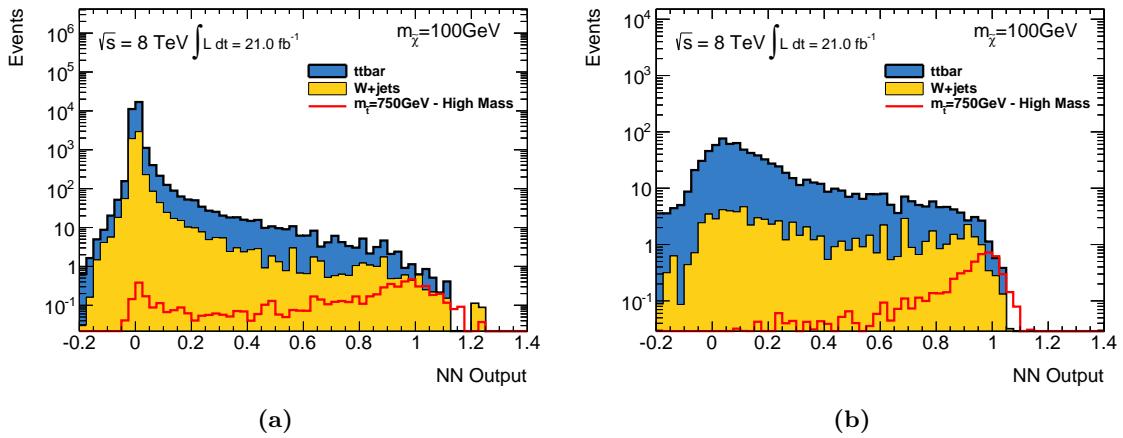


Figure 9.9: Expected number of events for an integrated luminosity of 21 fb^{-1} showed in the MLP distribution for the standard preselection (a) and the stronger preselection using a cut of $\cancel{E}_T > 150 \text{ GeV}$ and $m_T > 120 \text{ GeV}$ (b).

9.5. Results and conclusion

In the variable distributions (cf. chapter 9.1 or appendix A.9), there are differences between signal and $t\bar{t}$ background for samples using a center of mass energy of 14 TeV. Within a neural network these differences can be used to achieve a good separation, as shown in figure 9.5b.

Using simulated data samples with a center of mass energy of 14 TeV, a harder preselection was applied before the training of the neural network in order to improve the separation. This new preselection removes many $t\bar{t}$ and $W+\text{jets}$ events which are easy to differentiate from the signal events. After this additional cut, the neural network is optimised especially

against signal-like $t\bar{t}$ and $W+jets$ events which leads to an enhancement of the separation power and the N_S/N_B ratio.

In addition to this study, a tighter preselection was applied for the Monte Carlo samples using a center of mass energy of 8 TeV. Afterwards the events which survive the new preselection, namely $\cancel{E}_T > 150 \text{ GeV}$ and $m_T > 120 \text{ GeV}$, are used for the optimisation of a new neural network. Using the new neural network, the significances increase around 20% in comparison to the neural network using the standard preselection.

In a dedicated study an optimal preselection should be found in order to further improve the results as preparation for run II.

10. Conclusion and outlook

Supersymmetry is a possible extension of the Standard Model of particle physics; it can solve the hierarchy problem and it provides a possible candidate for dark matter. This analysis is designed to search for Supersymmetry in a simplified model. This model assumes the stop quark to be the lightest quark and is therefore able to conserve the naturalness. In addition, it assumes the lightest supersymmetric particle, the neutralino (LSP), to be stable. The model conserves the R parity, therefore stop quarks are only produced in pairs. They each decay into a top quark and a neutralino which is only visible as extra missing transverse energy.

Neural networks are used in order to separate between the supersymmetric signal and the two main backgrounds – $t\bar{t}$ and $W+jets$. For the optimisation, three different analysis regions are defined. One region is for high stop masses, the so-called high mass region. One is near the boundary of the decay, the so-called diagonal region. In between is the so-called bulk region. For each of these regions one neural network is optimised. For the optimisation, a strategy is developed to increase both the significance after a simple cut at the neural network distribution, and the stability of the neural network.

During the optimisation, different variables show good separation power for the single regions. The angular distributions are sensitive to the differences for the diagonal region, the kinematic variables improve the separation for the high mass region.

Comparing the optimised neural network to the cuts used before, an improvement of 20% to 120% is achieved in the significances. Using the neural network distribution and comparing the Standard Model prediction to the measured data, no excess is visible.

Therefore the neural network distributions are used in order to set limits on the stop and the neutralino masses. During the limit setting procedure, a data-driven method is used in order to determine the normalisation of the $t\bar{t}$ and the $W+jets$ background.

Compared to the official ATLAS results, the resulting limits using the full neural network distribution are stronger for the diagonal and the bulk region. For the high mass region the limits are similar for low neutralino masses and also become stronger for higher neutralino masses.

As preparation for run II, the optimisation procedure was applied to simulated data samples using a center of mass energy of 14 TeV. Using a stronger preselection, and therefore optimising a neural network against more signal-like events, improves the N_S/N_B ratio. A second optimisation was performed with a tighter preselection for Monte Carlo samples with a center of mass energy of 8 TeV, too. For the high mass region, the significances increase around 20% compared to the neural network with the standard preselection. For further improvement, a better preselection could be found and used for the complete optimisation procedure.

With the higher statistics and the higher center of mass energy for run II, new variables should be tested in order to improve the sensitivity, e.g. the additional ISR jet requirement. In the context of run II the analysis can be improved by a reoptimised preselection and a better measurement of the sensitivity. Instead of the significance calculation using a simplified consideration for background systematics, an approximation for the CL_s values can be used.

References

- [1] M. E. Peskin and D. V. Schroeder, *Quantum field theory*, Perseus Book, New York (1995).
- [2] S. Weinberg, *The quantum theory of fields*. Cambridge university press, 1996.
- [3] **Particle Data Group** Collaboration, J. Beringer et al., *Review of particle physics*, *Phys. Rev. D* **86** (Jul, 2012) 010001.
- [4] **ATLAS** Collaboration, *Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC*, *Physics Letters B* **716** (2012), no. 1 1 – 29.
- [5] ESA, *Planck reveals an almost perfect universe*, Juli, 2014.
- [6] H. Baer, C.-h. Chen, F. Paige, and X. Tata, *Signals for minimal supergravity at the cern large hadron collider: Multijet plus missing energy channel*, *Physical Review D* **52** (1995), no. 5 2746.
- [7] S. P. Martin, *A Supersymmetry primer*, [hep-ph/9709356](#).
- [8] J.-L. Caron, “Overall view of LHC experiments.. Vue d’ensemble des expériences du LHC..” AC Collection. Legacy of AC. Pictures from 1992 to 2002., May, 1998.
- [9] J. Pequenao, “Computer generated image of the whole ATLAS detector.” Mar, 2008.
- [10] J. Pequenao, “Computer generated image of the ATLAS inner detector.” Mar, 2008.
- [11] J. Pequenao, “Computer Generated image of the ATLAS calorimeter.” Mar, 2008.
- [12] J. Pequenao, “Event Cross Section in a computer generated image of the ATLAS detector..” Mar, 2008.
- [13] **ATLAS** Collaboration, G. Aad et al., *Search for top squark pair production in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector*, [arXiv:1407.0583](#).
- [14] Y. Abulaiti, M. Barisonzi, P. Butti, C. Deluca, T. Eifert, T. Golling, P. Grenier, A. Henrichs, P. Mättig, B. Nachman, P. Pani, S. Pataraia, K. Rosbach, S. Strandberg, V. Tsiskaridze, and X. Wang, *Search for top squark pair production in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 8$ TeV pp collisions using 20.3 fb^{-1} of ATLAS data*, Tech. Rep. ATL-COM-PHYS-2013-1490, CERN, Geneva, Nov, 2013.
- [15] **ATLAS** Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, *Eur. Phys. J. C* **73** (2013) 2518, [[arXiv:1302.4393](#)].
- [16] **GEANT4** Collaboration, S. Agostinelli et al., *GEANT4: A simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.

- [17] **ATLAS** Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATLAS-PHYS-PUB-2010-013, (2010).
- [18] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, [[arXiv:0709.2092](#)].
- [19] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, et al., *The CT10 NNLO Global Analysis of QCD*, *Phys. Rev. D* **89** (2014) 033009, [[arXiv:1302.6246](#)].
- [20] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1*, *Comput.Phys.Commun.* **184** (2013) 919, [[hep-ph/0405247](#)].
- [21] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07** (2002) 012, [[hep-ph/0201195](#)].
- [22] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5 : Going Beyond*, *JHEP* **06** (2011) 128, [[arXiv:1106.0522](#)].
- [23] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al., *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007, [[arXiv:0811.4622](#)].
- [24] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, [[arXiv:0710.3820](#)].
- [25] M. Bahr et al., *Herwig++ Physics and Manual*, *Eur. Phys. J. C* **58** (2008) 639, [[arXiv:0803.0883](#)].
- [26] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP* **04** (2008) 063, [[arXiv:0802.1189](#)].
- [27] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, *Tech. Rep. ATLAS-LARG-PUB-2008-002, CERN*, 2011.
- [28] <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTaggingCalibrationDataInterface>.
- [29] **ATLAS** Collaboration, *Measurement of the b-tag efficiency in a sample of jets containing muons with 5 fb^{-1} of data from the atlas detector*, ATLAS-CONF-2012-043, (2012).
- [30] <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/BtagAnalysis172Winter2013>.
- [31] D. Börner, P. Mättig, and S. Fleischmann, *Studien zur Unterdrückung von Untergrundprozessen in Zerfällen des supersymmetrischen Top-Squark in ein Top-Quark und ein Neutralino*, bachelorarbeit, Wuppertal U., Jun, 2012. Presented 26 Sep 2012.

- [32] M. L. Graesser and J. Shelton, *Hunting Asymmetric Stops*, *Phys. Rev. Lett.* **111** (2013) 121802, [[arXiv:1212.4495](https://arxiv.org/abs/1212.4495)].
- [33] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, *TMVA: Toolkit for Multivariate Data Analysis*, *PoS ACAT* (2007) 040, [[physics/0703039](https://arxiv.org/abs/physics/0703039)].
- [34] M. Barisonzi et al., *Search for top squark pair production in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 8$ TeV pp collisions using 21 ifb of ATLAS data*, . support note for ATLAS-CONF-2013-037.
- [35] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, *eConf C0303241* (2003) MOLT007, [[physics/0306116](https://arxiv.org/abs/physics/0306116)].
- [36] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf, *The roostats project*, *arXiv preprint arXiv:1009.1003* (2010).
- [37] **ROOT** Collaboration, K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, Tech. Rep. CERN-OPEN-2012-016, New York U., New York, Jan, 2012.
- [38] J. Neyman and E. S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*. Springer, 1992.
- [39] A. L. Read, *Presentation of search results: the cl s technique*, *Journal of Physics G: Nuclear and Particle Physics* **28** (2002), no. 10 2693.
- [40] <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/EfficiencyMeasurements>.
- [41] “Energy scale resolution recommendations.” <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EnergyScaleResolutionRecommendations>.
- [42] “Atlas muon combined performance guidelines for analyses of 2012 data.” <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MCPAnalysisGuidelinesData2012>.
- [43] **ATLAS** Collaboration, *ATLAS Inner Detector: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997.
- [44] **ATLAS** Collaboration, N. Wermes and G. Hallewell, *ATLAS Pixel Detector: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1998.
- [45] M. Capeans, G. Darbo, K. Einsweiller, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet, *ATLAS Insertable B-Layer Technical Design Report*, Tech. Rep. CERN-LHCC-2010-013. ATLAS-TDR-19, CERN, Geneva, Sep, 2010.

- [46] **ATLAS** Collaboration, *ATLAS Liquid-argon Calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996.
- [47] **ATLAS** Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996.
- [48] **ATLAS** Collaboration, *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1997. distribution.
- [49] **ATLAS** Collaboration, *ATLAS level-1 trigger: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1998.
- [50] **ATLAS** Collaboration, P. Jenni, M. Nessi, M. Nordberg, and K. Smith, *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 2003.

A. Appendix

A.1. Simulated data samples

Process	Name	ID
$t\bar{t}$	PowhegPythia P2011C	117050
$W+jets$	WenuMassiveCBPt0 BFfilter	167740
	WenuMassiveCBPt0 CJetFilterBVeto	167741
	WenuMassiveCBPt0 CJetVetoBVeto	167742
	WmunuMassiveCBPt0 BFfilter	167743
	WmunuMassiveCBPt0 CJetFilterBVeto	167744
	WmunuMassiveCBPt0 CJetVetoBVeto	167745
	WtaunuMassiveCBPt0 BFfilter	167746
	WtaunuMassiveCBPt0 CJetFilterBVeto	167747
	WtaunuMassiveCBPt0 CJetVetoBVeto	167748
	WenuMassiveCBPt70 140 BFfilter	167761
	WenuMassiveCBPt70 140 CJetFilterBVeto	167762
	WenuMassiveCBP70 140 CJetVetoBVeto	167763
	WmunuMassiveCBPt70 140 BFfilter	167764
	WmunuMassiveCBPt70 140 CJetFilterBVeto	167765
	WmunuMassiveCBPt70 140 CJetVetoBVeto	167766
	WtaunuMassiveCBPt70 140 BFfilter	167767
	WtaunuMassiveCBPt70 140 CJetFilterBVeto	167768
	WtaunuMassiveCBPt70 140 CJetVetoBVeto	167769
	WenuMassiveCBPt140 280 BFfilter	167770
	WenuMassiveCBPt140 280 CJetFilterBVeto	167771
	WenuMassiveCBP140 280 CJetVetoBVeto	167772
	WmunuMassiveCBPt140 280 BFfilter	167773
	WmunuMassiveCBPt140 280 CJetFilterBVeto	167774
	WmunuMassiveCBPt140 280 CJetVetoBVeto	167775
	WtaunuMassiveCBPt140 280 BFfilter	167776
	WtaunuMassiveCBPt140 280 CJetFilterBVeto	167777
	WtaunuMassiveCBPt140 280 CJetVetoBVeto	167778
	WenuMassiveCBPt280 500 BFfilter	167779
	WenuMassiveCBPt280 500 CJetFilterBVeto	167780
	WenuMassiveCBP280 500 CJetVetoBVeto	167781
	WmunuMassiveCBPt280 500 BFfilter	167782

Process	Name	ID
	WmunuMassiveCBPt280 500 CJetFilterBVeto	167783
	WmunuMassiveCBPt280 500 CJetVetoBVeto	167784
	WtaunuMassiveCBPt280 500 BFfilter	167785
	WtaunuMassiveCBPt280 500 CJetFilterBVeto	167786
	WtaunuMassiveCBPt280 500 CJetVetoBVeto	167787
	WenuMassiveCBPt500 BFfilter	167788
	WenuMassiveCBPt500 CJetFilterBVeto	167789
	WenuMassiveCBP500 CJetVetoBVeto	167790
	WmunuMassiveCBPt500 BFfilter	167791
	WmunuMassiveCBPt500 CJetFilterBVeto	167792
	WmunuMassiveCBPt500 CJetVetoBVeto	167793
	WtaunuMassiveCBPt500 BFfilter	167794
	WtaunuMassiveCBPt500 CJetFilterBVeto	167795
	WtaunuMassiveCBPt500 CJetVetoBVeto	167796
	WenuMassiveCBPt40 70 BFfilter	180534
	WenuMassiveCBPt40 70 CJetFilterBVeto	180535
	WenuMassiveCBPt40 70 CJetVetoBVeto	180536
	WmunuMassiveCBPt40 70 BFfilter	180537
	WmunuMassiveCBPt40 70 CJetFilterBVeto	180538
	WmunuMassiveCBPt40 70 CJetVetoBVeto	180539
	WtaunuMassiveCBPt40 70 BFfilter	180540
	WtaunuMassiveCBPt40 70 CJetFilterBVeto	180541
	WtaunuMassiveCBPt40 70 CJetVetoBVeto	180542
single top	t-channel	110101
	s-channel	110119
	Wt-channel	110140
$t\bar{t}W$	ttbarW	119353
	ttbarWjExcl	174830
	ttbarWjjIncl	174831
$t\bar{t}Z$	ttbarZ	119355
	ttbarZjExcl	174832
	ttbarZjjIncl	174833
$t\bar{t}WW$	ttbarWW	119583
tZ	tZ, Wt-channel	179991
	tZ, st-channel	17992

A.2. Cuts from the official analysis

Cuts for all three regions from the official analysis [34].

Requirement	diagonal	bulk	high mass
$\Delta\varphi(\cancel{E}_T, j_1) >$	0.8	–	0.8
$\Delta\varphi(\cancel{E}_T, j_2) >$	0.8	0.8	0.8
$\cancel{E}_T [\text{GeV}] >$	150	200	225
$\cancel{E}_T/\sqrt{H_T} [\text{GeV}^{1/2}] >$	8	13	11
$m_T [\text{GeV}] >$	140	140	180
$m_T [\text{GeV}] <$	250	–	–
$am_{T2} [\text{GeV}] >$	–	170	200
$m_{T2}^\tau [\text{GeV}] >$	–	–	120
$m_{t_{had}} [\text{GeV}] >$	130	130	130
$m_{t_{had}} [\text{GeV}] <$	205	205	205

A.3. Distributions with and without an ISR jet

The following distributions are shown for the diagonal region and the $t\bar{t}$ and the $W+\text{jets}$ background with an additional jet with $p_T > 150 \text{ GeV}$ (a) and without an additional ISR jet (b).

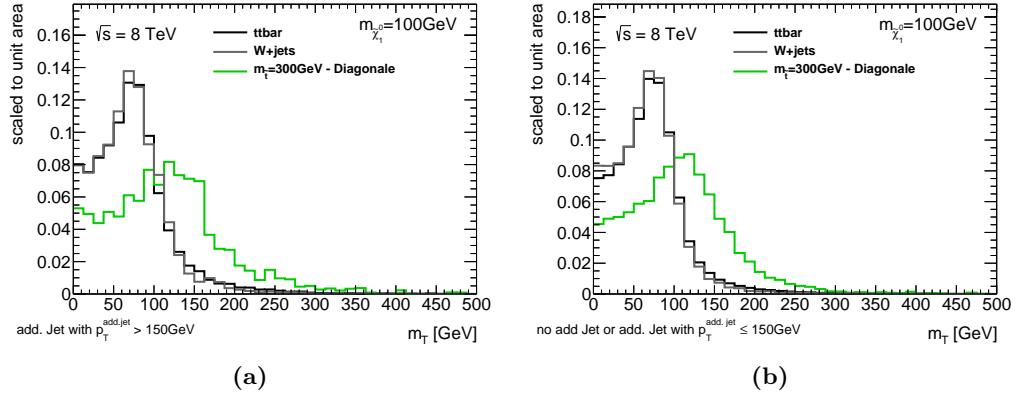


Figure A.1: Transverse mass m_T .

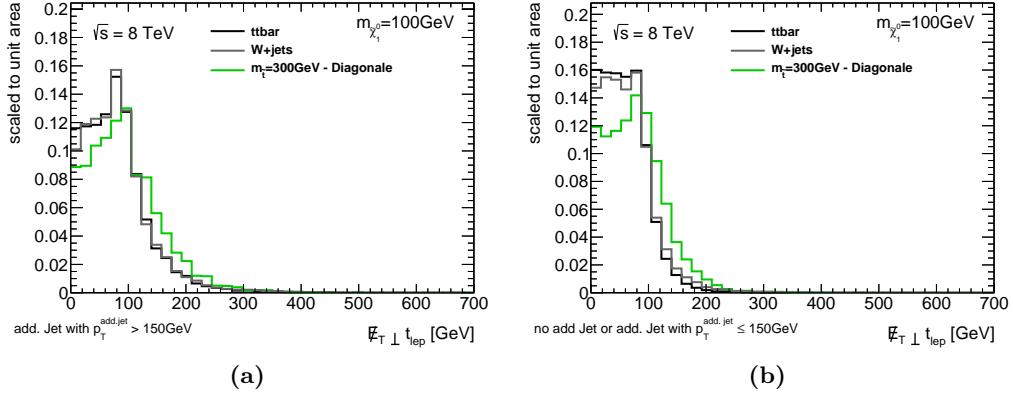


Figure A.2: Perpendicular component of the missing transverse energy to the leptonically decaying top quark $\cancel{E}_T \perp t_{lep}$.

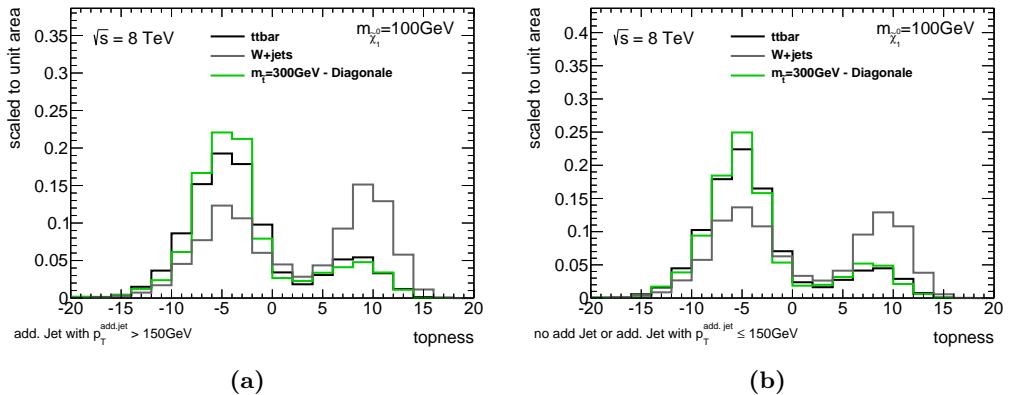


Figure A.3: Topness.

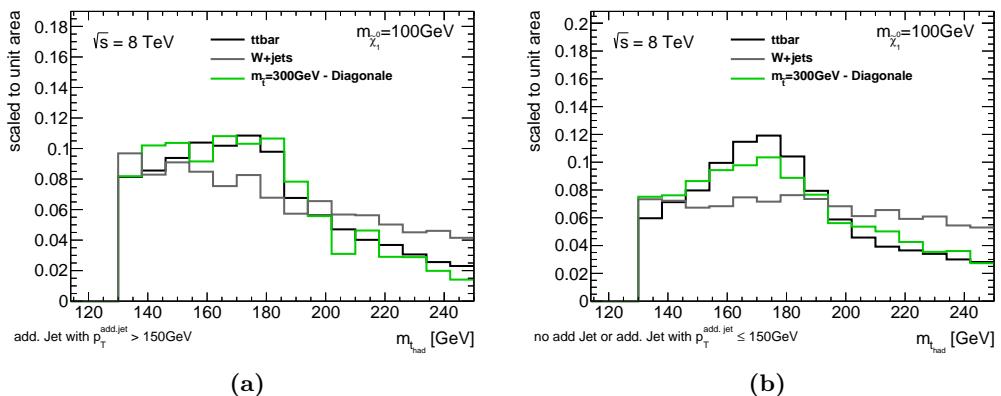
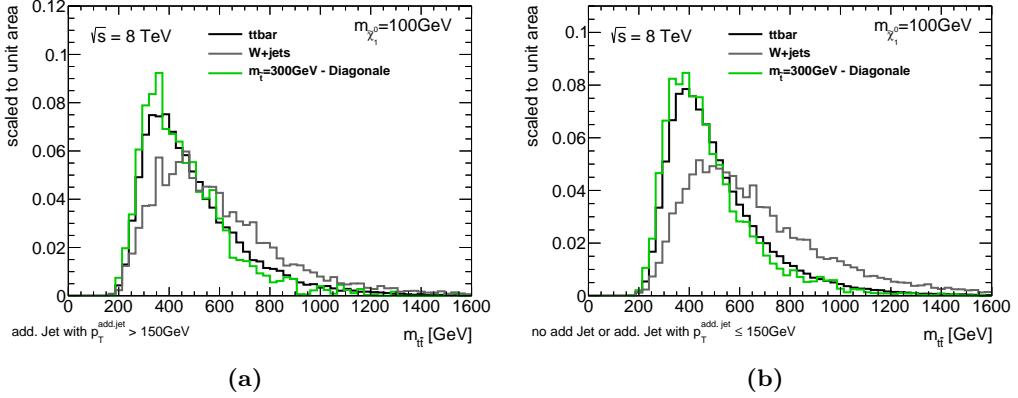
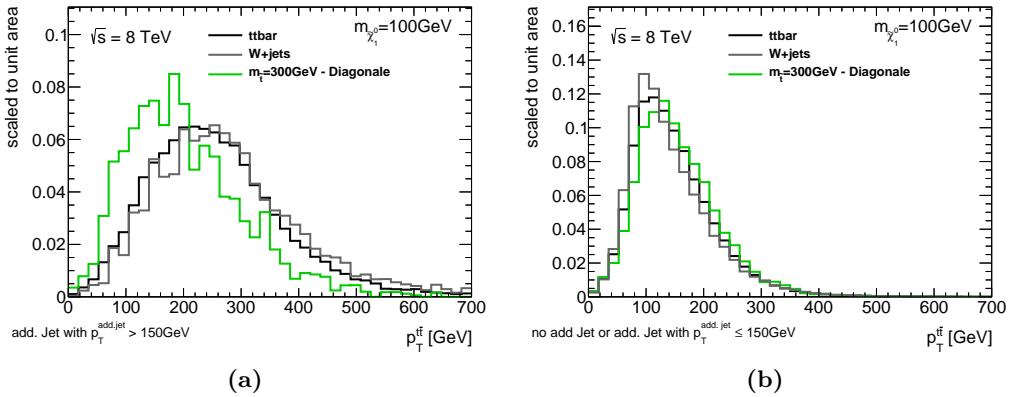
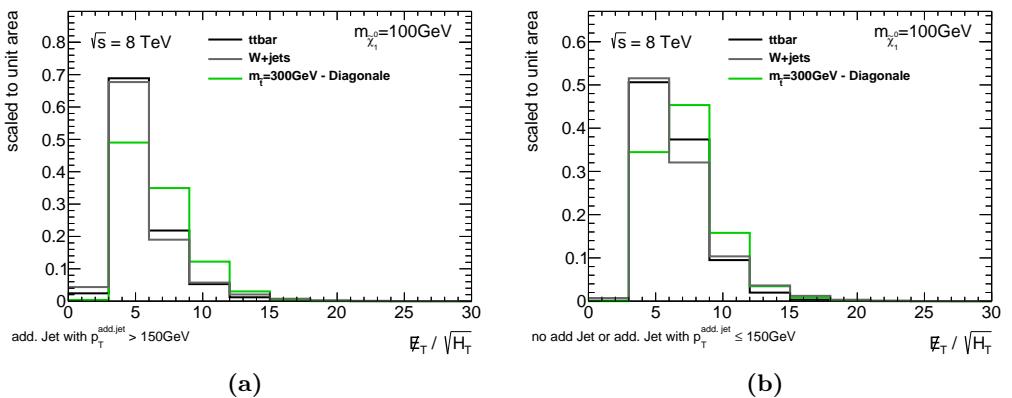
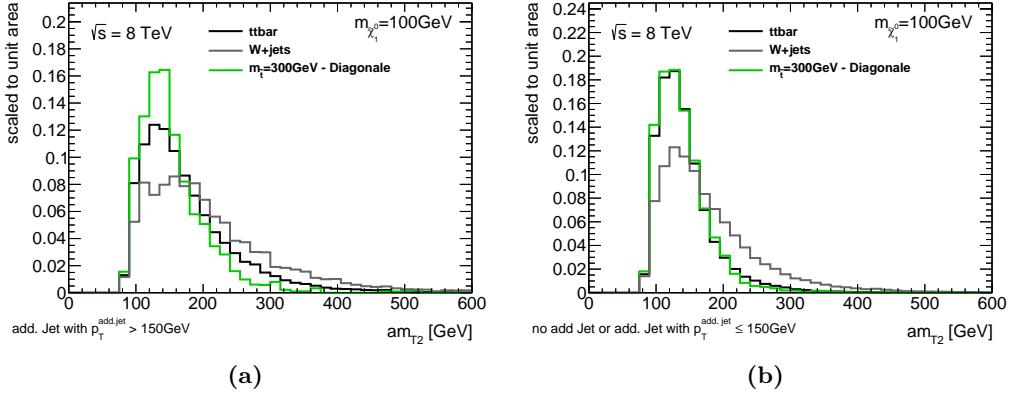
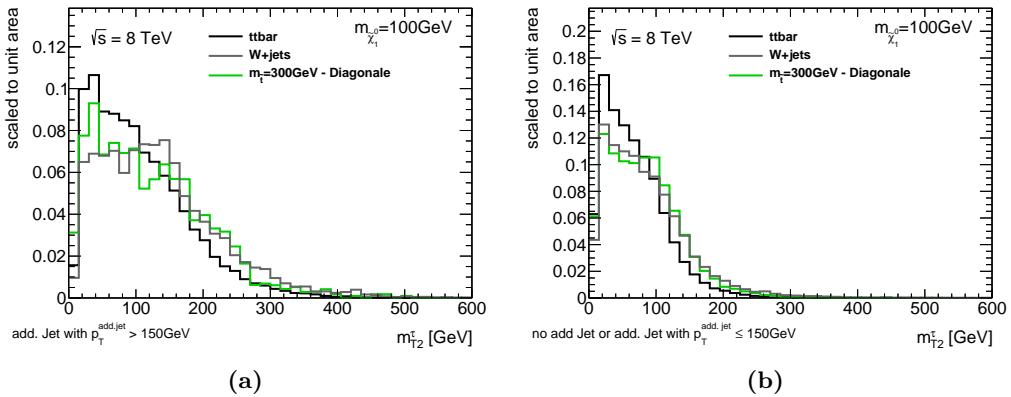
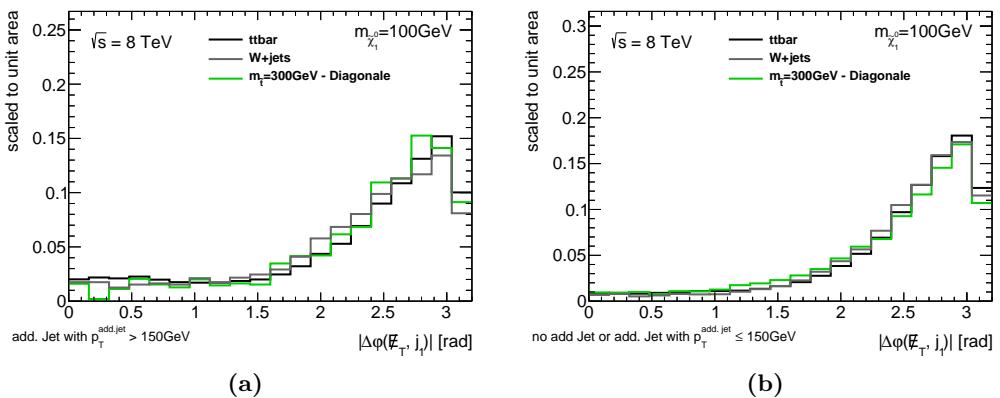


Figure A.4: Mass of the hadronically decaying top quark $m_{t_{had}}$.

**Figure A.5:** Mass of the $t\bar{t}$ system $m_{t\bar{t}}$.**Figure A.6:** Momentum of the $t\bar{t}$ system $p_T^{t\bar{t}}$.**Figure A.7:** \cancel{E}_T significance.

**Figure A.8:** Stransverse mass am_{T2} .**Figure A.9:** Stransverse mass m_{T2}^τ .**Figure A.10:** Angle between the \not{E}_T and the leading jet.

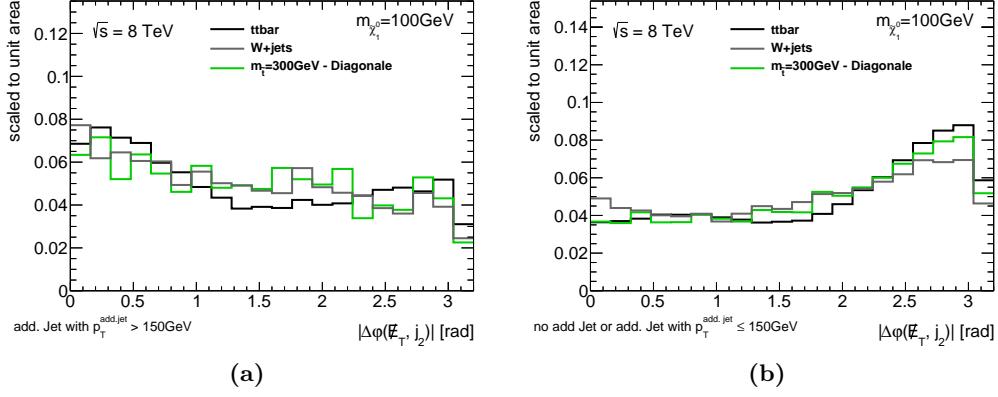


Figure A.11: Angle between the \not{E}_T and the second leading jet.

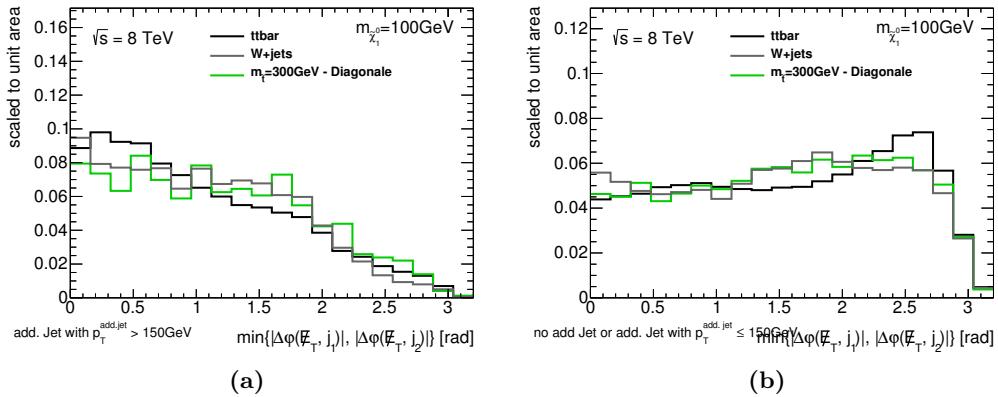


Figure A.12: Minimal angle between the \not{E}_T and the two leading jets.

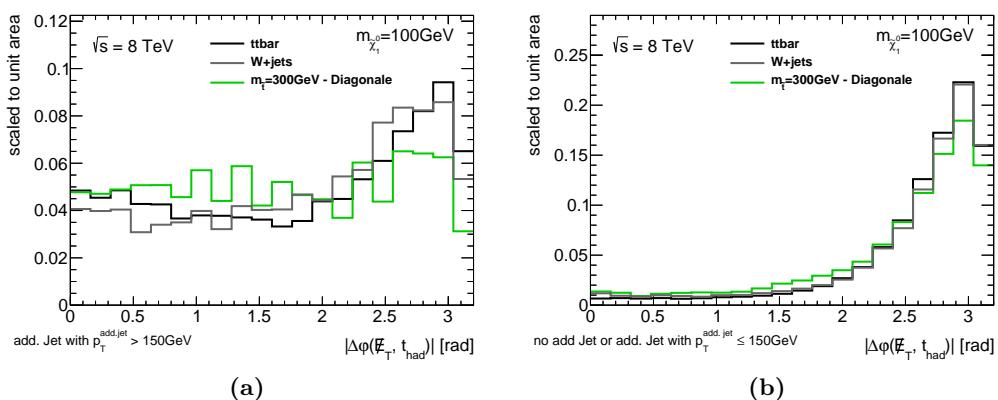


Figure A.13: Angle between the \not{E}_T and the hadronically decaying top quark.

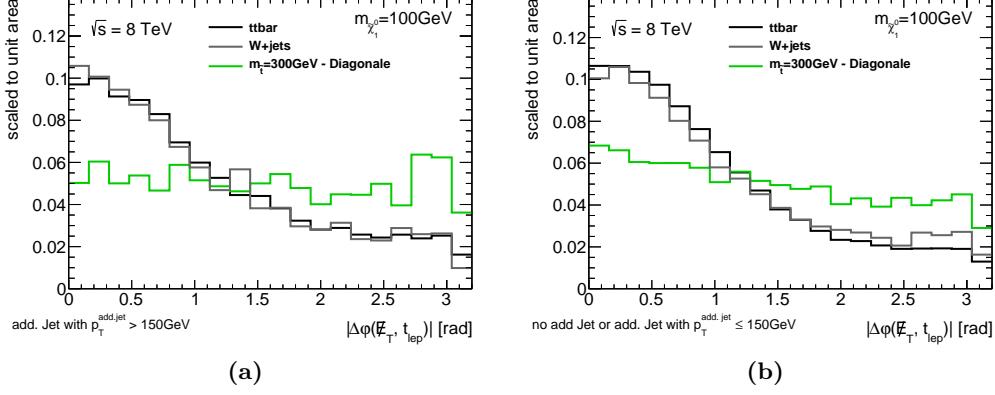


Figure A.14: Angle between the \not{E}_T and the leptonically decaying top quark.

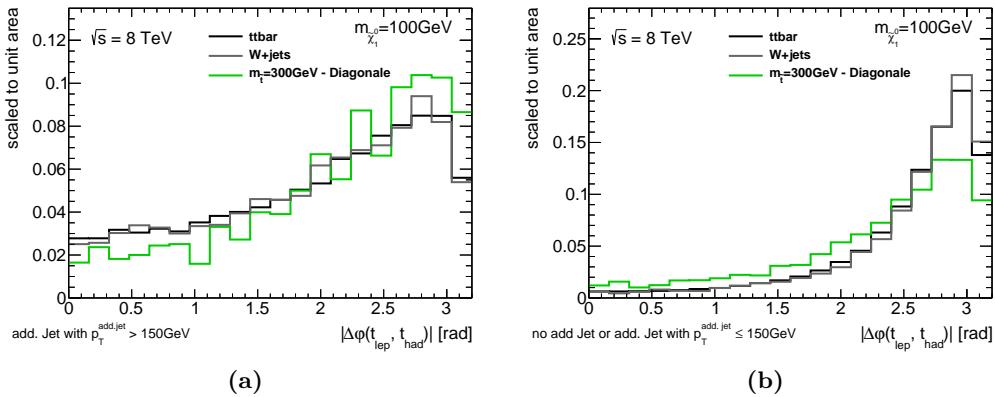


Figure A.15: Transverse angle between both top quarks $\Delta\varphi(t_{had}, t_{lep})$.

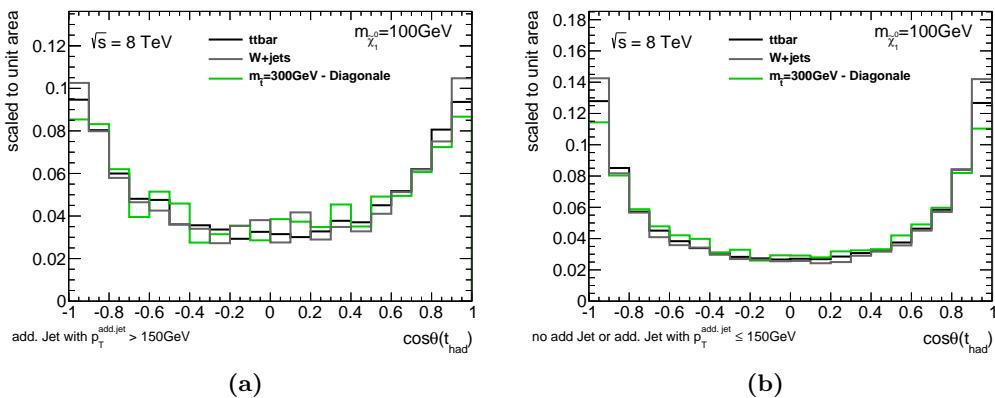


Figure A.16: Angle of the hadronically decaying top quark $\cos\theta_{t_{had}}$.

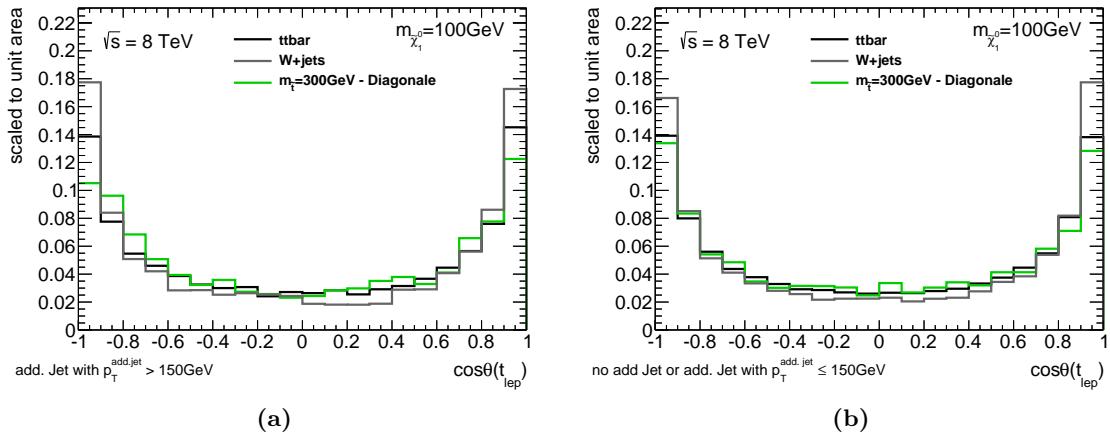


Figure A.17: Angle of the leptonically decaying top quark $\cos \theta_{t_{lep}}$.

A.4. Optimisation results for the muon channel

Bulk region The results for the bulk region using the muon channel are presented similar to the electron channel.

Table A.2: List of variables which show the best performance in the neural network for the bulk region, ranked depending on their importance for the MLP.

Rank	Variable	Importance
1	m_T	37.18
2	$m_{t_{had}}$	18.83
3	am_{T2}	14.79
4	m_{T2}^τ	11.9
5	\cancel{E}_T	10.25
6	$p_T^{t\bar{t}}$	7.065
7	$\cancel{E}_T \perp t_{lep}$	4.749
8	$\cancel{E}_T / \sqrt{H_T}$	4.062
9	$\Delta\varphi(t_{lep}, t_{had})$	0.3929
10	$\Delta\varphi(\cancel{E}_T, j_2)$	0.2769
11	topness	0.007138
12	$\min \left\{ \Delta\varphi(\cancel{E}_T, j_{1/2}) \right\}$	0.001717

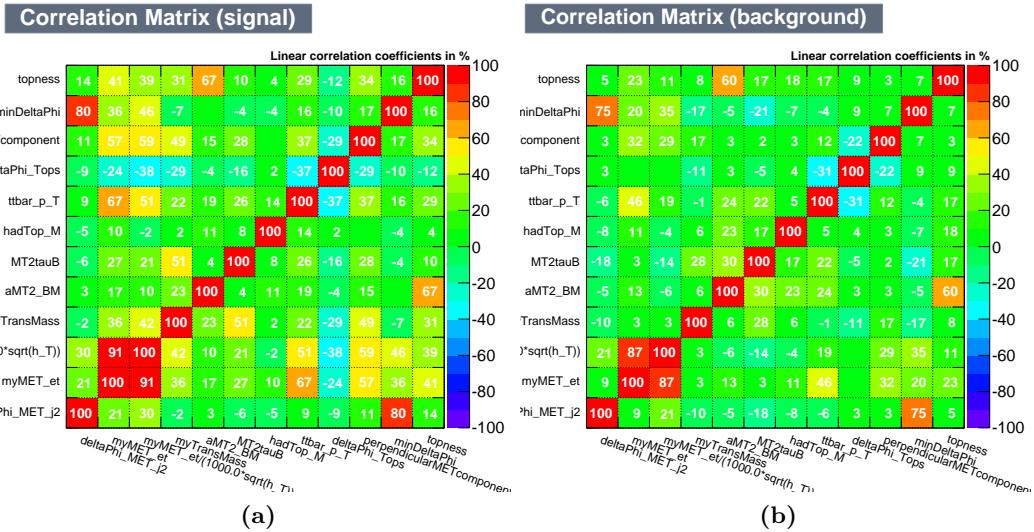


Figure A.18: Linear correlation coefficients for the remaining variables after the optimisation. The linear correlation coefficients are shown for both signal (a) and background (b).

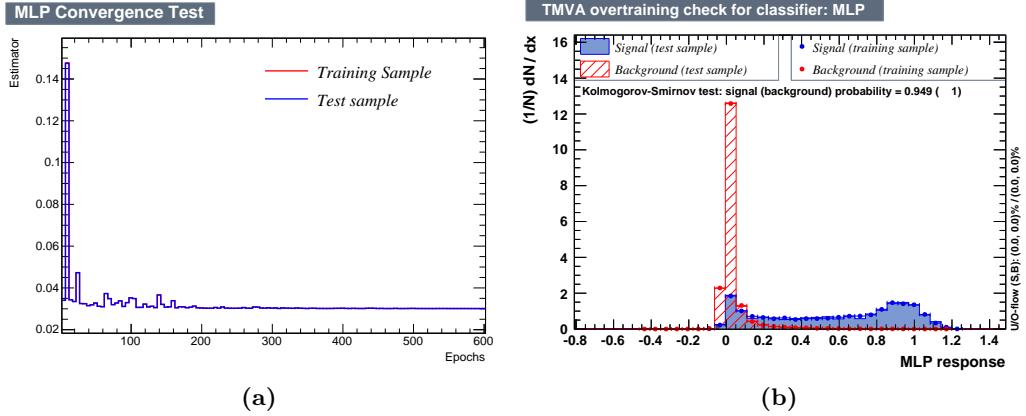


Figure A.19: Convergence (a) and overtraining check (b) for the MLP configuration of the bulk region.

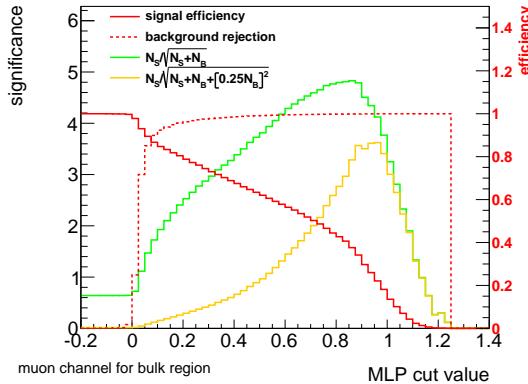


Figure A.20: The signal efficiency (red), the background rejection (red dashed) and two significances (yellow/green) depending on the MLP value are shown for the central signal sample ($\tilde{t}(500, 100)$).

Table A.3: Significances for the old cuts (cf. appendix A.2), using all old variables and using the optimised neural network.

Sample $(m_{\tilde{t}}, m_{\tilde{\chi}_1^0})$	cuts	MLP old variables	optimised MLP
(500, 100)	2.32	3.3	3.62
(450, 50)	2.91	4.65	4.98
(500, 200)	1.36	2.14	2.31

Diagonal region The results for the diagonal region using the muon channel are presented similar to the electron channel.

Table A.4: List of variables which show the best performance in the neural network for the diagonal region, ranked depending on their importance for the MLP.

Rank	Variable	Importance
1	m_T	20.02
2	tau-veto	9.223
3	$\cancel{E}_T / \sqrt{H_T}$	1.14
4	\cancel{E}_T	0.7535
5	$m_{t\bar{t}}$	0.1448
6	add. ISR jet	0.02026
7	$\cancel{E}_T \perp t_{lep}$	0.01212
8	$\Delta\varphi(\cancel{E}_T, j_1)$	0.009136
9	$\Delta\varphi(t_{lep}, t_{had})$	0.007115
10	$m_{t_{had}}$	0.006209

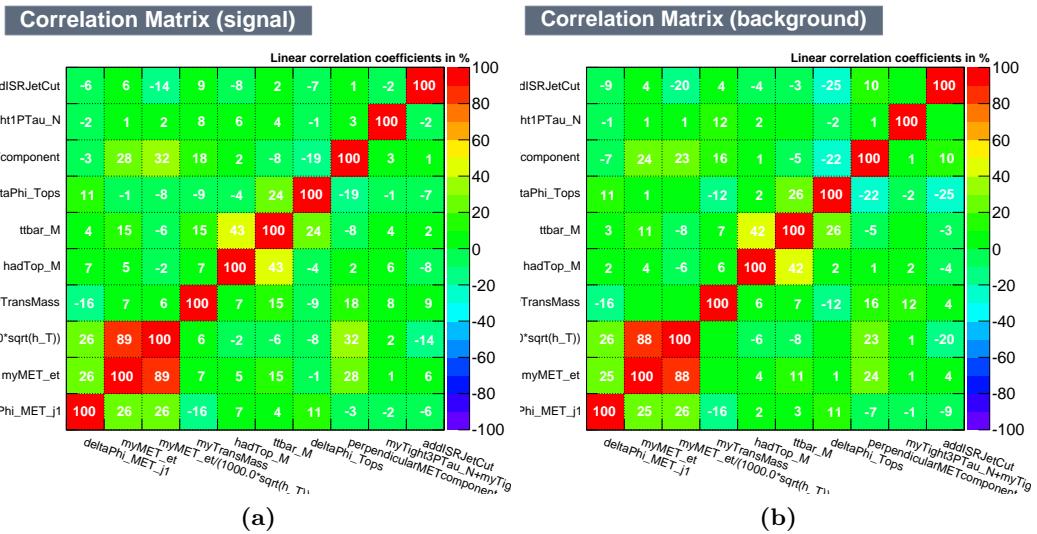


Figure A.21: Linear correlation coefficients for the remaining variables after the optimisation. The linear correlation coefficients are shown for both signal (a) and background (b).

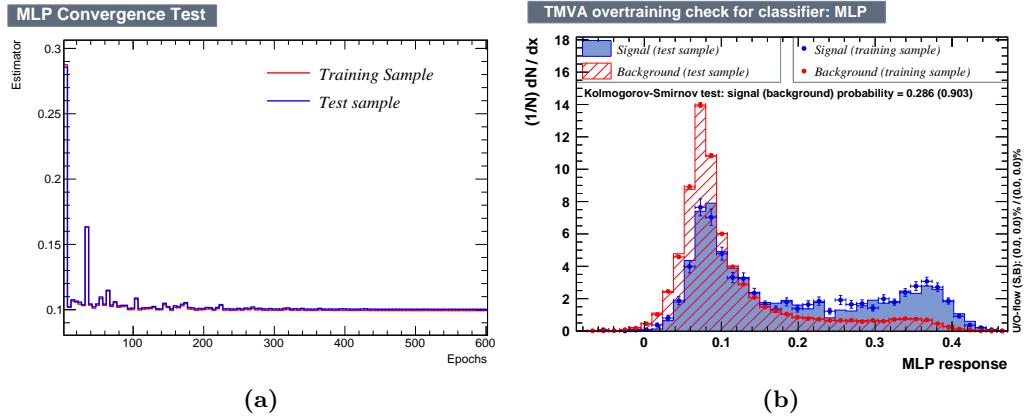


Figure A.22: Convergence (a) and overtraining check (b) for the MLP configuration of the diagonal region.

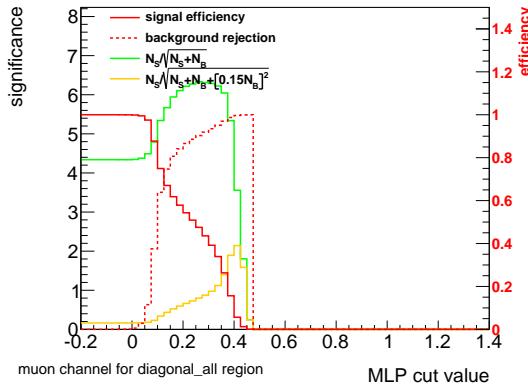


Figure A.23: The signal efficiency (red), the background rejection (red dashed) and two significances (yellow/green) depending on the MLP value are shown for the central signal sample ($\tilde{t}(300, 100)$).

Table A.5: Significances for the old cuts (cf. appendix A.2), using all old variables and using the optimised neural network.

Sample $(m_{\tilde{t}}, m_{\tilde{\chi}_1^0})$	cuts	MLP old variables	optimised MLP
$\tilde{t}(300, 100)$	1.58	1.88	2.14
$\tilde{t}(250, 50)$	1.62	2.55	2.81
$\tilde{t}(200, 1)$	0.81	1.18	1.31
$\tilde{t}(350, 150)$	1.01	1.11	1.3
$\tilde{t}(400, 200)$	0.79	0.72	0.92
$\tilde{t}(450, 250)$	0.36	0.35	0.45

High mass region The results for the high mass region using the muon channel are presented similar to the electron channel.

Table A.6: List of variables which show the best performance in the neural network for the high mass region, ranked depending on their importance for the MLP.

Rank	Variable	Importance
1	tau-veto	23.36
2	m_T	16.78
3	$p_T^{t\bar{t}}/\cancel{E}_T$	9.729
4	\cancel{E}_T	8.107
5	$\Delta\varphi(\cancel{E}_T, t_{had})$	3.313
6	$\cancel{E}_T/\sqrt{H_T}$	2.441
7	$\cancel{E}_T \perp t_{lep}$	2.424
8	$m_{t\bar{t}}$	2.214
9	$\Delta\varphi(\cancel{E}_T, j_2)$	0.6663
10	topness	0.2524

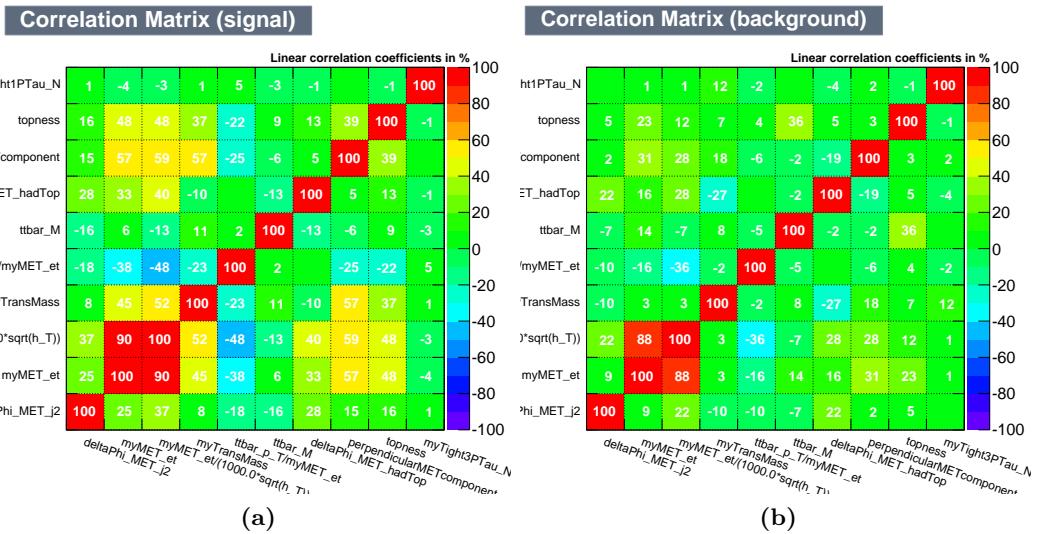


Figure A.24: Linear correlation coefficients for the remaining variables after the optimisation. The linear correlation coefficients are shown for both signal (a) and background (b).

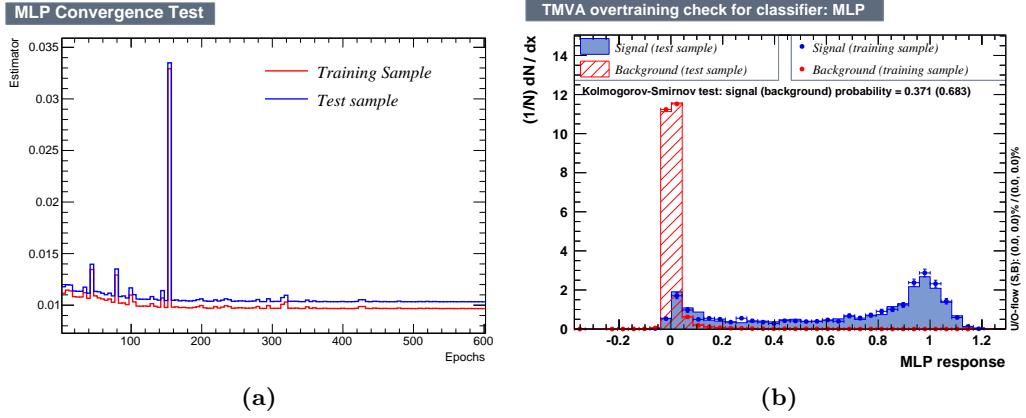


Figure A.25: Convergence (a) and overtraining check (b) for the MLP configuration of the high mass region.

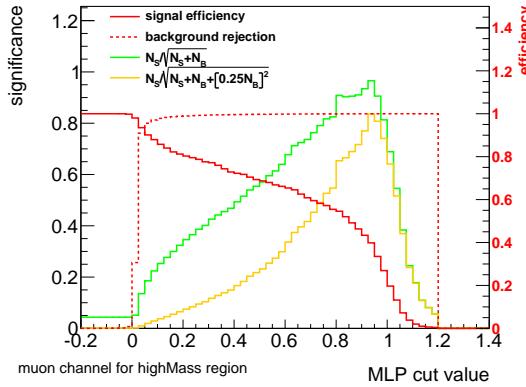


Figure A.26: The signal efficiency (red), the background rejection (red dashed) and two significances (yellow/green) depending on the MLP value are shown for the central signal sample ($\tilde{t}(750, 100)$).

Table A.7: Significances for the old cuts (cf. appendix A.2), using all old variables and using the optimised neural network.

Sample $(m_{\tilde{t}}, m_{\tilde{\chi}_1^0})$	cuts	MLP old variables	optimised MLP
$\tilde{t}(750, 100)$	0.66	0.83	0.84
$\tilde{t}(750, 150)$	0.39	0.77	0.76
$\tilde{t}(750, 50)$	0.43	0.81	0.84
$\tilde{t}(700, 100)$	0.36	0.97	1.22
$\tilde{t}(800, 100)$	0.28	0.54	0.57

A.5. Neural network distributions for all backgrounds

The neural network distributions for all backgrounds and a signal sample for the electron channel (a) and the muon channel (b).

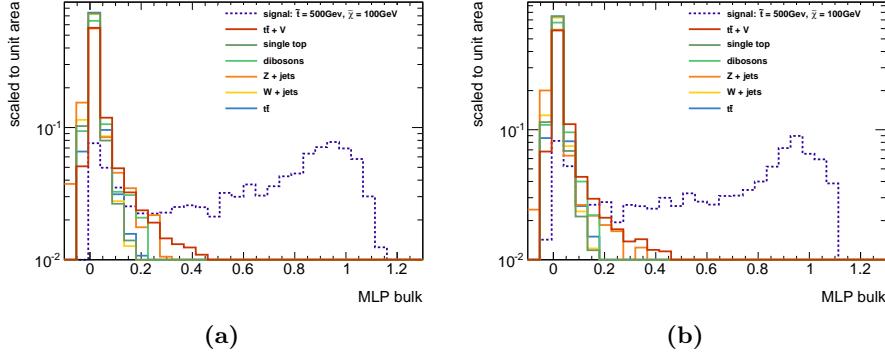


Figure A.27: Bulk region.

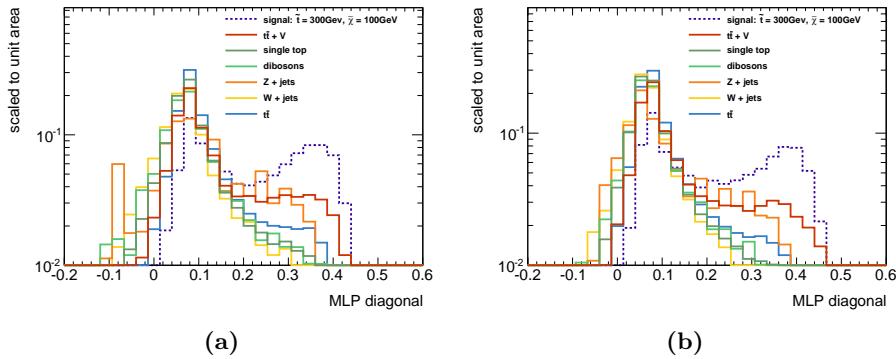


Figure A.28: Diagonal region.

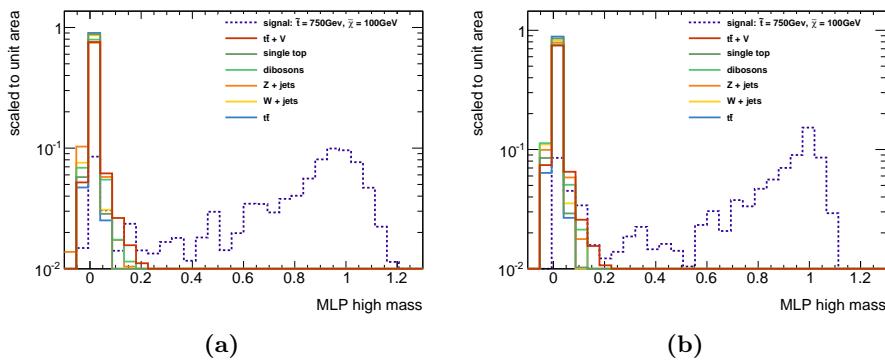


Figure A.29: High mass region.

A.6. Determination of normalisation factors

Diagonal region The normalisation factors for the diagonal region determined by `HistFitter`. The calculated values should match the given values which are used in order to produce the pseudo data.

Table A.8: `HistFitter` output for the determination of the W +jets normalisation factor. The μ_W should be the same value as it was used for producing the pseudo data (equation 27).

given μ_W	calculated μ_W
0.7	0.70 ± 0.04
1.0	1.00 ± 0.04
1.3	1.30 ± 0.04

Table A.9: `HistFitter` output for the determination of the W +jets normalisation using additional signal for the pseudo data. The μ_W should be the same value as it was used for producing the pseudo data (equation 28).

given μ_W	calculated μ_W
0.7	0.70 ± 0.04
1.0	1.00 ± 0.04
1.3	1.30 ± 0.04

Table A.10: `HistFitter` output for the determination of the W +jets and the $t\bar{t}$ normalisation factors. The μ_W and the $\mu_{t\bar{t}}$ should be the same as it was used for producing the pseudo data (equation 29).

given μ_W	given $\mu_{t\bar{t}}$	calculated μ_W	calculated $\mu_{t\bar{t}}$
0.7	0.7	0.70 ± 0.03	0.70 ± 0.01
	1.0	0.70 ± 0.04	1.00 ± 0.01
	1.3	0.70 ± 0.04	1.30 ± 0.01
1.0	0.7	1.00 ± 0.03	0.70 ± 0.01
	1.0	1.00 ± 0.04	1.00 ± 0.01
	1.3	1.00 ± 0.04	1.30 ± 0.01
1.3	0.7	1.30 ± 0.04	0.70 ± 0.01
	1.0	1.30 ± 0.04	1.00 ± 0.01
	1.3	1.30 ± 0.04	1.30 ± 0.01

Table A.11: HistFitter output for the determination of the W +jets and the $t\bar{t}$ normalisation factors using additional signal for the pseudo data. The μ_W and the $\mu_{t\bar{t}}$ should be the same as it was used for producing the pseudo data (equation 30).

given μ_W	given $\mu_{t\bar{t}}$	calculated μ_W	calculated $\mu_{t\bar{t}}$
0.7	0.7	0.70 ± 0.03	0.70 ± 0.01
	1.0	0.70 ± 0.04	1.00 ± 0.01
	1.3	0.70 ± 0.04	1.30 ± 0.01
1.0	0.7	1.00 ± 0.04	0.70 ± 0.01
	1.0	1.00 ± 0.04	1.00 ± 0.01
	1.3	1.00 ± 0.04	1.30 ± 0.01
1.3	0.7	1.30 ± 0.04	0.70 ± 0.01
	1.0	1.30 ± 0.04	1.00 ± 0.01
	1.3	1.30 ± 0.04	1.30 ± 0.01

High mass region The normalisation factors for the high mass region determined by HistFitter. The calculated values should match the given values which are used in order to produce the pseudo data.

Table A.12: HistFitter output for the determination of the W +jets normalisation factor. The μ_W should be the same value as it was used for producing the pseudo data (equation 27).

given μ_W	calculated μ_W
0.7	0.70 ± 0.11
1.0	1.00 ± 0.11
1.3	1.30 ± 0.11

Table A.13: HistFitter output for the determination of the W +jets normalisation factor using additional signal for the pseudo data. The μ_W should be the same value as it was used for producing the pseudo data (equation 28).

given μ_W	calculated μ_W
0.7	0.70 ± 0.11
1.0	1.00 ± 0.11
1.3	1.30 ± 0.11

Table A.14: `HistFitter` output for the determination of the W +jets and the $t\bar{t}$ normalisation factors. The μ_W and the $\mu_{t\bar{t}}$ should be the same as it was used for producing the pseudo data (equation 29).

given μ_W	given $\mu_{t\bar{t}}$	calculated μ_W	calculated $\mu_{t\bar{t}}$
0.7	0.7	0.70 ± 0.16	0.70 ± 0.04
	1.0	0.70 ± 0.17	1.00 ± 0.04
	1.3	0.70 ± 0.19	1.30 ± 0.05
1.0	0.7	1.00 ± 0.17	0.70 ± 0.04
	1.0	1.00 ± 0.18	1.00 ± 0.05
	1.3	1.00 ± 0.20	1.30 ± 0.05
1.3	0.7	1.30 ± 0.18	0.70 ± 0.04
	1.0	1.30 ± 0.19	1.00 ± 0.05
	1.3	1.30 ± 0.21	1.30 ± 0.05

Table A.15: `HistFitter` output for the determination of the W +jets and the $t\bar{t}$ normalisation factors using additional signal for the pseudo data. The μ_W and the $\mu_{t\bar{t}}$ should be the same as it was used for producing the pseudo data (equation 30).

given μ_W	given $\mu_{t\bar{t}}$	calculated μ_W	calculated $\mu_{t\bar{t}}$
0.7	0.7	0.70 ± 0.16	0.70 ± 0.04
	1.0	0.70 ± 0.18	1.00 ± 0.04
	1.3	0.70 ± 0.19	1.30 ± 0.05
1.0	0.7	1.00 ± 0.17	0.70 ± 0.04
	1.0	1.00 ± 0.18	1.00 ± 0.04
	1.3	1.00 ± 0.20	1.30 ± 0.05
1.3	0.7	1.30 ± 0.18	0.70 ± 0.04
	1.0	1.30 ± 0.19	1.00 ± 0.05
	1.3	1.30 ± 0.21	1.30 ± 0.05

A.7. Validation of the simulation

All distributions are shown for data with an integrated luminosity of 20.6 fb^{-1} and all Standard Model backgrounds stacked and scaled to this luminosity. The shown error are the statistical error and the systematic error due to the JES and JER uncertainty. The lower part of the plots shows the ratio of data to the Standard Model prediction.

Comparison of different normalisation factors

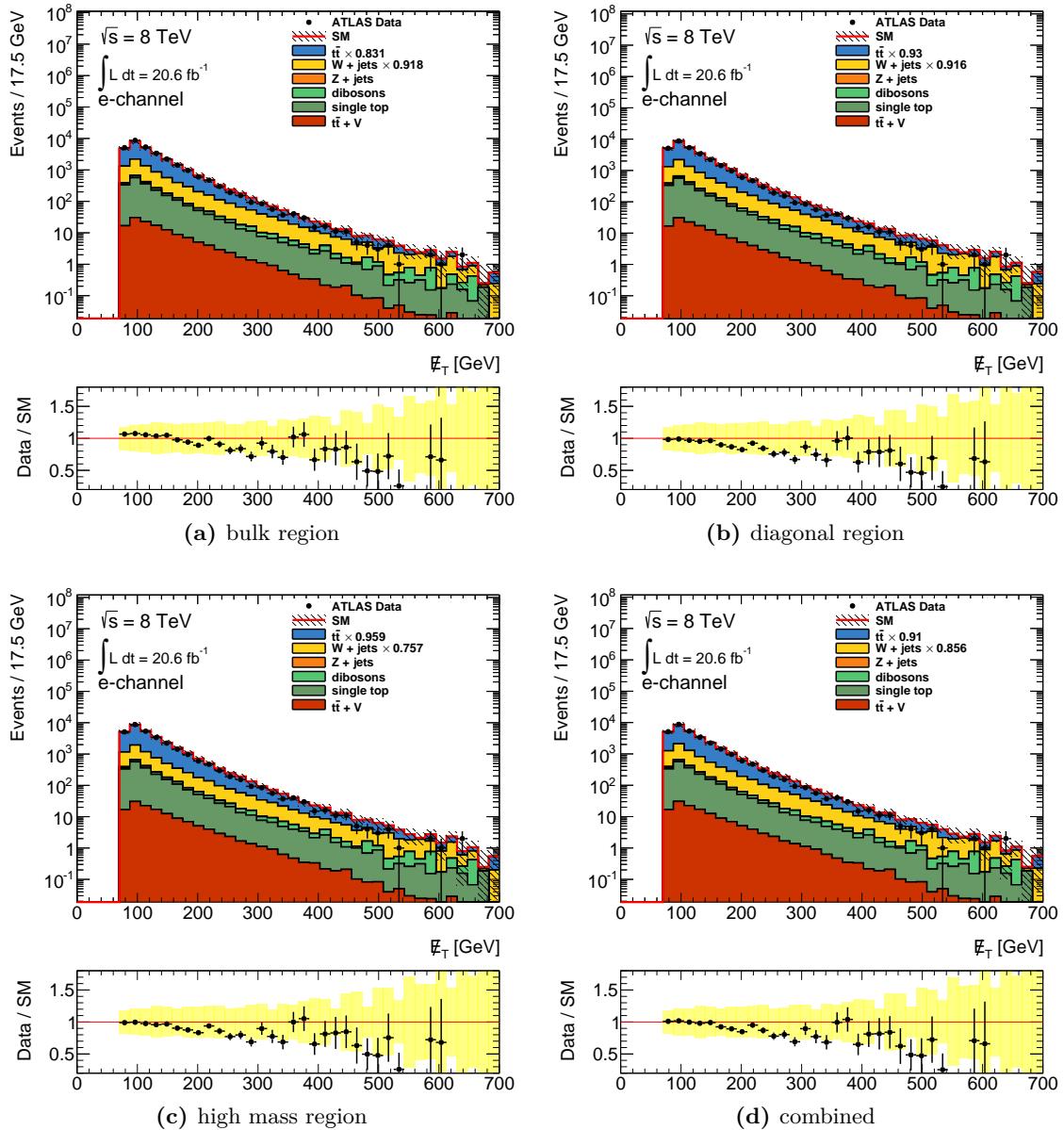


Figure A.30: The missing transverse energy distribution for the three different scaling factors from each region and for the mean value as combination.

Comparison of data and simulation The following distributions are shown using the combined normalisation factor for the electron channel (a) and the muon channel (b).

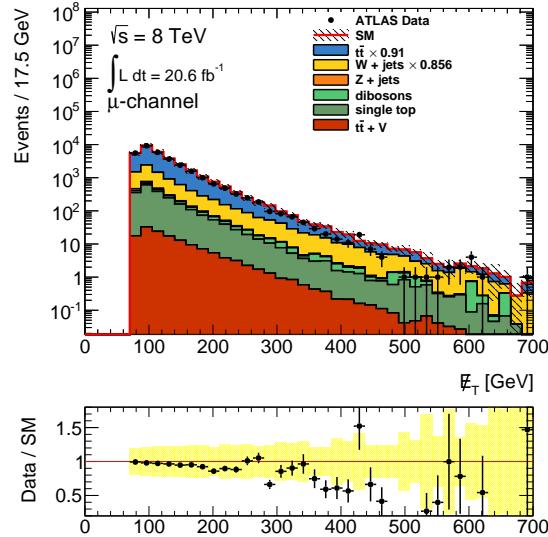


Figure A.31: The missing transverse energy \cancel{E}_T for the muon channel.

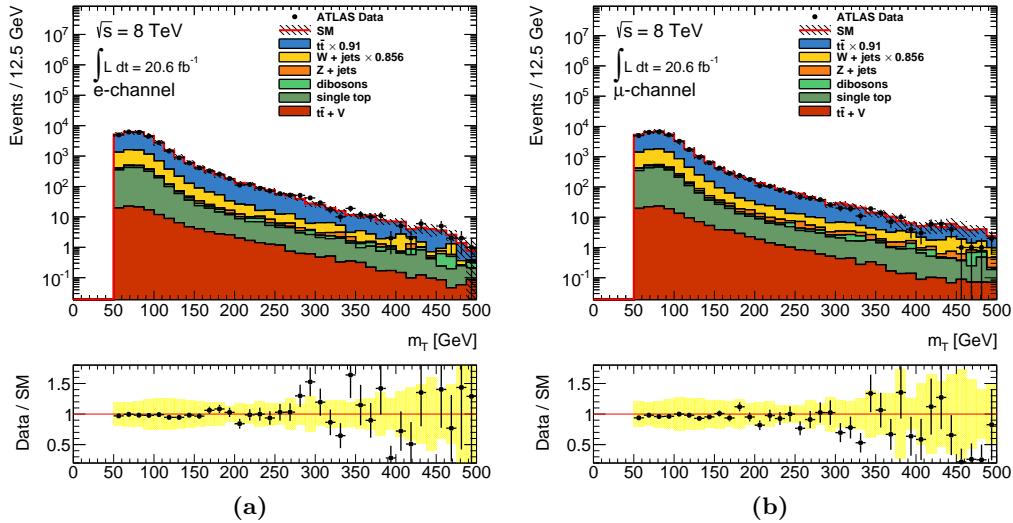
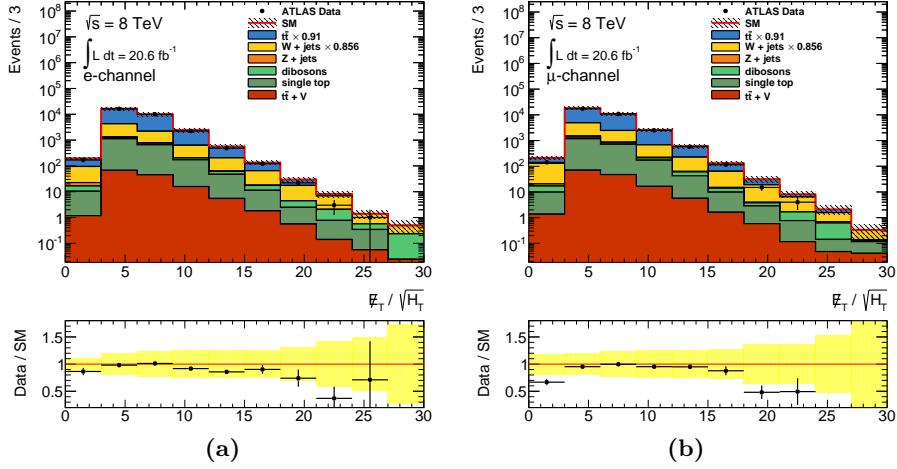
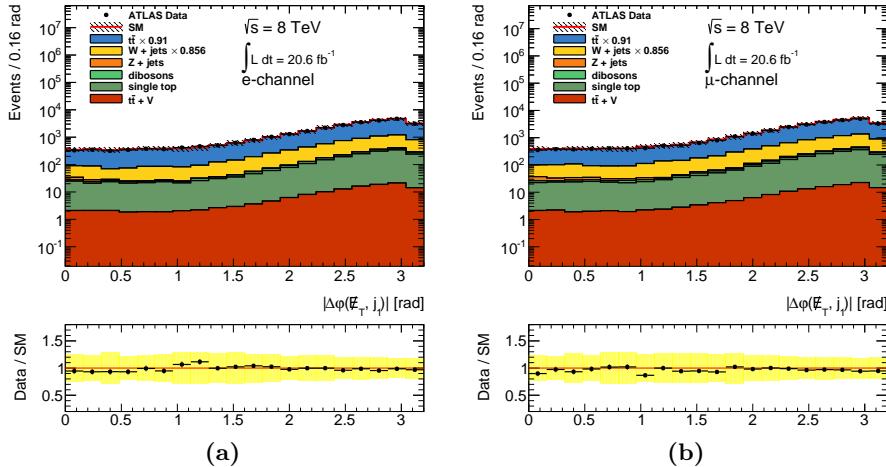
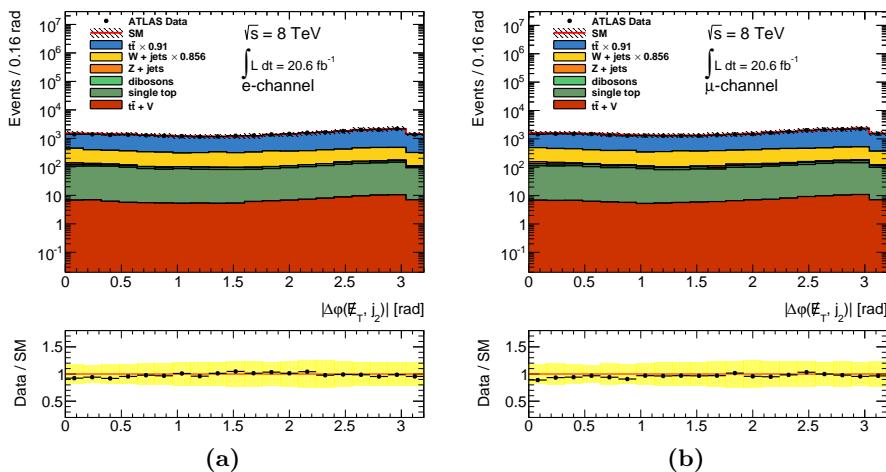


Figure A.32: The transverse mass m_T .

Figure A.33: The \not{E}_T significance.Figure A.34: Angle between \not{E}_T and the leading jet.Figure A.35: Angle between \not{E}_T and the second leading jet.

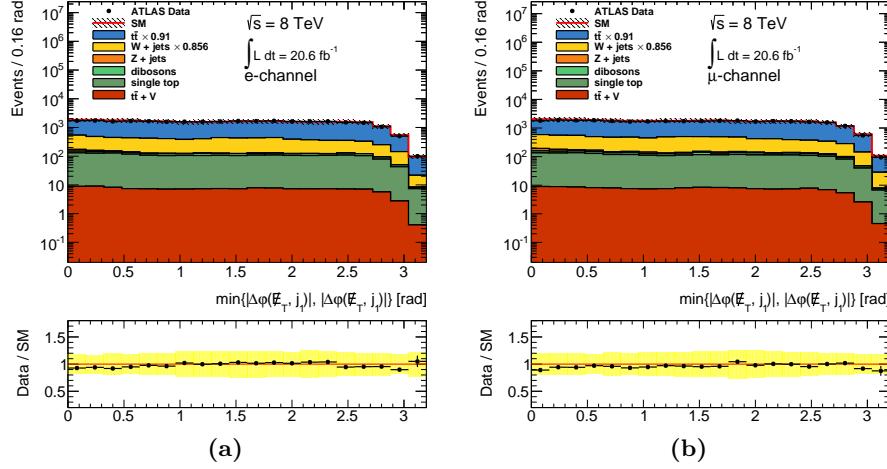


Figure A.36: The minimal angle between the \not{E}_T and the two leading jets.

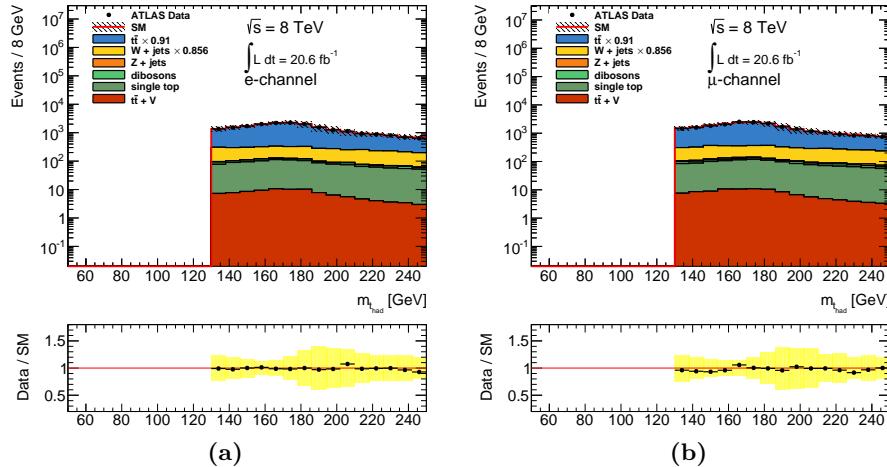


Figure A.37: The mass of the hadronically decaying top quark.

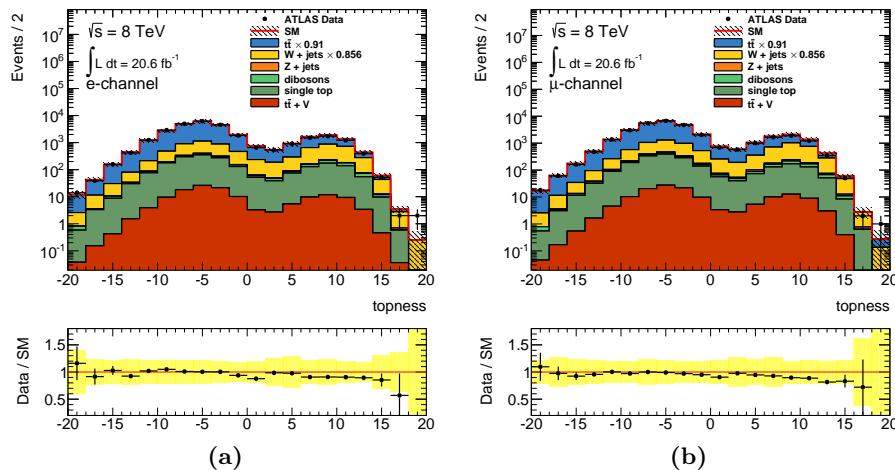
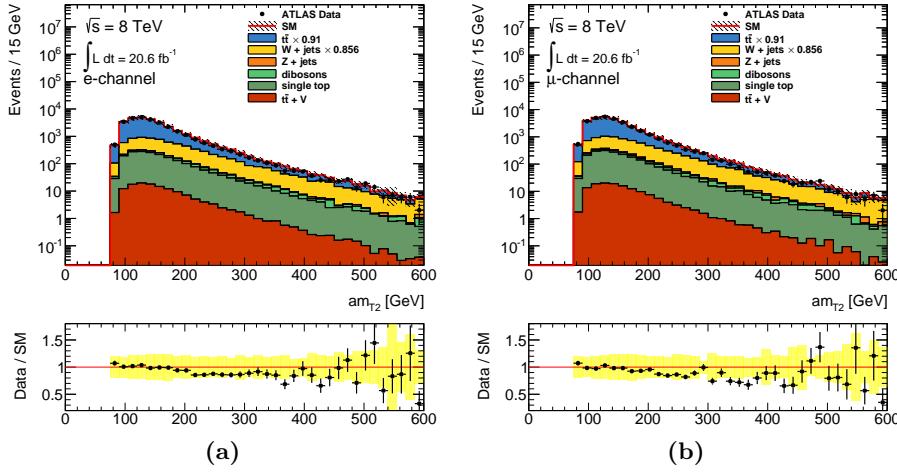
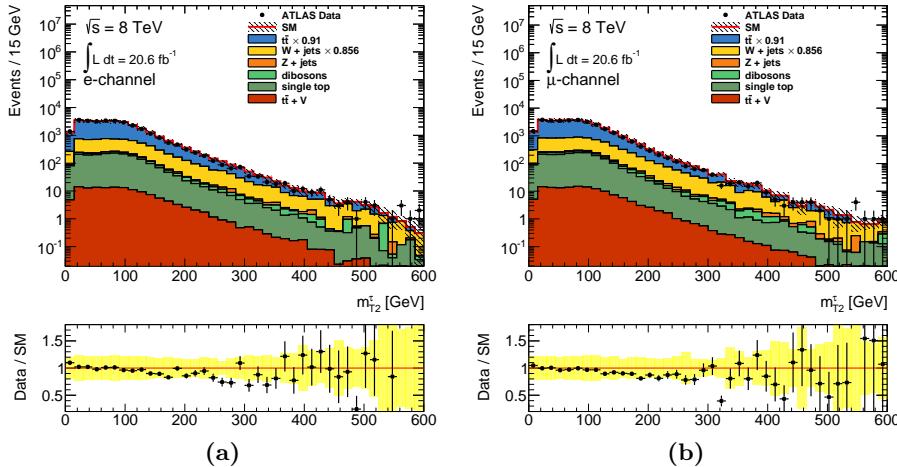
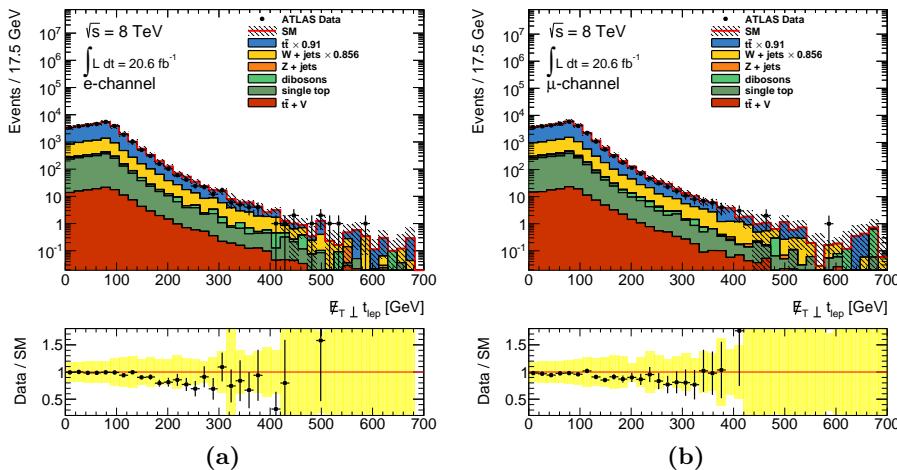
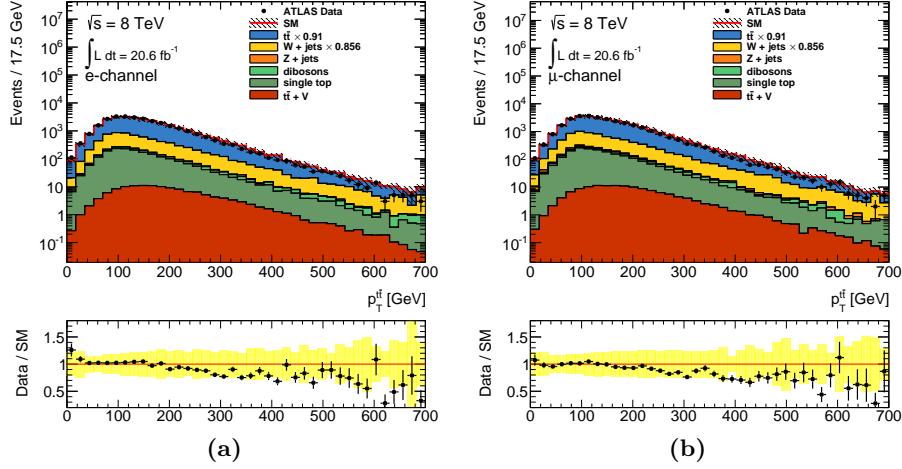
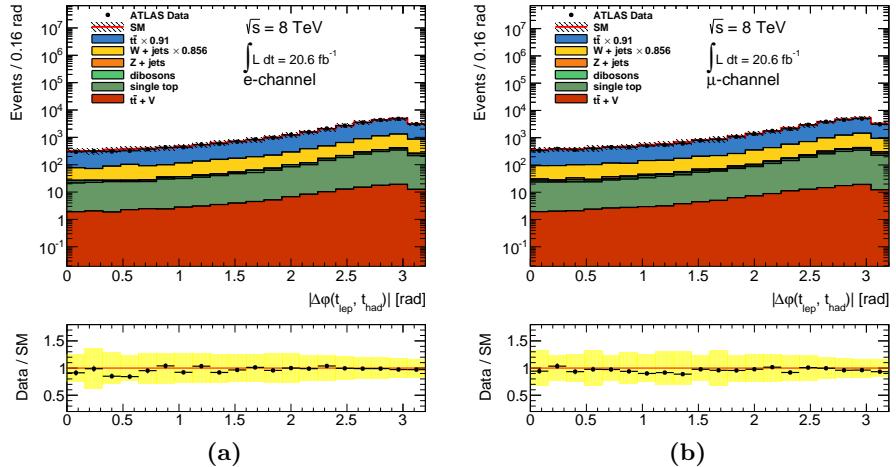
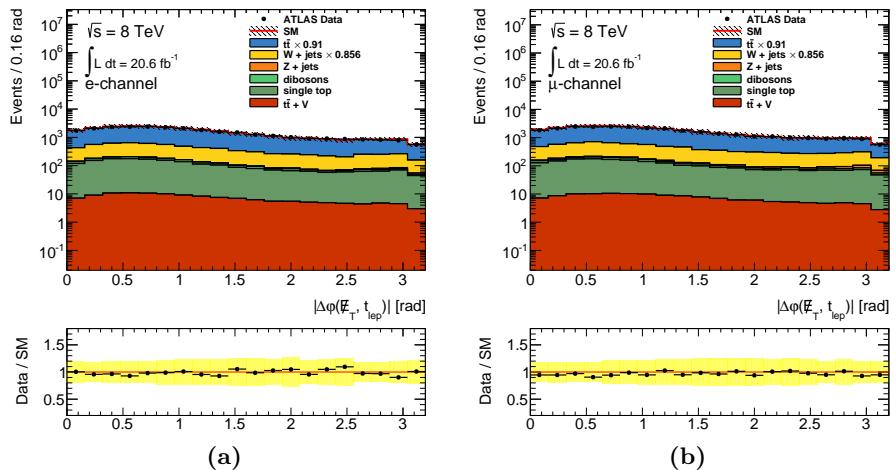


Figure A.38: The topness distribution.

**Figure A.39:** The am_{T2} distribution.**Figure A.40:** The m_{T2}^τ distribution.**Figure A.41:** The perpendicular \not{E}_T component to the leptonically decaying top quark.

**Figure A.42:** The transverse momentum of the $t\bar{t}$ pair.**Figure A.43:** The transverse angle between both top quarks.**Figure A.44:** The transverse angle between \not{E}_T and the leptonically decaying top quark.

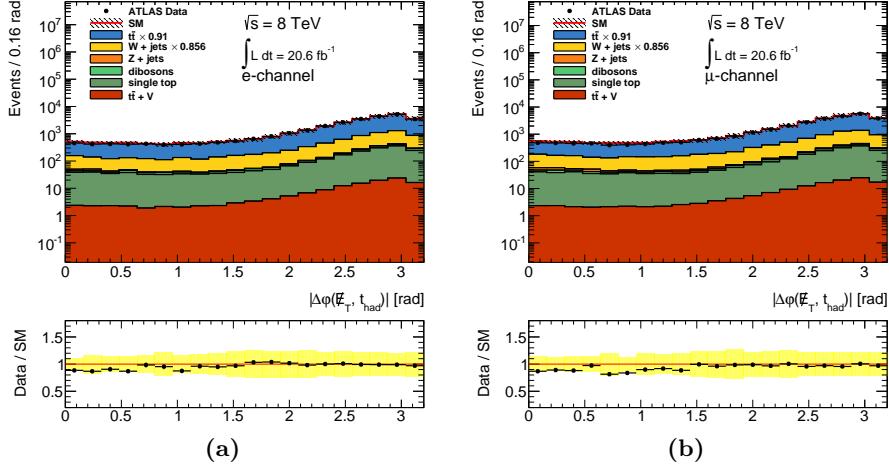


Figure A.45: The transverse angle between \vec{E}_T and the hadronically decaying top quark.

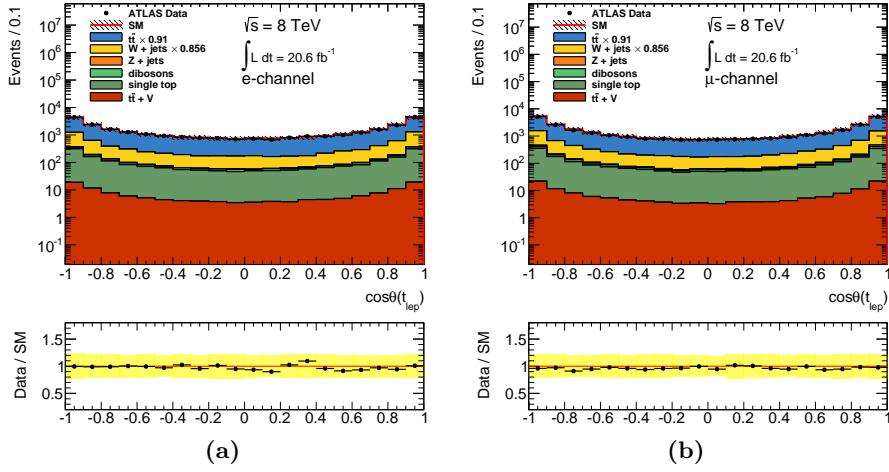


Figure A.46: The angle of the leptonically decaying top quark.

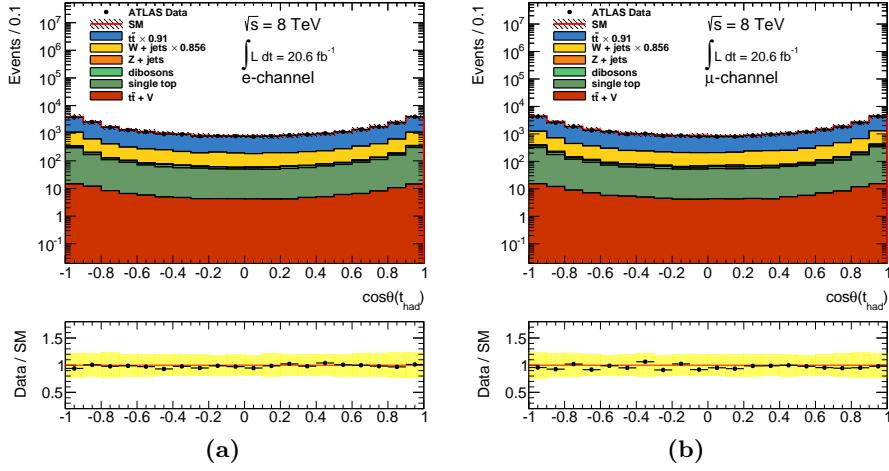


Figure A.47: The angle of the hadronically decaying top quark.

A.8. Limits on simplified models

Limits using data

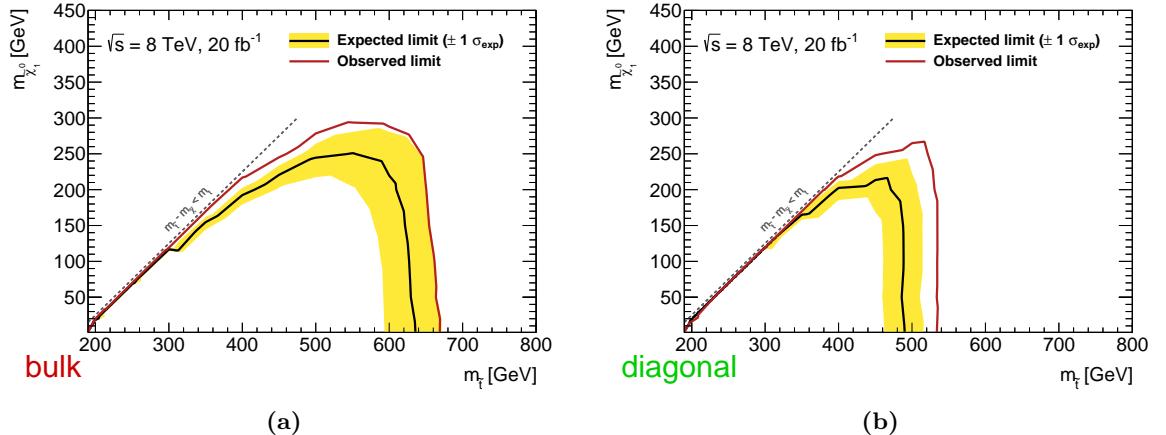


Figure A.48: Limits for the stop and the neutralino mass using the neural network output of the bulk (a) and the diagonal (b) region with systematics and measured data (cf. table 8.1).

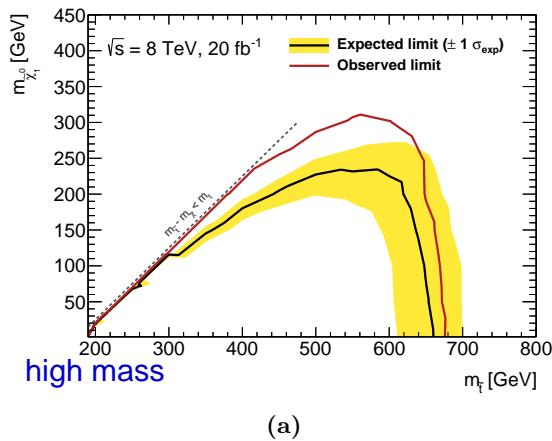


Figure A.49: Limits for the stop and the neutralino mass using the neural network output of the high mass region (a) with systematics and measured data (cf. table 8.1).

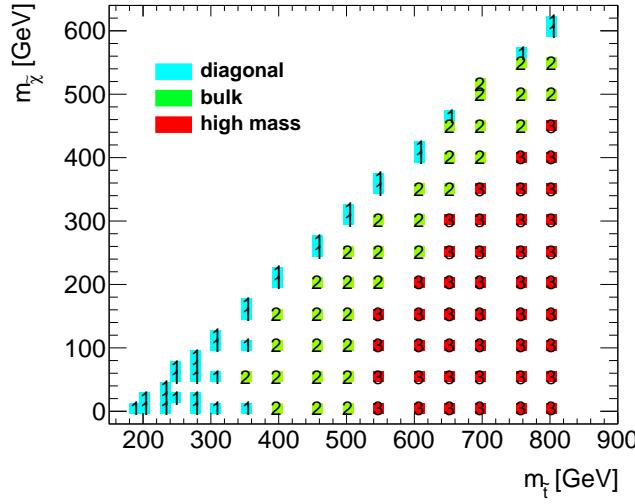


Figure A.50: The region which is used to get the best expected limit of all three analysis regions.

$t\bar{t}$ reweighting method

Table A.16: The $t\bar{t}$ reweighting method multiply the event weight with a factor depending on the transverse momentum of the truth $t\bar{t}$ pair.

$p_T^{t\bar{t}}$ [GeV]	factor
< 40	$0.5 \cdot \left(\frac{1}{1.01185} + \frac{1}{0.994193} \right)$
< 170	$0.5 \cdot \left(\frac{1}{1.09592} + \frac{1}{1.03448} \right)$
< 340	$0.5 \cdot \left(\frac{1}{1.40728} + \frac{1}{1.31911} \right)$
> 340	$0.5 \cdot \left(\frac{1}{1.79938} + \frac{1}{1.71078} \right)$

A.9. Preparation for run II and reoptimisation of the high mass region

Variables The following distributions are shown for a stop-signal with $m_{\tilde{t}} = 1500 \text{ GeV}$ and $m_{\tilde{\chi}_1^0} = 400 \text{ GeV}$ and the $t\bar{t}$ background and they are normalised to unit area.

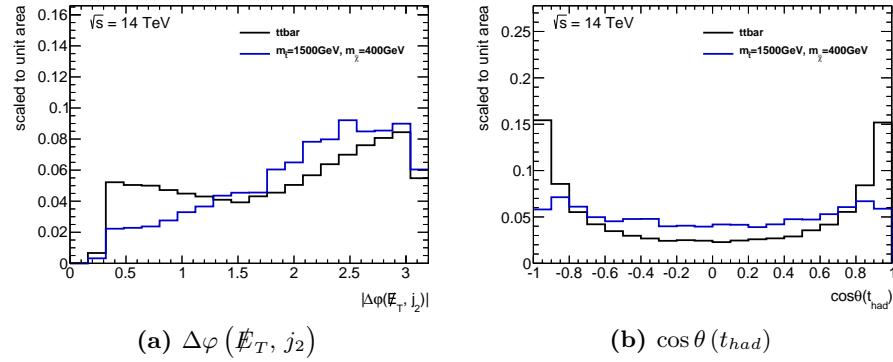


Figure A.51

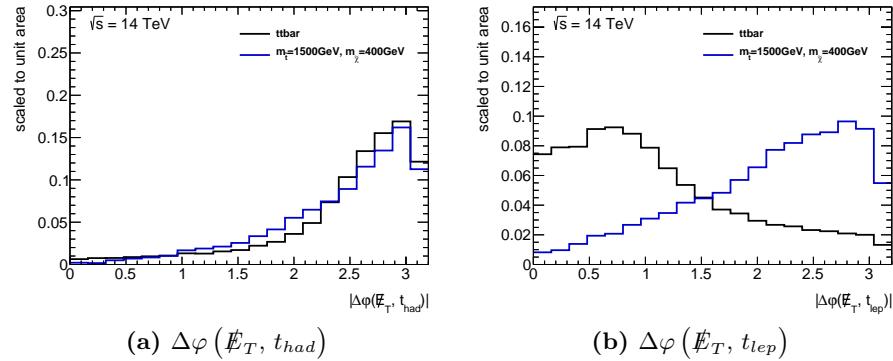


Figure A.52

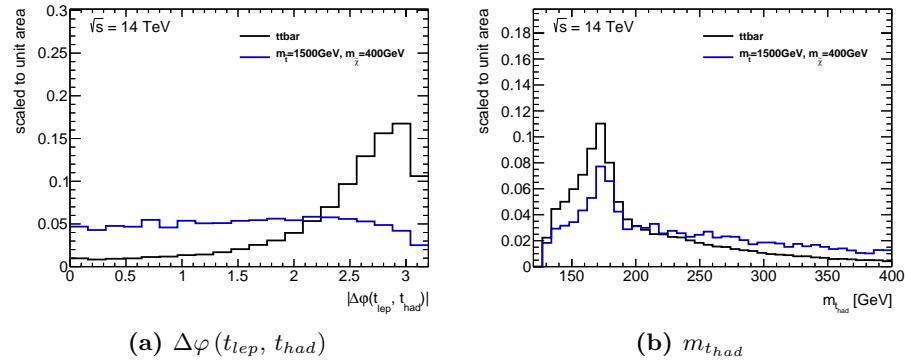


Figure A.53

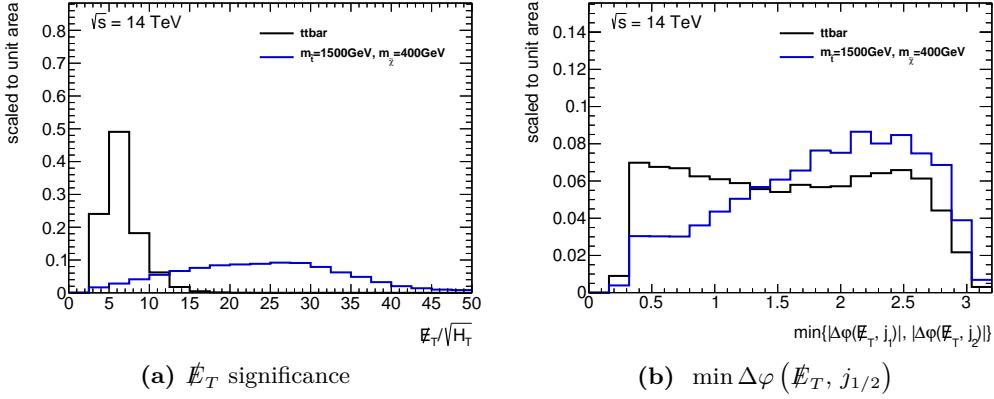


Figure A.54

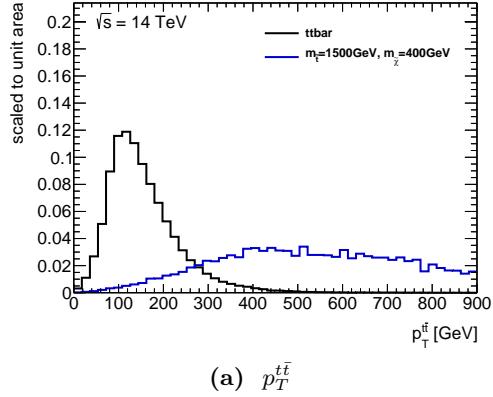


Figure A.55

Optimisation with other preselection

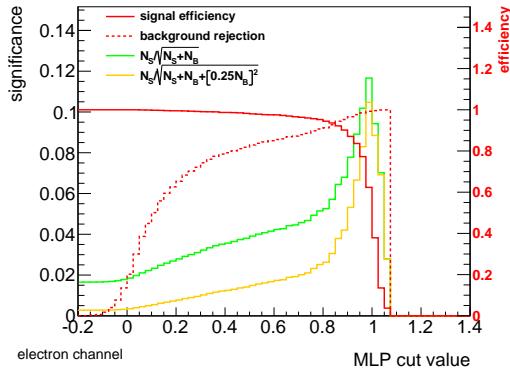


Figure A.56: Signal efficiency (red), background rejection (red dashed) and two significances (green, yellow) depending on the MLP value are shown for the centred stop-sample ($\tilde{t}(1500, 400)$).

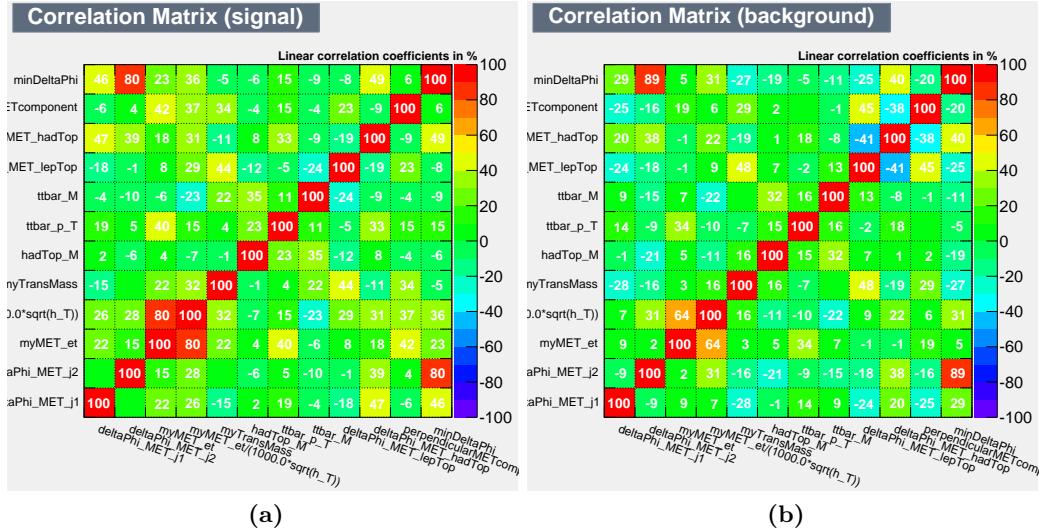


Figure A.57: Linear correlation coefficients for the remaining variables after the optimisation. The linear correlation coefficients are shown for both signal (a) and background (b).

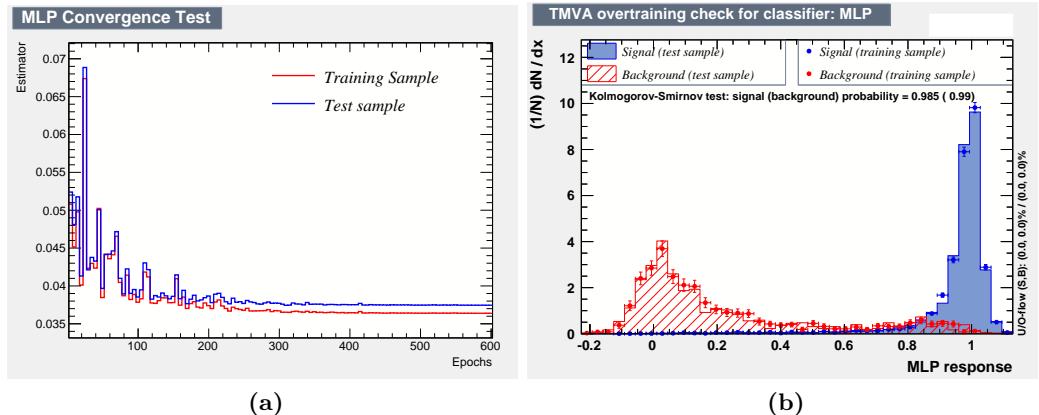


Figure A.58: Convergence and overtraining check for the MLP configuration. The estimator for the test and the training sample is shown as a function of the number of epochs (a). The MLP distribution is shown for the test and the training sample (b).

Reoptimisation for a center of mass energy of 8 TeV

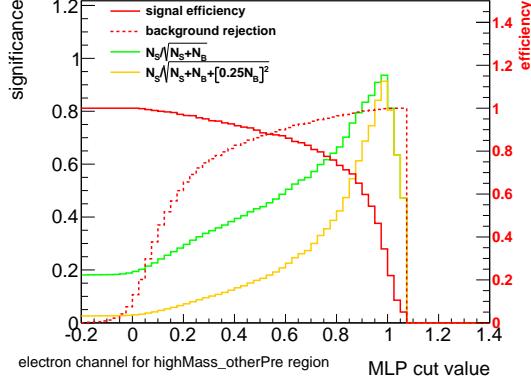


Figure A.59: Signal efficiency (red), background rejection (red dashed) and two significances (green, yellow) depending on the MLP value are shown for the centred stop-sample (\tilde{t} (750, 100)).

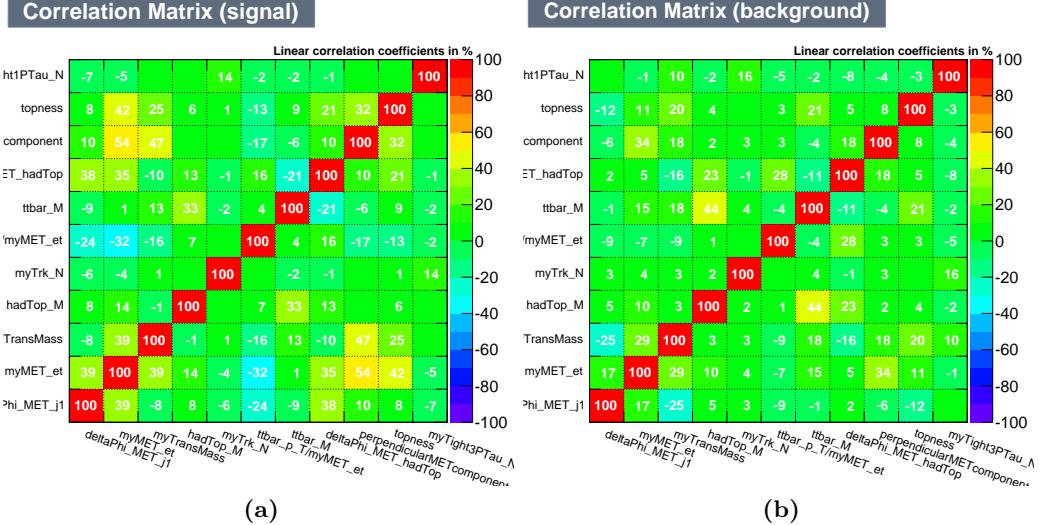


Figure A.60: Linear correlation coefficients for the remaining variables after the optimisation. The linear correlation coefficients are shown for both signal (a) and background (b).

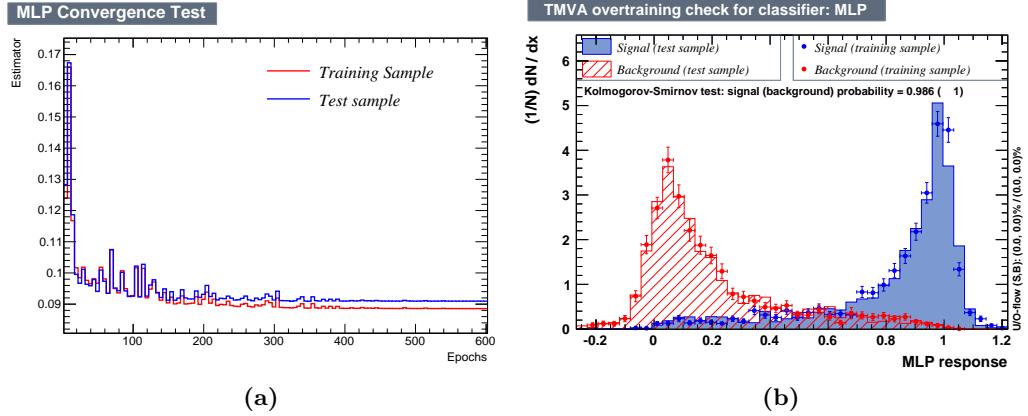


Figure A.61: Convergence and overtraining check for the MLP configuration. The estimator for the test and the training sample is shown as a function of the number of epochs (a). The MLP distribution is shown for the test and the training sample (b).

Erklärung

Hiermit erkläre ich, Daniela Börner, dass ich diese Arbeit zum Thema

**Search for the decay of a top quark superpartner
into a top quark and a neutralino using multivariate
techniques at the ATLAS experiment**

selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt, sowie Zitate
kenntlich gemacht habe.

Wuppertal, den 26. Februar 2015

(Daniela Börner)