Search for top squark pair production in the 3-body decay mode with a single lepton final state with the ATLAS detector

David Handl

SIGILL·UNIVERSITAT·LUDOVIC·MAXIM

München 2019

Search for top squark pair production in the 3-body decay mode with a single lepton final state with the ATLAS detector

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
David Handl
aus St.Margarethen

München, den 11. Juni 2019

Erstgutachter: Prof. Dr. Dorothee Schaile
Zweitgutachter: Prof. Dr. Martin Faessler
Tag der mündlichen Prüfung: 26. Juli 2019

"With a strong heart and a good mind, you can do it."

– Eliud Kipchoge

Dedicated to Juliane

Zusammenfassung

In dieser Arbeit werden die Ergebnisse zweier Suchen nach direkter Paarproduktion von Top-Squarks, den supersymmetrischen Partnern des Top-Quarks dokumentiert.

Beide Suchen konzentrieren sich auf Endzustände mit einem isolierten Elektron oder Myon, mehreren hadronischen Jets und großem fehlenden transversalen Impuls. Die erste Analyse wird an Daten von Proton-Proton-Kollisionen durchgeführt, die vom Large Hadron Collider mit einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV erzeugt und vom ATLAS-Detektor in den Jahren 2015 und 2016 aufgezeichnet wurden. Die Datenmenge entspricht einer integrierten Luminosität von 36.1 fb^{-1}. Die zweite Analyse wird am vollständigen Run 2 Datensatz von Proton-Proton-Kollisionen mit einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV durchgeführt und wurde vom ATLAS-Detektor im Zeitraum von 2015 bis 2018 aufgezeichnet. Die Datenmenge entspricht einer integrierten Luminosität von 139 fb^{-1}. Ein besonderer Top-Squark Zerfall wird berücksichtigt, bei dem die Massendifferenz zwischen dem Top-Squark und dem Neutralino kleiner ist als die Masse des Top-Quarks, sodass jedes Top-Squark daher über einen 3-Körper-Zerfall zu einem Bottom-Quark, einem W Boson und einem Neutralino zerfällt. In diesem Phasenraum ähneln die Ereignisse des Top Squark-Paares sehr den Prozessen der Top-Quark-Paar Produktion.

In beiden Suchen wird keine signifikante Abweichung von den erwarteten Standardmodell Ereignissen beobachtet, daher werden Ausschlussgrenzen bei 95 % Vertrauensniveau im Bezug auf das supersymmetrischen Modell bestimmt. In der ersten Analyse werden Top-Squarks mit Massen von bis zu 460 GeV ausgeschlossen. Mit den Ergebnissen der zweiten Analyse konnten die Ausschlussgrenzen erweitert werden und Top-Squarks mit Massen von bis zu 720 GeV und Neutralino Massen von bis zu 580 GeV können ausgeschlossen werden.

Abstract

In this work, the results from two searches for direct pair production of top squarks, the supersymmetric partner of the top quark, are reported.

Both searches focus on final states with one isolated electron or muon, multiple hadronic jets, and large missing transverse momentum. The first analysis is performed on data from proton-proton collisions delivered by the Large Hadron Collider at a centre-of-mass energy of $\sqrt{s} = 13$ TeV and recorded by the ATLAS detector within the years 2015 and 2016, corresponding to an integrated luminosity of 36.1 fb^{-1}. The second analysis is performed on the full Run 2 dataset of proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV recorded by the ATLAS detector within the period from 2015 to 2018, corresponding to an integrated luminosity of 139 fb^{-1}. A particular top squark decay mode is considered, where the mass difference between the top squark and the neutralino is smaller than the top quark and as a result each top squark decays via a 3-body process into a b quark, a W boson, and a neutralino. In this phase space, the top squark pair events closely resemble top quark pair processes.

No significant deviation from the predicted Standard Model background is observed in both searches. Hence, exclusion limits at 95 % confidence level on the supersymmetric model are determined. In the first analysis, top squarks with masses up to 460 GeV are excluded. With the results from the second analysis, the exclusion limit is extended and top squarks with masses up to 720 GeV and neutralino masses up to 580 GeV are excluded.

Acknowledgements

Without a doubt it is simply conceivable that without the substantial effort and encouragement of many people, the completion of this thesis would not have been possible. I would like to take the opportunity and express my deepest gratitude to some people now.

First of all, I would like to thank my thesis advisor Prof. Dorothee Schaile for her support, expert advice and for giving me the chance to conduct my research at the Institute of Experimental Particle Physics at the Ludwig-Maximilians University Munich as well as for the great opportunity to spend 6 month of my research abroad at CERN.

Thank you to Prof. Martin Faessler for writing the second expert opinion on my thesis and all members of the examining board for the willingness to participate to my defence.

I would also like to express my sincere gratitude to PD Jeannine Wagner-Kuhr for being an excellent supervisor. Her proficient guidance and encouragement throughout this project was invaluable for me. I also want to thank in particular PD Alexander Mann for taking over supervision from Jeannine during the last period of my research.

I am deeply grateful to the conveners of the Supersymmetry Working Group of the ATLAS collaboration, Zach and Federico, and special thanks goes out to Calum and Rosa, the conveners of the subgroup for searches of third generation squarks. I have greatly benefited from fruitful discussions and proficient advice about the analyses I contributed to. Especially in the last weeks before the analyses have been completed, I felt unlimited support and substantial encouragement. Furthermore, I am deeply grateful to the current and previous members of the $t\bar{t} + E_{\mathrm{T}}^{\mathrm{miss}}$ 1-lepton analysis who have accompanied me during the last three years, especially to the coordinators Keisuke and Michaela for their generous support. I really enjoyed being part of such a great team.

Thank you to all members of the Institute of Experimental Particle Physics at the Ludwig-Maximilians University Munich, especially to my office mates Clara, Michael and Paola for the great time we had during the last three years. I worked with great colleagues and I enjoyed working in such a supportive environment. We shared fruitful discussions about particle physics and sometimes conversations absolutely far from physics.

I want to express out appreciation to my family which is always there for me, for their unconditional support and that they always gave me the chance to pursue my personal goals.

Finally, my deepest gratitude goes to Juliane for her love and her support. She is always there for me. Life would not be the same without you.

– Thank you

Contents

Zusammenfassung **xi**

Abstract **xiii**

Acknowledgements **xv**

1 Introduction **1**

2 Theoretical foundations **3**
- 2.1 The standard model of particle physics . . . 3
  - 2.1.1 The standard model as a gauge theory . . . 5
  - 2.1.2 The Higgs mechanism and the associated Higgs boson . . . 9
- 2.2 Motivation for new physics beyond the standard model . . . 10
- 2.3 Supersymmetry . . . 13
  - 2.3.1 The minimal supersymmetric standard model . . . 13
  - 2.3.2 Phenomenology of top squarks . . . 16
  - 2.3.3 Production and decay of supersymmetric particles at the Large Hadron Collider . . . 16
  - 2.3.4 Simplified models . . . 18
- 2.4 Current status of searches for top squark pair production . . . 20

3 Experimental setup **23**
- 3.1 The Large Hadron Collider . . . 23
- 3.2 The ATLAS detector . . . 26
  - 3.2.1 Magnet system . . . 28
  - 3.2.2 Inner detector . . . 29
  - 3.2.3 Calorimetry . . . 30
  - 3.2.4 Muon spectrometer . . . 32
  - 3.2.5 Trigger and data acquisition . . . 33

4 Common aspects of the searches for top squark pair production **37**
- 4.1 Analysis strategy . . . 37
- 4.2 Data collection and simulated event samples . . . 38
  - 4.2.1 Data sample . . . 38
  - 4.2.2 Simulated event samples . . . 38
- 4.3 Object definition, event reconstruction and discriminating variables . . . 44
  - 4.3.1 Track and primary vertex reconstruction . . . 44
  - 4.3.2 Electrons . . . 45
  - 4.3.3 Muons . . . 46

4.3.4 Jets . . . 48
4.3.5 Flavour tagging . . . 49
4.3.6 Hadronically decaying τ leptons . . . 50
4.3.7 Missing transverse momentum . . . 51
4.3.8 Overlap removal . . . 51
4.3.9 Discriminating variables . . . 53

5 Fundamentals of machine learning 57
5.1 Supervised learning . . . 57
5.1.1 Classification . . . 57
5.2 Neural networks . . . 58
5.2.1 Feed-forward neural networks . . . 58
5.2.2 Recurrent neural networks . . . 60
5.3 Network training . . . 62
5.3.1 Loss function . . . 62
5.3.2 Parameter optimisation . . . 63
5.3.3 Stochastic gradient descent . . . 64
5.3.4 Error backpropagation . . . 65
5.4 Performance measure . . . 66

6 Statistical data analysis 69
6.1 Likelihood . . . 69
6.2 Test statistics and p-values . . . 70
6.3 The CL_s technique . . . 71
6.4 Systematic uncertainties . . . 71
6.4.1 Experimental uncertainties . . . 71
6.4.2 Theory uncertainties . . . 72

7 Search for top squark pair production using 36.1 fb^{-1} of *pp* collision data 75
7.0.1 Event preselection . . . 75
7.1 Signal region . . . 77
7.2 Background estimation and validation . . . 81
7.2.1 Control region . . . 81
7.2.2 Validation region . . . 83
7.3 Impact of systematic uncertainties . . . 85
7.4 Results and interpretation . . . 87

8 Search for top squark pair production using 139 fb^{-1} of *pp* collision data 93
8.1 Discrimination between signal and background . . . 93
8.1.1 Event preselection . . . 94
8.2 Signal extraction with machine learning techniques . . . 96
8.2.1 Truth-assisted ML training . . . 96
8.3 Signal region definition . . . 109
8.4 Background estimation and validation . . . 113
8.4.1 Control region . . . 113
8.4.2 Validation region . . . 117
8.5 Impact of systematic uncertainties . . . 120
8.6 Results and interpretation . . . 122

9 Comparison of current results from searches for top squark pair production 129

10 Conclusion 133

Bibliography 135

List of Figures 149

List of Tables 158

List of Abbreviations 163

A Supplement to `bWN-fullSet` 165

A.1 Additional kinematic distributions after the preselection 165
A.2 Comparison of particle-level and detector-level events 168
A.3 Linear correlation of input variables 169
A.4 Profile plots 180
A.5 Further details on the alternative analysis techniques 182
A.6 Kinematic distributions of input variables in the control region 184
A.7 Kinematic distributions of input variables in the validation region 187
A.8 Kinematic distributions of input variables in the signal region 190

Chapter 1

Introduction

Particle physics is one of the central pillars in our understanding of the universe. With the discovery of the Higgs boson by the ATLAS and CMS collaborations at the Large Hadron Collider in 2012 a new age in modern particle physics has begun. The standard model of particle physics describes the building blocks of matter and its interactions at the most fundamental level. Experiments have verified its predictions to incredible precision and all particles predicted by the standard model have been found. However, despite the remarkable success of the standard model, it fails to describe some physics phenomena and open questions remain. This opens up possibilities for theories that may describe the nature of physics beyond the standard model.

Supersymmetry provides elegant solutions to some unanswered problems. The fundamental principle of the theory is a symmetry between fermions and bosons, which also allows a cancellation of the quantum loop corrections to the Higgs boson mass. Solutions to the unification of the electroweak and the strong force at very high energy scales are also provided by supersymmetric theories. An additional neutral stable particle is also predicted in some supersymmetric models which might be an attractive candidate for dark matter.

Searches for new physics beyond the standard model, such as supersymmetry, have high priority at the ATLAS and CMS collaborations. Searches performed on data from proton-proton collisions at the centre-of-mass energies of 7 TeV and 8 TeV did not hint at any evidence for new physics phenomena. The recorded dataset from proton-proton collisions at a centre-of-mass energy of 13 TeV is currently intensively investigated and many exciting high precision measurements and results from various searches beyond the standard model are yet to come in the near future.

Many supersymmetric theories favour a relatively light supersymmetric partner of the top quark, referred to as the top squark, with a predicted mass that might be within the discovery reach of the Large Hadron Collider. In this work the results from two searches for direct top squark pair production are reported. Both searches focus on events with one isolated electron or muon, multiple hadronic jets, and large transverse momentum imbalance in the final state. A particular top squark decay mode is considered, where the mass difference between the top squark and the neutralino is smaller than the top quark. As a result, each top squark decays via a 3-body process into a b quark, a W boson, and a neutralino. In this phase space, the top squark pair events closely resemble top quark pair processes. As a consequence discrimination of supersymmetric signal events from the dominant standard model background is difficult. Advanced analysis techniques are used in this work in order to extend the search sensitivity.

The first analysis is performed on data from proton-proton collisions delivered by the Large Hadron Collider at a centre-of-mass energy of $\sqrt{s} = 13$ TeV and recorded by the ATLAS detector, corresponding to an integrated luminosity of 36.1 fb^{-1}. In this search, the sensitivity of the signal model is improved by utilising the shape differences of the signal model and the dominant standard model background in the distribution of the most discriminating kinematic variable.

The second analysis is performed on the full Run 2 dataset of proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV recorded by the ATLAS detector within the period from 2015 to 2018, corresponding to an integrated luminosity of 139 fb^{-1}. The analysis strategy in this search relies on a machine learning technique, which aims at a discrimination of signal and background events based on the kinematic properties of a given set of simulated signal and background events.

This thesis is structured as follows: Chapter 2 introduces the fundamental theories of the standard model of particle physics and supersymmetry. The Large Hadron Collider and the ATLAS detector, the experimental apparatus at which this work has been carried out, is described in Chapter 3. Common analysis aspects such as the collected dataset, simulated event samples as well as the reconstruction methods of the physics objects are discussed in Chapter 4, followed by an introduction to machine learning in Chapter 5. The statistical methods used to interpret the observed results are briefly described in Chapter 6. A search for direct top squark pair production using a sub-set of the full Run 2 data is discussed in Chapter 7. Chapter 8 continues with the search for direct top squark pair production using the full dataset of Run 2. The observed results from the two analyses are then compared with the results from other experiments in Chapter 9. Finally, a conclusive remark of this thesis is expressed in Chapter 10.

Chapter 2

Theoretical foundations

In this chapter the standard model of particle physics is first described and an overview of the fundamental particles and their interactions is given. Further on, a description of the standard model as a gauge theory is introduced and the principle of spontaneous symmetry breaking is shown. Later on, the Higgs mechanism and how the particles finally acquire masses are described. After a discussion on deficits of the standard model and a motivation for new physics that may appear beyond the standard model, the fundamentals of supersymmetry are introduced. Furthermore, phenomenological aspects such as assumptions of masses for supersymmetric particles as well as possible production and decays in hadron colliders are discussed and the framework of simplified models is summarised. The last section in this chapter is dedicated to the current search programme for a particular supersymmetric decay scenario at the ATLAS collaboration.

2.1 The standard model of particle physics

The standard model of particle physics (SM) [1–3] incorporates all fundamental constituents of ordinary matter, known as elementary particles. Likewise, interactions between the particles are described in the SM by an exchange of particles. Generally particles can be categorised into fermions and bosons, identified by either a half-integer spin or integer spin respectively. At microscopic scales, matter consists of fermions with spin $\frac{1}{2}$, whereas interactions are mediated by bosons of spin 1. In addition, the SM also includes one boson of spin 0, which takes on an important role in the theory, namely the Higgs boson.

Three of the four known fundamental forces of nature are of particular relevance to particle physics, each interaction is thereby described by a quantum field theory. Quantum electrodynamics (QED) is the corresponding theory that describes electromagnetism by the exchange of photons between charged particles. The weak nuclear force, which is for example responsible for the β-decay of nuclei, is mediated by the charged W^+ and W^- bosons. In addition, the weak interaction contains also a neutral force mediator, referred to as Z boson, which is, among other terms, the result of the unification of the weak force with QED to the electroweak theory, developed by S.L. Glashow, S. Weinberg and A. Salam [1, 2, 4]. Unlike the photon the $W^\pm$ and Z bosons are massive particles. The strong nuclear force is mediated by gluons which, like the photon, are massless, and its respective quantum theory is referred to as quantum chromodynamics (QCD). The fundamental interactions, included in the standard model, with their associated mediator are indicated in Table 2.1. In the case of gravity, a complete quantum field theory has yet to be worked out, likewise it is also assumed that gravity is too weak to play a significant role in particle physics, at least at the weak energy scale.

The building blocks of matter, the fermions, can be divided into leptons and quarks. Both can be further grouped in three generations containing two particles respectively. This results in a total content of twelve fundamental spin $\frac{1}{2}$ particles as listed with their respective electrical charge and mass in Table 2.2. Indeed, this is just half of the fermion content because additionally there exists an antiparticle state for each fermion with the

Interaction	Mediator		Spin	Mass [GeV/c^2]
Strong	Gluon	g	1	0
Electromagnetic	Photon	γ	1	0
Weak	W/Z boson	W$^\pm$ / Z	1	80.379 ± 0.012 / 91.1876 ± 0.0021

Table 2.1: The three fundamental forces contained in the SM. Their associated mediators, together with their spin and mass are shown [5].

same mass but opposite electrical charge. This is the consequence of the Dirac equation of relativistic quantum mechanics, which describes the dynamics of each of the fermions.

Leptons (Spin 1/2)				Quarks (Spin 1/2)			
Particle		Charge	Mass [GeV/c^2]	Particle		Charge	Mass [GeV/c^2]
electron	e$^-$	-1	0.0005	down	d	$-1/3$	0.005
neutrino	ν_e	0	$< 10^{-9}$	up	u	$+2/3$	0.003
muon	μ^-	-1	0.106	strange	s	$-1/3$	0.1
neutrino	ν_μ	0	$< 10^{-9}$	charm	c	$+2/3$	1.3
tau	τ^-	-1	1.78	bottom	b	$-1/3$	4.5
neutrino	ν_τ	0	$< 10^{-9}$	top	t	$+2/3$	174

Table 2.2: Basic properties of the twelve fundamental fermions separated into leptons and quarks. Masses are rounded, precises values can be found in [5].

The electron (e^-), the muon (μ^-) and the tau-lepton (τ^-) together with their associated neutrinos (ν_e, ν_μ, ν_τ) make up the three generations of leptons. Electron, muon and tau-lepton have very similar characteristics but differ in mass. Since they carry electrical charge of -1 they participate in the electromagnetic interaction as well as in the weak interactions. On the other hand neutrinos only interact weakly with matter, they are electrically neutral and from observations of neutrino oscillations it is known that neutrinos must have a non-zero mass [6, 7]. However, their exact masses have yet to be determined and only upper limits are reported [5].

The second group of fundamental fermions are the quarks. Each quark generation contains a doublet of particles with electrical charges of $+2/3$ and $-1/3$ and the respective names are up and down, charm and strange, top and bottom. As can be seen in Table 2.2, their masses span over a wide range. Apart from the top quark, which has a very short lifetime, isolated quarks cannot be observed. Moreover, they can only be observed as bound states which are collectively referred to as hadrons. The mechanism that prevents the observation of isolated quarks or gluons can be described by the hypothesis of colour confinement, although there is no analytic proof of this concept until now. The two main types of hadrons are composed of either a quark anti-quark pair, called mesons, or three quarks or three anti-quarks referred to as baryons. However, there exist also other, "exotic" bound states possible. For example bound states composed of two quark anti-quark pairs from the first two quark generations or the composition of a quark anti-quark pair together with three quarks or anti-quarks which has been recently announced by the LHCb collaboration [8]. In general, the Pauli exclusion principle of fermions would be violated in bound states composed of two or more identical quarks, hence the introduction of an additional quantum number, denoted as colour-charge, as an allusion to the electrical charge, is necessary. Quarks, which have a colour-charge and an electromagnetic charge, interact with each other through all three forces incorporated in the SM, the strong nuclear force, the electromagnetic and the weak interaction.

2.1.1 The standard model as a gauge theory

The SM can be described as a non-Abelian Yang-Mills gauge theory based on the symmetry group:

$$SU(3)_\mathrm{C} \otimes SU(2)_\mathrm{L} \otimes U(1)_\mathrm{Y} \tag{2.1}$$

Where $SU(3)_\mathrm{C}$ represents the three colour degrees of freedom and $SU(2)_\mathrm{L} \otimes U(1)_\mathrm{Y}$ is the underlying symmetry of the electroweak theory. Y in this nomenclature represents the weak hypercharge and $SU(2)_\mathrm{L}$ is the symmetry group of the weak interaction which only couples via the weak isospin to left-handed (L) particles and right-handed (R) antiparticles.

Gauge theories build the mathematical foundation to describe interactions in the SM. A comprehensive derivation of gauge theories will be omitted in this document, however the basic concept of the Lagrangian formalism will be outlined because it is essential for the understanding of the relationship between particles and forces. Detailed descriptions may be found in [9–13].

Elementary particles are considered as excitations of relativistic quantum fields. The dynamics of quantum fields are expressed by the Lagrangian density $\mathcal{L}$ and the equation of motion, known as the Euler-Lagrange equation, is derived by requiring an infinitesimal variation of the action S to vanish:

$$S = \int d^4x\mathcal{L} \rightarrow \delta S = 0, \tag{2.2}$$

with the 4-dimensional space-time x. In natural units ($\hbar = c = 1$), the action S is dimensionless and there is only one independent unit left, which is [mass] (or energy). It follows that the Lagrangian density has to be of dimension $[\text{mass}]^4$. Exemplary, the Dirac Lagrangian of a free fermion described by the four-component complex spinor field $\psi(x)$ and mass m looks like:

$$\mathcal{L} = i\bar{\psi}(x)\gamma^\mu\partial_\mu\psi(x) - m\bar{\psi}(x)\psi(x) \tag{2.3}$$

Here, γ^μ represents the four Dirac-matrices, ∂_μ is the four-derivative and $\bar{\psi}$ is the adjoint spinor ($\bar{\psi} \equiv \psi \dagger \gamma^0$).

Gauge principle

A gauge theory is commonly defined by a symmetry group G and the principle of local gauge symmetry requires that the physics does not change under the symmetry of the group. In other words the physics is invariant under local phase transformations $\hat{U}$ and so is the fermion field ψ:

$$\psi(x) \rightarrow \psi'(x) = \hat{U}(x)\psi(x) \tag{2.4}$$

The physics of electromagnetism (QED), for example, is invariant under local $U(1)$ phase transformations and thus may be represented by the symmetry group $U(1)$. The local phase transformation (Equation 2.4) can be written as:

$$\psi(x) \rightarrow \psi'(x) = e^{iq\chi(x)}\psi(x) \tag{2.5}$$

q represents the coupling, which is equal to the electrical charge ($q = -e$) in electromagnetism. $\chi(x)$ is a phase factor. The local phase invariance also affects the Lagrangian, because $\mathcal{L}$ depends on the spinor field $\psi(x)$. Since the local phase factor χ is a function of x, the derivative ∂_μ acts not only on the field $\psi(x)$, but also on the phase factor itself. Thus, the Lagrangian from Equation 2.3 transforms:

$$\mathcal{L} \to \mathcal{L}' = \mathcal{L} - q\bar{\psi}\gamma^\mu(\partial_\mu\chi)\psi \tag{2.6}$$

Hence, it can be easily seen that the Dirac Lagrangian of a free fermion from Equation 2.3 is not invariant under local $U(1)$ transformations. In general, local phase invariance is not possible for a free theory without any interactions. However, the required symmetry can be preserved if $\mathcal{L}$ will be extended by a further degree of freedom A_μ:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - q\bar{\psi}\gamma^\mu A_\mu\psi \tag{2.7}$$

The second term in the equation above is necessary to compensate the offended term in Equation 2.6 and describes the interaction of a fermion with the newly introduced gauge field A_μ, which can be interpreted as the massless photon. A_μ itself has to satisfy following transformation property:

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu\chi \tag{2.8}$$

Formally, the required gauge invariance can be restored by substituting the derivative ∂_μ by the so called covariant derivative D_μ:

$$\partial_\mu \to D_\mu = \partial_\mu + iqA_\mu \tag{2.9}$$

Finally, the Lagrangian of the QED can be written as:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu\partial_\mu - m_f)\psi - e\bar{\psi}\gamma^\mu A_\mu\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.10}$$

The electrical charge e represents the coupling, m_f the mass of the fermion and the term $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ is referred to as the kinetic term, which is already invariant under $U(1)$ transformations and contains the field strength tensor of the electromagnetism:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{2.11}$$

Quantum chromodynamics $SU(3)_\mathrm{C}$

The mathematical formalism of the strong interaction between quarks and gluons is incorporated in the field theory of quantum chromodynamics (QCD) which is based on a non-Abelian $SU(3)$ gauge group. As described in the previous paragraph, local gauge invariance is satisfied by the use of the covariant derivative:

$$D_\mu = \partial_\mu + ig_S T^a G^a_\mu \tag{2.12}$$

Here, g_S is the strong coupling constant (usually defined as $\alpha_S \equiv g_S^2/4\pi$), while T^a represents the generators of the $SU(3)$ symmetry group with the index a running up to eight. Again new vector fields G^a_μ have to be introduced corresponding to eight massless gluon fields. The Lagrangian density for n quarks with masses m_q is given by:

$$\mathcal{L}_{QCD} = \sum_q^n \bar{\psi}_{q,A}(i\gamma^\mu D_\mu - m_q)\psi_{q,A} - \frac{1}{4}F^a_{\mu\nu}F^{a\ \mu\nu}, \tag{2.13}$$

where the index A represents three conserved colour-charges. The notation colour is simply a label for three orthogonal states in the $SU(3)$. Couplings to gluons are only possible for non-zero colour-charged particles, e.g. leptons are colour-neutral and hence do not interact via the strong interaction. The last term in Equation 2.13 represents the kinetic term of QCD with the non-Abelian field strength tensor $F^a_{\mu\nu}$. Compared to QED an extra term is necessary in the definition of $F^a_{\mu\nu}$ to keep the kinetic term invariant under $SU(3)$ transformations:

$$F^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f_{abc} G^b_\mu G^c_\nu \tag{2.14}$$

Here, f_{abc} are the structure constants of the $SU(3)$ symmetry. The last term in Equation 2.14 is responsible for the non-Abelian nature of QCD and describes the self-interaction of gluons. It is exactly the gluon-gluon self-interaction which induces important features of QCD, because additional bosonic loops are possible and contribute positively to the momentum scale (q^2) of the interaction. Furthermore, the coupling strength becomes dependent on the momentum scale ($\alpha_S(q^2)$) due to the renormalisation of the theory [14]. Hence, virtual loops from gluon self-interactions directly affect the strength of the coupling. As a consequence, in the high energy regime, as present in modern collider experiments, α_S becomes sufficiently small so that QCD processes are calculable using perturbation theory and a very good predictive power of the theory is revealed. In the asymptotic limit ($q^2 \to \infty$), the coupling vanishes completely and quarks can be treated as free particles. This property of QCD is known as asymptotic freedom. This is in contrast to the low energy limit where α_S increases. Quarks and gluons do not propagate freely and are confined to colourless bound states. If quarks are separated by distance, the potential energy increases until it becomes energetically more favourable to form a new quark-antiquark pair of opposite colour-charges. This process is usually referred to as hadronisation. In high energy collider experiments quarks and gluons are usually observed as jets which are collimated showers of colourless hadrons.

At energy scales where the coupling is large ($\alpha_S \sim \mathcal{O}(1)$), QCD processes are not computable in perturbation theory anymore. At these energy scales, calculations are performed using lattice QCD [15].

Electroweak unification $SU(2)_\mathrm{L} \times U(1)_\mathrm{Y}$

The electroweak (EWK) theory, developed by Glashow, Weinberg and Salam [1, 2, 4, 16], unifies the electromagnetic and weak interactions in the $SU(2) \times U(1)$ symmetry group. From the β decay it is well known that only left-handed particles (and right-handed antiparticles) interact weakly [17], whereas the electromagnetic interaction couples to both particle states. Hence, the key to a unified model is to combine those structures in an appropriate fashion. The conserved quantum number of the weak interaction is the weak isospin I_W or more precisely its third component I^3_W. In the weak isospin space, the generations of the lepton content can be written as left-handed doublets (e.g. $L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L$) with eigenstates $I^3_W = \pm\frac{1}{2}$, for the upper and lower particle states respectively. Right-handed particle states are singlets with $I^3_W = 0$ (e.g. e_R for the electron, right-handed neutrinos do not exist because in the EWK theory they are assumed to be massless). As a result right-handed particle states do not couple to the gauge bosons of the $SU(2)$ symmetry. For the quark content the chiral structure is similar, however a left-handed isospin doublet is constructed by an up-type quark and a down-type anti-quark (e.g. $\begin{pmatrix} u \\ d' \end{pmatrix}_L$). Right-handed quark states are also singlets (e.g. u_R and d_R, representing up-type and down-type quarks of all generations).

The procedure to form a gauge invariant theory is the same as shown before for QED and QCD. Local gauge invariance for both symmetries, $SU(2)$ and $U(1)$, is satisfied by introducing the covariant derivative as follows:

$$D_\mu = \partial_\mu + ig' T^a W_\mu^a + ig\frac{Y}{2}B_\mu, \tag{2.15}$$

g' and g denote the corresponding coupling constants, T^a are the three generators of the $SU(2)$ symmetry which can be also written in terms of the Pauli matrices, $T^a = \frac{1}{2}\sigma^a$, and W_μ^a (with $a = 1, 2, 3$) are three newly introduced gauge fields. The $U(1)$ symmetry group with the weak hypercharge Y is represented by the third term in Equation 2.15. B_μ is the fourth bosonic field introduced in the EWK model. Y is the conserved quantum number of the $U(1)$ symmetry and a relation to the weak isospin and the electromagnetic charge q is given by the Gell-Mann-Nishijima relation [18–20]:

$$q = I_W^3 + \frac{1}{2}Y \tag{2.16}$$

Together with the kinetic terms for the four gauge fields the Langrangian density for the EWK model results:

$$\mathcal{L}_{EWK} = \sum_f \bar{\psi}_f i\gamma^\mu D_\mu \psi_f - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}^a W^{a\ \mu\nu} \tag{2.17}$$

Here, the sum over f represents the fermion content, including left- and right-handed states, while $W_{\mu\nu}^a$ and $B^{\mu\nu}$ are the field strength tensors, which are related to the respective gauge fields:

$$\begin{aligned} B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\ W_{\mu\nu}^a &= \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g'\epsilon_{abc}W_\mu^b W_\nu^c \end{aligned} \tag{2.18}$$

Once more, the last term in the latter equation illustrates the non-Abelian structure of the $SU(2)$ gauge group which results in gauge self-interactions with ϵ_{abc} being the total antisymmetric Levi-Civita tensor.

The physical $W^\pm$ bosons, corresponding to the positively and negatively charged currents of the weak interaction can be identified as linear combinations of W_μ^1 and W_μ^2,

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \tag{2.19}$$

while the other two neutral gauge fields W_μ^3 and B_μ mix to the physical states A_μ and Z_μ, which are associated with the photon and the Z^0 boson:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \tag{2.20}$$

θ_W is the electroweak mixing angle, often referred to as Weinberg angle, and establishes a relation between the couplings g, g' and the electrical charge q:

$$q = g'\sin\theta_W = g\cos\theta_W \tag{2.21}$$

Certainly there are further terms required in the EWK theory, because a $SU(2) \times U(1)$ symmetry like the one described above would predict massless fermions and gauge bosons, which is unambiguously in contradiction with experimental observations. The concept of spontaneous symmetry breaking provides a solution to this delicate issue [21–24].

2.1.2 The Higgs mechanism and the associated Higgs boson

The masses of the associated gauge bosons ($W^{\pm}$ and Z) spontaneously break the gauge symmetry in the EWK model. The concept of spontaneous symmetry breaking can be described by introducing a complex $SU(2)$ scalar doublet:

$$\phi = \begin{pmatrix} \phi^{+} \\ \phi^{0} \end{pmatrix}. \tag{2.22}$$

The scalar field contributes to the SM Lagrangian as:

$$\mathcal{L} = (D^{\mu}\phi)^{\dagger}(D_{\mu}\phi) - V(\phi), \tag{2.23}$$

in which the † symbol is the Hermitian conjugate of a four-component particle state ($\psi^{\dagger} = (\psi^{*})^{T}$) and D_{μ} is the covariant derivative from Equation 2.15 and induces the coupling of the gauge fields with the scalar field. $V(\phi)$ represents the Higgs potential which is restricted to the form

$$V(\phi) = \mu^{2}(\phi^{*}\phi) + \lambda(\phi^{*}\phi)^{2}, \tag{2.24}$$

due to renormalisability and gauge invariance in $SU(2)\times U(1)$. The vacuum expectation value, v, of the scalar field ϕ corresponds to the minimum of the potential $V(x)$. The shape of the potential $V(x)$ is defined by the sign of μ^{2}. If $\mu^{2} < 0$, the potential has an infinite set of minima ($\phi^{0} \neq 0$) and the model is spontaneously broken. λ describes a quartic self-interaction. The ground state is only stable if the potential term has a finite minimum and $\lambda > 0$. This means that the choice of the minimum breaks the symmetry of the Lagrangian. However, the direction of the minimum is not defined and can be arbitrarily chosen. In the unitary gauge the scalar field can be written by the vacuum state v and an excitation $h(x)$ around the vacuum state:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{2.25}$$

$h(x)$ is also referred to as the physical scalar Higgs field. The product $(D_{\mu}\phi)^{\dagger}(D_{\mu}\phi)$ in Equation 2.23 causes the interaction of the gauge fields with the non-zero expectation value of the Higgs field and leads to the representations determined in Equations 2.19 and 2.20, whereas their masses can be derived as:

$$\begin{aligned} M_{W}^{2} &= \frac{1}{4}g^{2}v^{2}, \\ M_{Z}^{2} &= \frac{1}{4}(g^{2} + g'^{2})v^{2}, \\ M_{A} &= 0. \end{aligned} \tag{2.26}$$

It can be seen that gauge boson masses arise from the product of the respective gauge coupling with the non-zero vacuum expectation value of the initially introduced scalar field. From an experimental point of view, the masses of the W and Z gauge bosons as well as the couplings g and g' can be measured precisely and thus the ground state v can be determined.

So far, fermions have not been considered in this context. However, the Higgs mechanism can describe the masses of the fermions as well by introducing an interaction term, $\mathcal{L}_{int}$, for fermions with the Higgs field. Such an interaction between a Dirac field and a scalar field is commonly referred to as a Yukawa coupling. As mentioned before, fermions are expressed by left-handed chiral $SU(2)$ doublets and right-handed chiral $SU(2)$ singlets. The interaction Lagrangian for example of the electron is defined as:

$$\mathcal{L}_{int} = g_e(\bar{L}\phi e^-_R + \bar{e}^-_R\phi^\dagger L). \tag{2.27}$$

g_e is an arbitrary constant. Using Equation 2.25, Equation 2.27 can be written as:

$$\begin{aligned}\mathcal{L}_{int} &= \frac{g_e v}{\sqrt{2}}(\bar{e}^-_L e^-_R + \bar{e}^-_R e^-_L) + \frac{g_e}{\sqrt{2}}(\bar{e}^-_L e^-_R + \bar{e}^-_R e^-_L)h \rightarrow \\ \mathcal{L}_{int} &= m_e\bar{e}e + \frac{m_e}{v}\bar{e}eh.\end{aligned} \tag{2.28}$$

Similar to the massive bosonic fields, the electron acquires its mass by the relation of g_e with the non-zero vacuum expectation value ($m_e = \frac{g_e v}{\sqrt{2}}$). Due to the absence of right-handed neutrinos there is no mass term, which implies they do not interact with the Higgs boson.

The interaction term for quarks can be derived in a similar fashion, nonetheless the right-handed particle states for up-type (u_R) as well as down-type quarks (d_R) have to be kept in mind:

$$\begin{aligned}\mathcal{L}_{int} &= g_d\bar{Q}_L\phi d_R + g_u\bar{Q}_L\phi u_R + \text{hermitian conjugate} \rightarrow \\ \mathcal{L}_{int} &= m_d\bar{d}d + m_u\bar{u}u + \frac{m_d}{v}\bar{d}dh + \frac{m_u}{v}\bar{u}uh.\end{aligned} \tag{2.29}$$

g_u and g_d are again arbitrarily chosen and Q_L represents the left-handed quark doublet. Masses of the second and third generations of fermions can be derived similarly by repeating the procedure above.

Unlike the masses of the heavy $W^\pm$ and Z bosons, the fermion masses can not be predicted from the theory, instead their masses need to be explicitly measured in experiments.

The actual mass of the Higgs boson as well as the vacuum expectation value can be derived from the Higgs potential using the two free parameters, μ and λ:

$$\begin{aligned}v^2 &= -\frac{\mu^2}{2\lambda} \approx (246\text{GeV})^2, \\ M_h^2 &= 2v^2\lambda\end{aligned} \tag{2.30}$$

Recent precision measurements performed by the ATLAS and CMS collaboration result in a Higgs boson mass of $m_\text{H} = 125.09 \pm 0.24\text{GeV}$ [25].

2.2 Motivation for new physics beyond the standard model

The standard model has been completed with the discovery of the Higgs boson by the ATLAS and CMS collaboration in 2012 [26, 27]. All fundamental particles had been observed in various experiments over a period of many decades and most of the presently known phenomena of particle physics may be precisely described by the SM. Notable to mention, that the predictions from the SM are in excellent agreement with the results from high precision experiments. Despite this remarkable success, some phenomena had been observed, which can not be described by the theory, indicating that the standard model is incomplete.

In the following a few phenomena will be addressed and serve as motivation for the necessity of theories beyond the standard model (BSM) with particular emphasis on supersymmetry. Nevertheless, this collection of phenomena is by far not exhaustive and there still remain open questions.

Hierarchy problem

If quantum loops are considered in the theory, which may be represented as higher-order terms, the Higgs boson mass (see Equation 2.30) receives corrections and needs to be modified, in order to obtain finite results for all higher-orders. This fact is attributed to the requirement that the SM is renormalisable [14]. Large quantum corrections contribute to the Higgs boson mass from the coupling of the Higgs boson to all massive particles in the SM. As a brief example, the coupling of a SM fermion f with mass m_f to the Higgs field h, which may be characterised by the term $-\lambda_f \bar{f} f h$ in the Lagrangian and is illustrated in Figure 2.1(left), induces a quadratically divergent correction [28]:

$$\Delta M_h^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda^2 + ... \tag{2.31}$$

As shown in the previous section, the coupling λ_f is fixed by the fermion mass m_f and v, where Λ denotes the cut-off which represents the boundary where possible new physics may appear and the SM needs to be modified. Below this cut-off the theory remains well defined. Despite it is not strictly clear at which energy scales new physics may become valid, there must be certainly new physics at the order of magnitude where quantum gravitational effects become considerable. This energy scale is assumed to be at the Planck scale, $M_\text{P} \simeq G_\text{Newton}^{-1/2} \simeq 10^{18}$ GeV. If, in the implausible case that there are no physics until the Planck scale and the cut-off Λ is at such exceedingly large energy scales, the loop corrections are then some orders of magnitude larger than the Higgs boson mass itself. This would require an exceptional cancellation, or extreme fine-tuning, to cancel out the corrections to get down to the weak energy scale ($\sim 10^2$ GeV).

In place of adopting such unnatural tuning, supersymmetric models suggest an elegant solution to this problem [29–31]. Supposing there exists a scalar particle S with mass m_S which also couples to the Higgs field like it is shown in Figure 2.1(right). As a consequence such scalar particle would induce a further quantum correction to M_h:

$$\Delta M_h^2 = \frac{\lambda_S}{16\pi^2}\Big[\Lambda^2 - 2m_S^2 \ln(\Lambda/m_S) + ...\Big]. \tag{2.32}$$

λ_S represents again the coupling between the scalar particle and the Higgs field. From the comparison of the terms from Equation 2.31 and Equation 2.32 it appears that the fermionic loop contributes negatively to the correction, while the bosonic loop yields a positive correction and thus at least the quadratic dependence to the huge cut-off Λ would cancel out. This only holds if the couplings λ_f and λ_S are at the same energy scale. The gauge bosons in the SM model are massive and thus their coupling to the Higgs field also contributes to the quadratic correction ΔM_h^2. The procedure mentioned above also holds for the massive SM gauge bosons, but in that case the quadratic dependence on the cut-off Λ would cancel if there would be additional fermions that also couple to the Higgs field with a compatible coupling strength as the SM gauge bosons.

Furthermore, the second term in Equation 2.32 can be very large as well, thus further restrictions on the theory are required. In general, some fine-tuning is still required for heavy fermions that couple to the Higgs field and the scale of the tuning relates to the masses of such heavy fermions. Moreover, the size of the tuning is still subject of discussions.

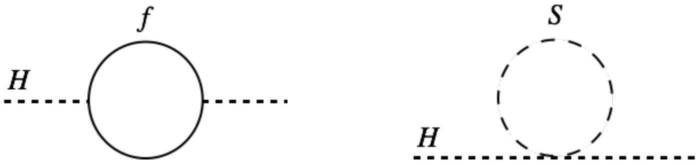


Figure 2.1: One-loop corrections to M_h^2. For a fermion f (left) and a scalar particle S (right) [28].

Dark matter

It is well known for many decades that the largest fraction of mass in the entire universe is non-luminous. This has been proven in early cosmological measurements such as rotation curves of spiral galaxies [32] as well as in modern observations like the cosmic microwave background (CMB). The non-luminous amount of matter, that do not interact with ordinary matter via the fundamental forces is referred to as dark matter.

Recent measurements from the Planck collaboration of the cosmological microwave background suggest that the matter density in the universe is about $\Omega = 0.315$ [33]. This result is in consistency with a well accepted cosmological model named the Lambda cold dark matter model (ΛCDM) [34], in which the letter Λ represents a cosmological constant and is associated with dark energy. Moreover, the matter density can be divided into ordinary baryonic matter, that can be described by the standard model of particle physics and which only takes up ~ 4.5 % of the universe, while the other ~ 27 % corresponds to dark matter. The residual amount of the energy-matter density in the universe are associated to dark energy (~ 68.5 %). The SM fails to describe dark matter completely, because there is no particle in the SM that can be considered as a cold dark matter candidate. In contrast, new weakly interacting massive particles are favoured candidates as dark matter particles.

Weakly interacting massive particles are typically predicted in BSM models, in which the lightest of the newly proposed particles is stable. Therefore, such stable neutral particles are also predicted in many supersymmetric extensions of the standard model with conserved R-parity (see Section 2.3.1). In fact, lightest neutral supersymmetric particles are often considered as favourable dark matter candidates [35, 36].

Grand unification

The following phenomenon is not necessarily a motivation for supersymmetry, it is also not essential for any theory to be valid, though in supersymmetry it is possible to unify the gauge couplings from the SM into a single coupling constant. This means that the $SU(3) \times SU(2) \times U(1)$ symmetry can be embedded into a more general gauge group. Such theories, that attempt to unify the fundamental forces, are commonly termed as grand unified theories (GUTs) [37]. Whether gauge unification is feasible in a certain theory depends, among other requirements, on the particle content of the theory. This simply favours supersymmetric theories, where at least twice as much particles are included, compared to the standard model. Provided that those additional particles are not too heavy, with masses up to TeV. However, gauge unification in supersymmetric GUT models, which contain particles with masses in the TeV range, is assumed to arise not until extremely high energy scales, $\sim 2 \times 10^{16}$ GeV, denoted as the GUT scale [38–41].

2.3 Supersymmetry

The origins of supersymmetry (SUSY) date back to the early 1970s [42, 43]. By that time, first theoretical ideas had been discussed and the mathematical framework had been developed. The first context between supersymmetry and particle physics had been proposed a few years later [44, 45]. However, it took another few years for the construction of a realistic SUSY model, leading to the first supersymmetric extension of the standard model [46–48].

As already discussed earlier in Section 2.2, a symmetry between fermions and bosons can potentially cure the fine-tuning problem of the SM, given their coupling to the Higgs boson is similar. Supersymmetry represents exactly such a symmetry between fermions and bosons. A supersymmetric transformation, represented by a SUSY generator Q, converts fermionic states into bosonic states and vice versa:

$$Q\,|\text{Fermion}\rangle = |\text{Boson}\rangle\,, \qquad Q\,|\text{Boson}\rangle = |\text{Fermion}\rangle\,. \tag{2.33}$$

Unlike the bosonic generators of the Poincaré algebra, which represent the algebra of translations, rotations and boosts in four-dimensional space-time, Q (as well as its hermitian conjugate $Q^\dagger$) is a fermionic operator. This has an important consequence. A SUSY transformation changes the spin of the respective particle state by $\frac{1}{2}$, while all residual quantum numbers remain identical. The generators of the SUSY algebra must satisfy the following anticommutation and commutation relations [28]:

$$\begin{aligned} \{Q, Q^\dagger\} &= P^\mu, \\ \{Q, Q\} &= \{Q^\dagger, Q^\dagger\} = 0, \\ [P^\mu, Q] &= [P^\mu, Q^\dagger] = 0. \end{aligned} \tag{2.34}$$

Where P^μ represents the four-momentum generator of space-time and thus, the SUSY algebra extends the ordinary space-time to super-space. According to Equation 2.33, supersymmetry introduces a supersymmetric particle state for each particle degree of freedom in the SM. Supersymmetric particles are also referred to as superpartners. All SM and SUSY particles are organised in supermultiplets, which are irreducible representations of the SUSY algebra. Furthermore, leptons and quarks, which are chiral fermions of spin $\frac{1}{2}$, can be described by chiral or matter supermultiplets and obtain scalar superpartners of spin 0, referred to as sleptons and squarks (the 's' in front of the conventional name stands for scalar). More precisely, each chiral state of the SM fermions receives a supersymmetric state due to the chiral structure of the $SU(2)$ in the SM, e.g the first generation of the leptons is the doublet (e_L, ν_e), accompanied by their supersymmetric scalars $(\tilde{e}_L, \tilde{\nu}_e)$, called selectron and sneutrino. The $\sim$ in this notation refers to supersymmetric particle states. The standard model gauge bosons with spin 1 are included in vector or gauge supermultiplets, which also contain their fermionic superpartners, denoted as gauginos. The Higgs boson, which is a spin 0 particle, obtains a fermionic superpartner called higgsino and hence is accommodated in a chiral supermultiplet.

2.3.1 The minimal supersymmetric standard model

The minimal supersymmetric standard model (MSSM) [46–48] builds up the simplest supersymmetrisation of the SM and is minimal in the sense of the particle content.

The SM fermions together with their supersymmetric scalar partners, the sfermions, are members of chiral supermultiplets. Due to the chiral structure of the SM fermions, left-handed and right-handed particle states are described by separate two-component Weyl spinors with its own supersymmetric scalar partner. For example, the left-handed and right-handed superpartners of the SM electron are the selectrons denoted as $\tilde{e}_L$ and $\tilde{e}_R$. Although, the nomenclature L and R does not refer to the their chiral state since they are spin 0 particles. The same convention is adapted to the smuons, staus and also to the squarks $\tilde{q}_L$ and $\tilde{q}_R$ (with $q = u, d, s, c, t, b$). The notation for sneutrinos is slightly simplified ($\tilde{\nu}_e$, $\tilde{\nu}_\mu$, $\tilde{\nu}_\tau$), because there are no right-handed neutrino states in the standard model.

The SM Higgs boson with spin 0 obtains supersymmetric fermionic partners with spin $\frac{1}{2}$, referred to as higgsinos which are also integrated in a chiral supermultiplet. In fact, instead of one Higgs scalar, two complex Higgs doublets, denoted as H_u and H_d, are needed, with hypercharges $Y = \pm\frac{1}{2}$, to avoid gauge anomalies. The Higgs supermultiplet with $Y = \frac{1}{2}$ only gives masses to the up-type quarks, whereas the Higgs supermultiplet with $Y = -\frac{1}{2}$ gives masses to the down-type quarks and the leptons. The complex scalar H_u doublet has a positively charged and a neutral component (H_u^+, H_u^0), while the H_d doublet has a neutral and a negatively charged component (H_d^0, H_d^-). Therefore, the nomenclature for the higgsinos is $\tilde{H}_u^+$, $\tilde{H}_u^0$ and $\tilde{H}_d^0$, $\tilde{H}_d^-$.

This set of particle degrees of freedom defines all chiral supermultiplets of a viable minimal extension of the standard model and is listed in Table 2.3

Name	spin 0	spin 1/2
squarks, quarks (× 3 families)	$(\tilde{u}_L\ \tilde{d}_L)$ $\tilde{u}_R^*$ $\tilde{d}_R^*$	$(u_L\ d_L)$ $u_R^\dagger$ $d_R^\dagger$
sleptons, leptons (× 3 families)	$(\tilde{\nu}\ \tilde{e}_L)$ $\tilde{e}_R^*$	$(\nu\ e_L)$ $e_R^\dagger$
Higgs, higgsinos	$(H_u^+\ H_u^0)$ $(H_d^0\ H_d^-)$	$(\tilde{H}_u^+\ \tilde{H}_u^0)$ $(\tilde{H}_d^0\ \tilde{H}_d^-)$

Table 2.3: Chiral supermultiplets in the MSSM.

The bosonic gauge fields of the SM are members of so-called vector or gauge supermultiplets. Their associated supersymmetric fermionic fields are denoted as gauginos. The eight gluons (g), mediating QCD interactions acquire superpartners called gluinos $\tilde{g}$. The spin 1 gauge bosons, $W^\pm$, W^0 and B^0, associated to the electroweak interaction have spin $\frac{1}{2}$ superpartners referred to as winos, $\tilde{W}^\pm$, $\tilde{W}^0$, and binos $\tilde{B}^0$. Due to the electroweak symmetry breaking, W^0 and B^0 gauge fields mix to form Z^0 and γ bosons with their respective supersymmetric partners denoted as zino $\tilde{Z}^0$ and photino $\tilde{\gamma}$. However, an essential fact is, that because of the electroweak symmetry breaking, also the supersymmetric partners of the SM bosons can mix to form mass eigenstates. This means that the neutral gauginos and higgsinos $\tilde{B}$, $\tilde{W}^0$, $\tilde{H}_u^0$ and $\tilde{H}_d^0$ mix to four mass eigenstates, referred to as neutralinos $\tilde{\chi}^0_{1,\ldots,4}$ and the charged bosonic fields $\tilde{W}^\pm$, $\tilde{H}_u^+$ and $\tilde{H}_d^-$, mix to mass eigenstates denoted as charginos $\tilde{\chi}^\pm_{1,2}$.

The vector supermultiplets of the minimal supersymmetric standard model are listed in Table 2.4.

The masses and couplings of the matter fields are determined in the superpotential of the MSSM which contains the fields corresponding to the chiral supermultiplets from Table 2.3, the Yukawa coupling parameters and an additional mass parameter μ corresponding to the Higgs mass in a supersymmetric extension of the SM. μ should not be

name	spin 1/2	spin 1
gluino, gluon	$\tilde{g}$	g
winos, W bosons	$\tilde{W}^{\pm}\ \tilde{W}^{0}$	$W^{\pm}\ W^{0}$
binos, B bosons	$\tilde{B}^{0}$	B^{0}

Table 2.4: Gauge supermultiplets in the MSSM.

confused with the muon particle and the μ-parameter in the Higgs potential (see Equation 2.24).

R parity

The superpotential of the MSSM is sufficient to produce a phenomenologically viable model, however additional terms, which would be potentially gauge-invariant, are omitted because those contribtutions would violate the lepton (L) and baryon number (B). So far, L- and B-violating processes have never been observed. In addition, one of the most obvious experimental evidences for lepton and baryon number violating processes would be the proton decay which has not been observed either. For example, in a hypothetical SUSY model a proton could eventually decay with a very short lifetime, mediated by a squark. As a consequence, the interactions described in the MSSM respect a further symmetry denoted as R parity [48]. R parity is a multiplicative conserved quantum number and is defined for each particle with spin s as follows:

$$P_R = (-1)^{3(B-L)+2s} \tag{2.35}$$

It follows that fermionic and bosonic fields from the SM have even R parity, $P_R = +1$, while their superpartners have odd R parity, $P_R = -1$. Furthermore, there are particular phenomenological consequences for R parity conserving supersymmetric models. Particle mixing between SM and SUSY particles is forbidden and each interaction vertex in SUSY contains an even number of supersymmetric particles. Therefore, supersymmetric particles can only be produced in pairs and any decay of a SUSY particle results in an odd number of supersymmetric particles. The lightest supersymmetric particle in a decay cascade (LSP) does not decay any further and hence must be stable. If in addition the lightest neutralino, $\tilde{\chi}_1^0$, is the LSP in a given SUSY model, as it is the case in the searches described in Chapter 7 and Chapter 8, it would interact only weakly and can be a suitable candidate for dark matter. However, in a collider experiment such a LSP candidate cannot be detected directly. The signature can be observed indirectly via the momentum imbalance it causes when it escapes the detector.

R parity conservation is in fact not a strict requirement for a supersymmetric model to be valid. There still exist also R parity violating theories, but since it has no further relevance for the work in this thesis a discussion about those models will be left out.

Soft supersymmetry breaking in the MSSM

A conserved and unbroken supersymmetry would imply that the masses of the supersymmetric particles are identical with their SM counterparts, but this is obviously not the case, since none of the proposed SUSY particles have been discovered so far. In a realistic model, supersymmetry must be broken and as a consequence the supersymmetric particles must be heavier than the SM particles.

A feature which makes supersymmetry attractive is that it provides a solution to the hierarchy problem by the cancellation of fermionic and scalar couplings of the same size. A broken supersymmetry can still provide a solution to this problem, however the relation between the two couplings must be adjusted, in order to avoid quadratically divergent radiative corrections.

The origin of the mechanism that lead to spontaneous SUSY breaking is not understood completely and hence an accurate theory has yet to be defined. A widely accepted assumption is that the breaking mechanism results from physics at higher energy scales. Concise discussions about SUSY breaking may be found in [28].

Nevertheless, the effective MSSM Lagrangian can be adapted by an additional term, which considers additional parameters that introduce a soft breaking of supersymmetry [49]:

$$\mathcal{L}_{\mathrm{MSSM}} = \mathcal{L}_{\mathrm{SUSY}} + \mathcal{L}_{\mathrm{soft}}, \tag{2.36}$$

in order to avoid quadratic divergences. The first term, $\mathcal{L}_{\mathrm{SUSY}}$, includes gauge and Yukawa interactions, while $\mathcal{L}_{\mathrm{soft}}$ contains following mass parameters and coupling parameters:

Gaugino mass terms M_1, M_2 and M_3 which correspond to the $U(1)$, $SU(2)$ and $SU(3)$ gauge groups.

Trilinear scalar coupling parameters included in 3×3 matrices in family space $\mathbf{a_u}$, $\mathbf{a_d}$ and $\mathbf{a_e}$. The coupling parameters in the 3×3 matrices correspond to squark, slepton and Higgs fields and are proportional to the Yukawa coupling parameters.

3×3 squark and slepton mass matrices $\mathbf{m^2_Q}$, $\mathbf{m^2_{\bar{u}}}$, $\mathbf{m^2_{\bar{d}}}$, $\mathbf{m^2_L}$ and $\mathbf{m^2_{\bar{e}}}$.

SUSY breaking contributions to the Higgs potential $m^2_{H_u}$, $m^2_{H_d}$ and b.

Before the introduction of the soft supersymmetry breaking Lagrangian $\mathcal{L}_{\mathrm{soft}}$, there was only a single free parameter in the theory which is not present in the standard model, namely the mass parameter μ. Although, $\mathcal{L}_{\mathrm{soft}}$ introduces no less than 104 additional free parameters, which make experimental searches for supersymmetric particles rather complicated.

2.3.2 Phenomenology of top squarks

The masses of the squarks and sleptons of the MSSM can be determined from their respective mass matrices. Mass eigenstates of the first and second generation sfermions are assumed to be almost degenerate, the mixing angles are assumed to be small and their Yukawa couplings are negligible. For sparticles of the third generation, a substantial mixing of left-handed and right-handed particle states ($\tilde{t}_{\mathrm{L}}$, $\tilde{t}_{\mathrm{R}}$), ($\tilde{b}_{\mathrm{L}}$, $\tilde{b}_{\mathrm{R}}$) and ($\tilde{\tau}_{\mathrm{L}}$, $\tilde{\tau}_{\mathrm{R}}$) is possible. The left- and right handed top squarks, which are the supersymmetric partners of the left- and right-handed top quark particle degrees of freedom, mix to form the two mass eigenstates $\tilde{t}_1$ and $\tilde{t}_2$. Significant mass splitting between $\tilde{t}_1$ and $\tilde{t}_2$ is possible due to the large Yukawa coupling of the top quark. Many models suggest a relatively light top squark mass which is assumed to be significantly lower than those of the masses of the squarks from other generations [50, 51]. Due to this fact, the top squark is a promising candidate to search for in high energy physics experiment.

2.3.3 Production and decay of supersymmetric particles at the Large Hadron Collider

At hadron colliders such as the Large Hadron Collider (see Chapter 3), a large variety of supersymmetric particles might be produced. Especially gluinos and squarks should be dominantly produced via the strong interaction from gluon-gluon fusion and gluon-quark-fusion. Furthermore, production of gluinos and squarks is also possible in association with charginos and neutralinos but those mechanisms are suppressed if gluinos and squarks are supposed to be very heavy.

The most favoured production mechanisms are gluino pair production ($pp \rightarrow \tilde{g}\tilde{g}$), squark pair production ($pp \rightarrow \tilde{q}\tilde{q}$) and gluino squark associated production ($pp \rightarrow \tilde{g}\tilde{q}$). The corresponding Feynman diagrams of the production mechanisms via the strong interaction are shown in Figure 2.2.

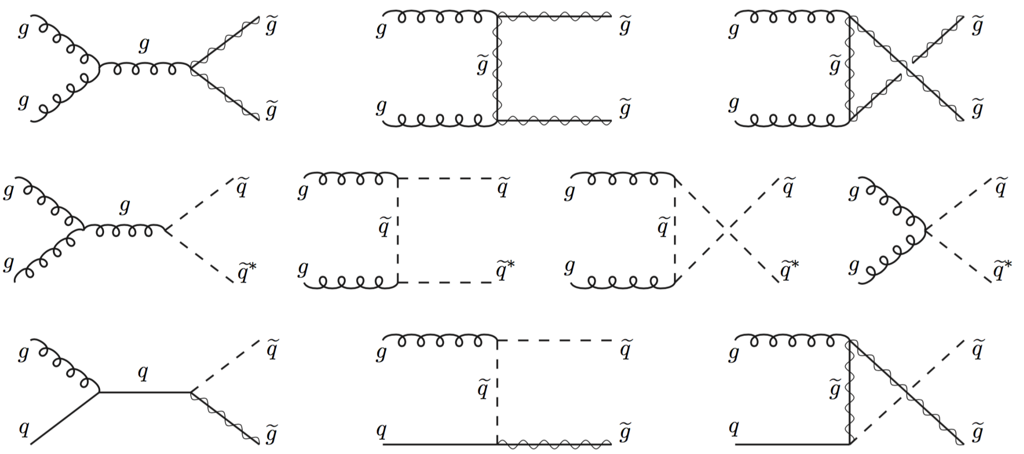


Figure 2.2: Leading order Feynman diagrams for likely production mechanisms via the strong interaction from gluon-gluon and gluon-quark fusion at hadron colliders [28]. $\tilde{g}$ and $\tilde{g}^*$ represent the gluino and its conjugate, and similarly the squark and its conjugate are denoted as $\tilde{q}$ and $\tilde{q}^*$.

Electroweak production of charginos and neutralinos from quark-antiquark annihilation is also possible. Especially if gluinos and squarks are very heavy, the electroweak production of charginos and neutralinos may give the first evidence for supersymmetric particles.

In order to conduct searches for supersymmetric particles at collider experiments, experimental signatures of the decays of supersymmetric particles are important. A large variety of hypothetical supersymmetric signatures may be realised. However, only R parity conserving models, in which the neutralino $\tilde{\chi}_1^0$ is assumed to be the lightest supersymmetric particle, are briefly described in the following.

Gluinos always decay through a squark since they only interact via the strong force. If the gluino is heavier than the squark, the decay $\tilde{g} \rightarrow q\tilde{q}$ is always favoured. Decays into top squarks $\tilde{t}$ or bottom squarks $\tilde{b}$ are more likely, because third generation squarks are assumed to be lighter than squarks from the first two generations. Gluino decays involving charginos, $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}^{\pm}$ and $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}^0$, become more dominant if the squarks are heavier.

Due to the QCD interaction, squarks decay predominantly into gluinos and quarks, $\tilde{q} \rightarrow q\tilde{g}$ if kinematically allowed. Alternatively, if the squark is lighter than the gluino, a squark decays into a quark and a neutralino $\tilde{q} \rightarrow q\tilde{\chi}^0$, or via an intermediate chargino decay $\tilde{q} \rightarrow q\tilde{\chi}^{\pm}$.

Similar as for the squarks, two-body decays of sleptons into a lepton and a neutralino or a chargino are kinematically favoured ($\tilde{l}^{\pm} \rightarrow l^{\pm}\tilde{\chi}^0$, $\tilde{l}^{\pm} \rightarrow \nu\tilde{\chi}^{\pm}$, $\tilde{\nu} \rightarrow \nu\tilde{\chi}^0$, $\tilde{\nu} \rightarrow l^{\pm}\tilde{\chi}^{\pm}$).

Many different decay modes of charginos and neutralinos can be realised, due to the gaugino and higgsino mixing as well as their mass hierarchy. However, only 2-body decays are briefly outlined:

$$\begin{aligned}\tilde{\chi}^0 &\rightarrow \tilde{\chi}^0 Z,\ \tilde{\chi}^0 h^0,\ \tilde{\chi}^{\pm} W^{\pm},\ \tilde{l}^{\pm} l^{\pm},\ \tilde{\nu}\nu \\ \tilde{\chi}^{\pm} &\rightarrow \tilde{\chi}^0 W^{\pm},\ \tilde{\chi}^{\pm} Z,\ \tilde{\chi}^{\pm} h^0,\ \tilde{l}^{\pm}\nu,\ \tilde{\nu} l^{\pm}\end{aligned}$$

Similar as for the squarks and sleptons, three-body decays may also be observed if the aforementioned two-body decays are kinematically forbidden ($\tilde{\chi}^0 \rightarrow ff\tilde{\chi}^0$, $\tilde{\chi}^{\pm} \rightarrow ff\tilde{\chi}^{\pm}$).

A large variety of potential supersymmetric signatures may be realised from the various decay modes mentioned above. Typical searches for supersymmetry are characterised by a large number of energetic jets, since the dominant production mechanism in hadron colliders is the production of gluinos and squarks which interact via the strong force and thus, they likely produce hadrons. Furthermore, the final state also contains the neutral, weakly interacting LSP, which cannot not be detected directly and results in a large momentum imbalance. This is another attribute of a characteristic signature. Additionally, searches for supersymmetry are commonly classified by the number of leptons in the final state.

2.3.4 Simplified models

The large number of free parameters in the MSSM, the huge amount of possible decay scenarios as well as the ignorance about the likely masses of the supersymmetric particles make experimental searches for supersymmetry rather challenging. However, to evaluate searches for new physics and interpret experimental results, an effective framework referred to as the concept of simplified models is used. Instead of an exact Lagrangian of a given BSM theory that describes a set of particles, their interactions and which includes a large number of free parameters, a simplified model is defined by an effective Lagrangian that only describes the interactions of particles of interest. Simplified models only characterise a small number of parameters, such as particle masses, cross-sections or branching ratios, which are directly related to collider physics observables [52–54].

The analyses described in Chapter 7 and Chapter 8 target searches for top squark pair production with a particular decay scenario. In general, the favoured decay structure of top squarks depends on the gaugino content of the neutralino LSP, including the gaugino mass parameters M_1 and M_2, and the mass parameter related to the Higgs boson in the MSSM μ. If the relations of the mass parameters of the target model are defined as $M_1 < M_2$ and $M_1 < |\mu|$, the dominant contribution to the neutralino mass eigenstate comes from the bino $\tilde{B}^0$. The neutralino is predominantly bino-like. As a consequence, the direct decay of the top squark to a top quark and the neutralino $\tilde{t}_1 \rightarrow t\tilde{\chi}_1^0$ is always favoured. A typical simplified model of such a decay scenario is illustrated in Figure 2.3 and referred to as 2-body decay scenario. Usually, the top squark pair production consists of a top squark and its respective antiparticle, although the antiparticle notation (using $\bar{\ }$) for the anti-top squark and its decay products is omitted in the following for simplicity.

However, such a decay can be kinematically forbidden. The decay structure of top squarks is characterised by the mass difference $\Delta m = m(\tilde{t}_1) - m(\tilde{\chi}_1^0)$. If the mass difference of the two supersymmetric particles is less than the mass of the top quark but larger than the masses of the W boson and the b quark, $m(W) + m(b) < \Delta m < m(t)$, the top quark cannot be produced on-shell and the top squark decays via a 3-body decay directly into a b quark, a W boson and the neutralino LSP, $\tilde{t}_1 \rightarrow bW\tilde{\chi}_1^0$. Such a 3-body decay scenario is illustrated in Figure 2.4.

There are even more compressed decay scenarios possible, if the masses of the $\tilde{t}_1$ and the $\tilde{\chi}_1^0$ are closer, $\Delta m < m(W) + m(b)$. In this case the top quark as well as the W boson can only be produced off-shell leading to a direct 4-body decay of the top squark $\tilde{t}_1 \rightarrow bff'\tilde{\chi}_1^0$ as shown in Figure 2.5.

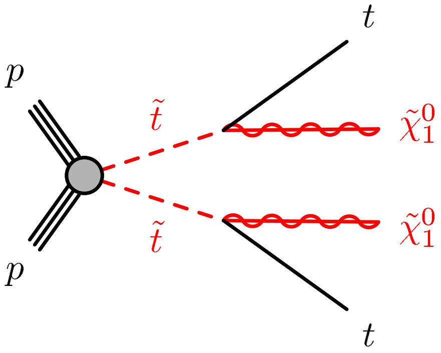

p
p
$\tilde{t}$
$\tilde{t}$
t
$\tilde{\chi}_1^0$
$\tilde{\chi}_1^0$
t

Figure 2.3: Simplified model topology of direct top squark pair production. The top squark directly decays to a top quark and the neutralino LSP. $\tilde{\chi}_1^0$ is predominantly bino-like. By courtesy of the ATLAS supersymmetry physics analysis group.

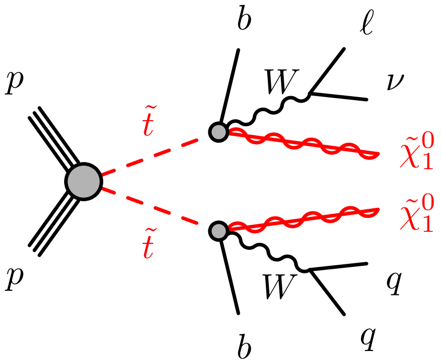

b
ℓ
W
ν
p
$\tilde{t}$
$\tilde{\chi}_1^0$
$\tilde{\chi}_1^0$
$\tilde{t}$
p
W
q
b
q

Figure 2.4: Simplified model topology of direct top squark pair production in a kinematically compressed phase space, referred to as 3-body decay scenario. The decay structure is characterised by $m(W) + m(b) < \Delta m < m(t)$. The top quark is produced off-shell and thus the top squarks directly decay to a b quark, W boson and the neutralino LSP. By courtesy of the ATLAS supersymmetry physics analysis group.

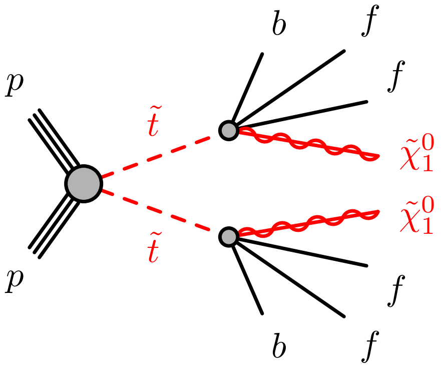

b
f
f
p
$\tilde{t}$
$\tilde{\chi}_1^0$
$\tilde{\chi}_1^0$
$\tilde{t}$
p
f
b
f

Figure 2.5: Simplified model topology of direct top squark pair production in a highly compressed phase space. The mass difference between the supersymmetric particles is so small ($\Delta m < m(W) + m(b)$), that the top quark as well as the W boson can only be produced off-shell and thus the top squarks decay via a 4-body decay mode. By courtesy of the ATLAS supersymmetry physics analysis group.

The kinematic phase space for direct top squark pair production which dominantly decay into a bino-like neutralino LSP is shown in Figure 2.6. The searches presented in this work focus only on top squark pair produced processes in the phase space where the 3-body decay mode is kinematically favoured. This region of interest is indicated in the figure as $\tilde{t}_1 \to bW\tilde{\chi}_1^0$ within the mass range of $m(W) + m(b) < \Delta m < m(t)$.

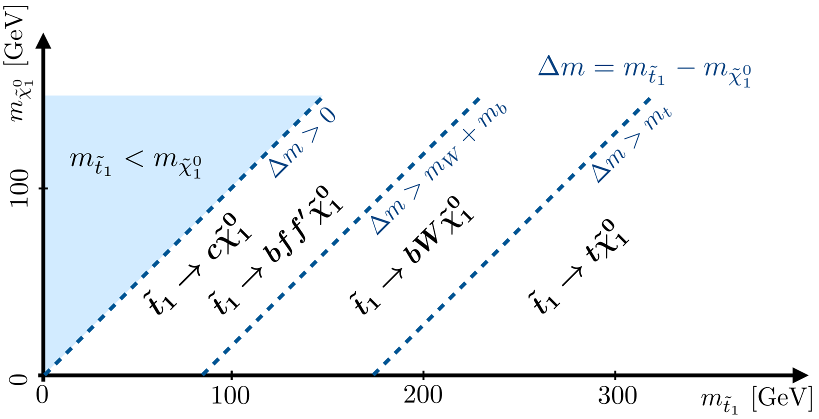

$m_{\tilde{\chi}_1^0}$ [GeV]
100
0
$\Delta m = m_{\tilde{t}_1} - m_{\tilde{\chi}_1^0}$
$m_{\tilde{t}_1} < m_{\tilde{\chi}_1^0}$
$\Delta m > 0$
$\Delta m > m_W + m_b$
$\Delta m > m_t$
$\tilde{t}_1 \to c\tilde{\chi}_1^0$
$\tilde{t}_1 \to bff'\tilde{\chi}_1^0$
$\tilde{t}_1 \to bW\tilde{\chi}_1^0$
$\tilde{t}_1 \to t\tilde{\chi}_1^0$
0
100
200
300
$m_{\tilde{t}_1}$ [GeV]

Figure 2.6: Illustration of the preferred top squark decay modes, assuming that the top squarks decay into a bino-like neutralino. The phase space of interest is spanned by the masses of the top squark ($m(\tilde{t}_1)$) and the lightest neutralino ($m(\tilde{\chi}_1^0)$) [55].

2.4 Current status of searches for top squark pair production

A comprehensive search programme for direct top squark pair production is conducted by the ATLAS and CMS collaborations. Since neither of the performed searches showed evidence of the existence of top squarks, the observed results are interpreted in terms of exclusion limits on the masses of the top squark and the neutralino.

Figure 2.7 illustrates a snapshot of the current exclusion limits at 95 % confidence level from individual searches for top squark pair production from the ATLAS collaboration as of May 2018. The searches are performed on 36.1 fb^{-1} of pp collision data recorded by the ATLAS experiment at a centre-of-mass energy of 13 TeV. Different colours indicate the excluded region from individual analyses, except the blue shaded region which illustrates the summary from dedicated searches conducted at pp collision data at a centre-of-mass energy of 8 TeV. Solid lines represent the observed limit, while the expected limits are shown as dashed lines. The three aforementioned decay modes, each assuming a branching ratio of 100 %, are considered in the various analyses. In addition, a top squark decay to charm quarks ($\tilde{t}_1 \to c\tilde{\chi}_1^0$) is considered. The latter decay scenario is superimposed by the 4-body decay mode. The different searches are separated by the number of leptons in the final state, except of the analysis whose exclusion limit is represented by the grey region, this search is performed on events with one energetic jet and large transverse momentum imbalance [56]. In the phase space of the 2-body decay mode, top squark masses up to 1 TeV are excluded depending on the mass of $\tilde{\chi}_1^0$ [57]. The limit for 3-body decay modes, which is derived by the analysis explained in Chapter 7, extends to top squark masses up to 460 GeV [58] with a neutralino mass of approximately 320 GeV. In the compressed region, 4-body decays of top squarks are excluded up to 400 GeV with a neutralino mass of 360 GeV [59], and potential decays to charm quarks ($\tilde{t}_1 \to c\tilde{\chi}_1^0$) are excluded up to

$m(\tilde{t}_1) = 500$ GeV with $m(\tilde{\chi}_1^0) = 430$ GeV [60].

The CMS collaboration reports exclusion limits up to 1120 GeV on the top squark mass in the phase space of a potential 2-body decay mode [61], while top squarks are excluded up to masses of approximately 580 GeV in the 3-body decay scenario [62]. In the region where 4-body decays are favoured, exclusion limits are determined up to a top squark mass of 590 GeV [63].

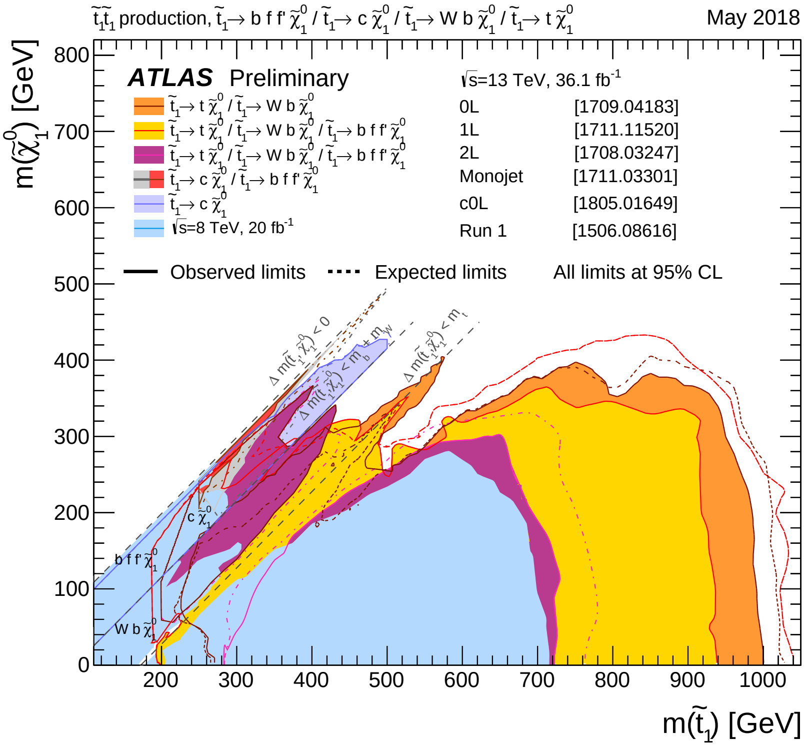


Figure 2.7: Summary of the individual ATLAS searches for top squark pair production based on 36.1 fb^{-1} of *pp* collision data at a centre-of-mass energy 13 TeV [64]. Exclusion limits at 95 % confidence level are shown in the $m(\tilde{t}_1)$ versus $m(\tilde{\chi}_1^0)$ mass plane. The dashed and solid lines show the expected and observed limits, respectively. For the derived limits, statistical as well as systematic uncertainties except for the theoretical signal cross section uncertainty are included. Different decay modes are considered separately with a branching ratio of 100 %, respectively: $\tilde{t}_1 \rightarrow t\tilde{\chi}_1^0$, $\tilde{t}_1 \rightarrow bW\tilde{\chi}_1^0$, $\tilde{t}_1 \rightarrow bff'\tilde{\chi}_1^0$ and $\tilde{t}_1 \rightarrow c\tilde{\chi}_1^0$. The latter two decay modes are superimposed.

Chapter 3

Experimental setup

This section briefly summarises the accelerator complex at CERN, culminating in the Large Hadron Collider (LHC) which accelerates and collides particles at unprecedented energies. Furthermore, essential features of the ATLAS experiment with its sub-detectors and magnet system are described in adequate detail.

3.1 The Large Hadron Collider

The LHC [65] at CERN (derived from the acronym for the French *Conseil Européen pour la Recherche Nucléaire*, or *European Council for Nuclear Research*), located near Geneva at the French-Swiss border, is the world's largest and most powerful particle physics research facility. The machine is installed underground in the existing 26.7 km long circular tunnel which was initially constructed to host the Large Electron Positron Collider (LEP). The particle accelerator is designed to collide proton beams at a centre-of-mass energy of $\sqrt{s} = 14$ TeV. In addition, it is also possible to generate heavy ion collisions with an energy of 2.76 TeV per nucleon.

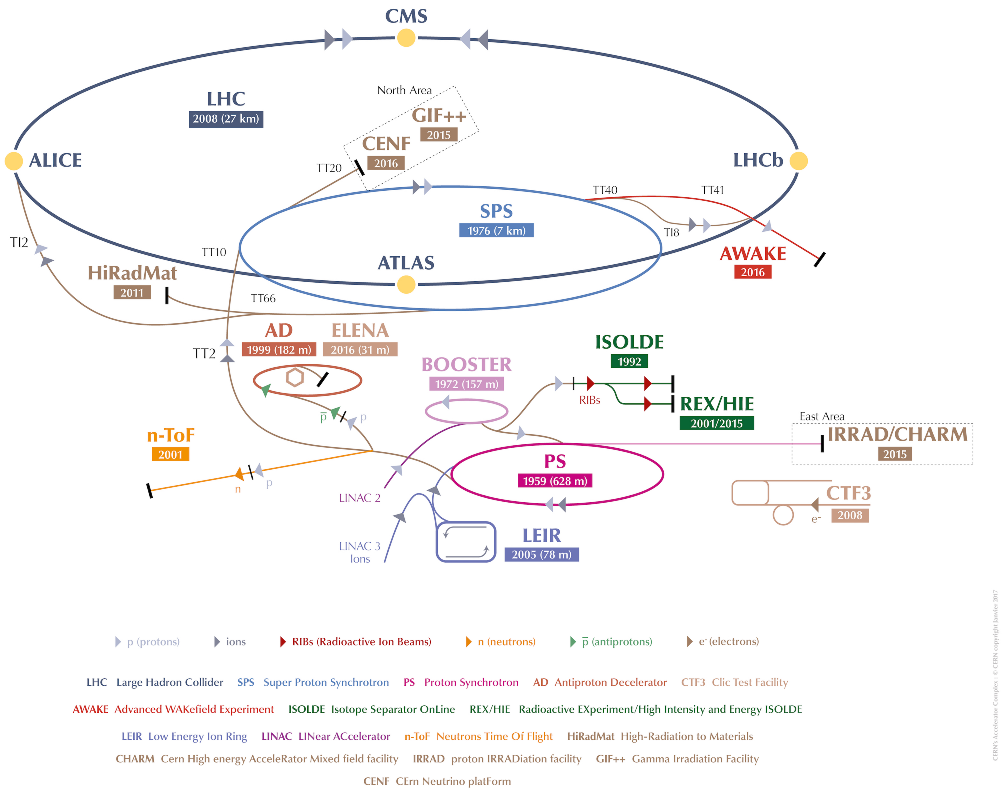

CMS
LHC
2008 (27 km)
North Area
GIF++
2015
CENF
2016
ALICE
LHCb
TT20
TT40
TT41
SPS
1976 (7 km)
TI8
TI2
TT10
ATLAS
AWAKE
2016
HiRadMat
2011
TT66
AD
1999 (182 m)
ELENA
2016 (31 m)
ISOLDE
1992
TT2
BOOSTER
1972 (157 m)
RIBs
REX/HIE
2001/2015
East Area
IRRAD/CHARM
2015
p̄
p
n-ToF
2001
n
p
PS
1959 (628 m)
LINAC 2
CTF3
2008
e⁻
LINAC 3
Ions
LEIR
2005 (78 m)
p (protons)
ions
RIBs (Radioactive Ion Beams)
n (neutrons)
p̄ (antiprotons)
e⁻ (electrons)
LHC Large Hadron Collider
SPS Super Proton Synchrotron
PS Proton Synchrotron
AD Antiproton Decelerator
CTF3 Clic Test Facility
AWAKE Advanced WAKefield Experiment
ISOLDE Isotope Separator OnLine
REX/HIE Radioactive EXperiment/High Intensity and Energy ISOLDE
LEIR Low Energy Ion Ring
LINAC LINear ACcelerator
n-ToF Neutrons Time Of Flight
HiRadMat High-Radiation to Materials
CHARM Cern High energy AcceleRator Mixed field facility
IRRAD proton IRRADiation facility
GIF++ Gamma Irradiation Facility
CENF CErn Neutrino platForm
CERN's Accelerator Complex : © CERN copyright Janvier 2017

Figure 3.1: CERN accelerator complex [66]. At first, protons are accelerated by the LINAC2 to energies up to 50 MeV, afterwards further acceleration up to 1.4 GeV is achieved by the BOOSTER followed by the PS (26 GeV). In the SPS they got accelerated up to 450 GeV and finally injected in the LHC where they are able to achieve a maximum energy of 7 TeV.

Such high particle energies cannot be accomplished by a single source, thus the particle beams have to be pre-accelerated by a series of accelerator systems, which were already used for several other experiments and have been upgraded to meet the needs of LHC. The layout of the injection and accelerator chain at CERN is illustrated in Figure 3.1. The LHC is supplied by bunches of protons from a simple hydrogen source, electrons are stripped away via a powerful electrical discharge and the protons are first concentrated and accelerated up to 50 MeV by the LINAC2 linear accelerator. In the next step, the bunches are accelerated up to 1.4 GeV by the Proton Synchrotron Booster (BOOSTER) followed by the Proton Synchrotron (PS) (26 GeV). Afterwards they are further accelerated to 450 GeV in the Super Proton Synchrotron (SPS) and finally injected in the LHC [67].

The LHC machine itself is built-up of 1232 superconducting dipole magnets. The key feature of those magnets is their twin bore magnet design to accommodate two beam channels side by side, where the bunches of particles are circulated in clockwise and counter-clockwise direction. The protons in the beam channels are collided in vacuum. The maximum magnetic field of the dipole magnets is 8.33 T, necessary to keep the bunches in the orbit. The magnet system is maintained by NbTi Rutherford cables and cooled down to a temperature of 1.9 K using superfluid helium. In addition to the dipole magnets, several quadrupole and correction magnets are located along the beam pipe to focus the particle beams, while the acceleration of the particles is achieved by a superconducting radio frequency cavity system concentrated at one point along the collider.

A maximum number of 2808 bunches can circulate at a time in the LHC with a bunch spacing of 25 ns. This results in a design value for the instantaneous luminosity for proton-proton collisions of $\mathcal{L} = 10^{34}$ cm^{2}s^{-1}. The luminosity $\mathcal{L}$ is another essential parameter of the collider, apart from the centre-of-mass energy. It is characterised by beam parameters only and is defined as [65]:

$$\mathcal{L} = \frac{N_b^2 n_b f_{\mathrm{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} F. \tag{3.1}$$

N_b is the number of particles per bunch and n_b is the number of bunches per beam. f_{rev} is the revolution frequency, γ_r the relativistic gamma factor and ϵ_n is the normalised transverse beam emittance. The beta function at the interaction point (IP) is denoted as β^* and F is a geometric luminosity reduction factor based on beam crossing angles at the collision point. Furthermore, $\mathcal{L}$ provides information about the produced event rate $\frac{dN}{dt}$, which can be expressed as:

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma. \tag{3.2}$$

σ is the cross section of a certain process under study given in barn (b $= 10^{-24}$ cm^2) and reflects the probability of the process to be produced in the collision. The integrated luminosity L can be determined by integration over time:

$$L = \int \mathcal{L} dt = \frac{N}{\sigma} \tag{3.3}$$

This quantity is commonly used to specify the number of collisions, its unit is given in b^{-1}. It also allows to derive the expected number of events of a specific process from its cross section.

Since the protons are grouped to bunches of high density as well as the high frequency of collisions, any hard interaction is accompanied by many concurrent soft *pp* collisions, referred to as pileup. Proper triggering and reconstruction of the hard interaction event is therefore more difficult and thus pileup collisions have also a strong impact on the physics measurements and searches.

There are four points along the collider ring where the particle beams are brought to collisions, also called interaction points. Following four main experimental insertions are located at these points to record and investigate the particle collisions:

- **ALICE** (A large Ion Collider Experiment). This heavy-ion detector is designed to address the physics of the quark-gluon plasma at extreme energy densities and temperatures in heavy-ion collisions [68].

- **ATLAS**. One of the two general purpose experiments. Physics motivations include precise investigations of SM phenomena such as electroweak symmetry breaking and searches for physics beyond the Standard Model [69]. A brief overview of the ATLAS detector follows in Section 3.2.

- **CMS** (Compact Muon Solenoid). Another general purpose detector that shares a similar purpose, namely high precision measurements of SM processes and searches for new physics phenomena [70].

- **LHCb** (LHC beauty). The physics program of LHCb comprises investigations of rare decays of B-hadrons as well as CP violating processes to study the differences of matter and antimatter [71].

The LHC successfully circulated beams for the first time in September 2008 reaching a novel energy frontier and originating a new era in modern particle physics. The main research programme was initiated in 2010 with proton-proton collisions at a centre-of-mass energy of 7 TeV. After a technical stop at the end of 2009 required for maintenance work and an unintentional halt due to an incident back in 2008 which caused a severe mechanical damage in one sector of the collider [72]. After another year of delivering pp collisions at $\sqrt{s} = 7$ TeV in 2011, the LHC increased the beam energy further and continued to operate at a centre-of-mass energy of 8 TeV in 2012. In February 2013 the LHC successfully finished its first period of operation, referred to as Run 1, and went in a shut down period lasting about two years which was mandatory for upgrade and maintenance work of the collider as well as the experiments. Several outstanding achievements have been reached with the highlight of the discovery of the Higgs boson by the ATLAS and CMS collaborations in 2012 [26, 27]. From 2015 until 2018 the LHC succeeded with the Run 2 campaign and provided pp collisions at the unprecedented centre-of-mass energy of $\sqrt{s} = 13$ TeV, while in December 2018 the second long shut down period has begun and will last another two years. Again, this shut down period is dedicated to maintenance work on the collider ring as well as on the experiments along the collider ring. Figure 3.2 illustrates the pp collision data delivered to the ATLAS experiment for each year respectively given in integrated luminosity and represents the very successful data taking periods. The total recorded data from pp collisions by the ATLAS experiment amounts to 45.0 pb^{-1} in 2010, 5.1 fb^{-1} in 2011 and 21.3 fb^{-1} in the year 2012. In Run 2, data corresponding to an integrated luminosity of 3.9 fb^{-1}, 35.6 fb^{-1}, 46.9 fb^{-1} and 60.6 fb^{-1} for the years between 2015 until 2018 have been recorded. Figure 3.3 shows a summary of the average pileup events for Run 1 and Run 2 respectively. The increase of the mean pileup between the two periods is due to the decrease of the bunch spacing from 50 ns (Run 1) to 25 ns (Run 2).

A more comprehensive description of the LHC and the injection chain may be found in [65, 67, 73].

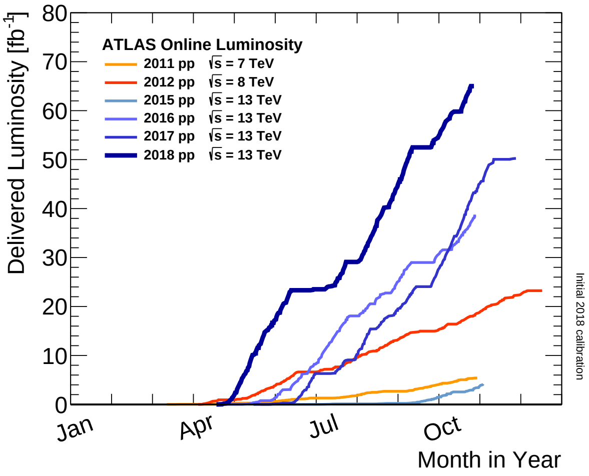
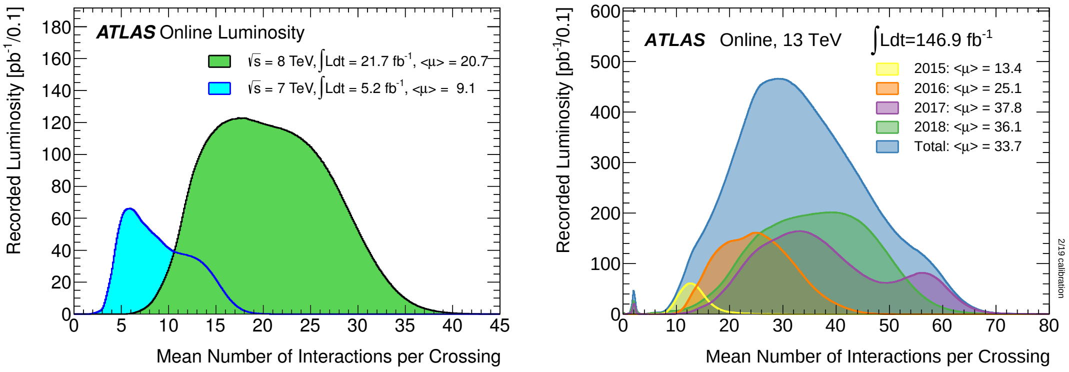


Figure 3.2: Integrated luminosity of stable beams for *pp* collisions delivered to ATLAS over the period of twelve months [74].



(a) Run 1

(b) Run 2

Figure 3.3: Mean number of interactions per bunch crossing for Run 1 (a) and Run 2 (b) of *pp* collision data at 7, 8 and 13 TeV centre-of-mass energy. Data recorded by the ATLAS detector during stable beams is shown, and the total integrated luminosity and the mean number of interactions per bunch crossing $< \mu >$ per year are given in the figures [74, 75].

3.2 The ATLAS detector

This section gives a brief summary of [69, 76, 77], where comprehensive and precise descriptions of the ATLAS experiment can be found.

The ATLAS detector hosted at Point 1 of the LHC collider ring is one of the two general purpose experiments of the LHC research program. A schematic layout of the detector is shown in Figure 3.4. Its total size is approximately 44 m in length and 25 m in height. It is built up of various sub-detector systems. The individual sub-detectors are either cylindrically arranged around the beam axis in the central part of the detector, denoted as barrel region, or as disks perpendicular to the beam axis on both sides of the interaction point, referred to as end-caps. Hence, hermetic geometric coverage is guaranteed.

Closest to the collision point is the inner detector (ID) located. It is surrounded by a superconducting solenoid which provides a 2 T axial magnetic field. The ID consists of silicon-based high-resolution pixel and strip detectors in combination with straw-tube tracking detectors for momentum and vertex measurements of charged particles as well as pattern recognition.

The calorimeter system assures excellent energy and position resolution and contains electromagnetic and hadronic sampling calorimeters based on liquid-argon (LAr) technology as well as a scintillator-tile hadronic calorimeter in the central part of the detector.

At last, the muon spectrometer (MS) is located furthest from the collision point. To achieve good bending power and acquire excellent muon momentum resolution a toroidal magnetic field provided by an air-core superconducting toroid system is incorporated in this part of the detector. The provided magnetic field strength in the toroid system ranges from $0.5 - 1$ T, depending on the location in the detector.

Since only a fraction of all the collisions are of particular interest, ATLAS uses a two-level trigger system in Run 2 [78] in order to reduce the number of events to less than a few hundreds per second for storage and subsequent data analysis.

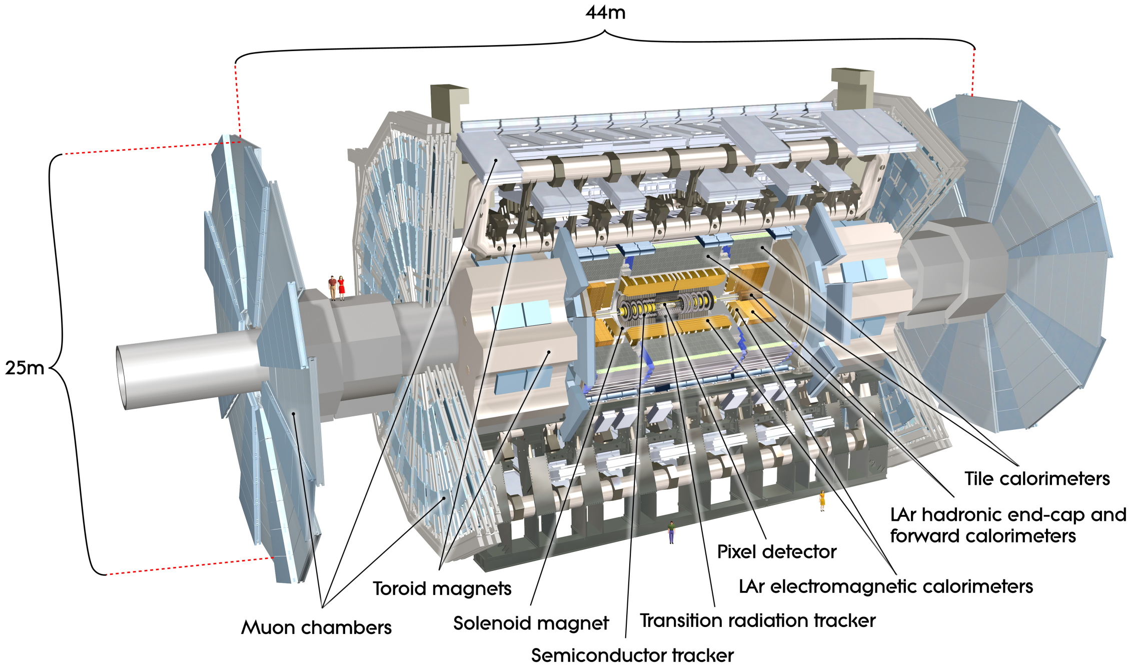

44m
25m
Tile calorimeters
LAr hadronic end-cap and forward calorimeters
Pixel detector
LAr electromagnetic calorimeters
Toroid magnets
Muon chambers
Solenoid magnet
Transition radiation tracker
Semiconductor tracker

Figure 3.4: ATLAS detector layout [79].

Coordinate system

The coordinate system used by ATLAS is right-handed with its origin at the nominal interaction point in the centre of the detector. The x-axis points radially from the IP to the centre of the LHC, the y-axis is pointing upwards and the z-axis points along the beam pipe. The plane spanned by the x- and y-axes is referred to as the transverse plane. The azimuthal angle, denoted as ϕ is measured from the x-axis around the beam axis, and the polar angle θ is measured from the z-axis. More conventional in high energy physics is the pseudorapidity which is defined as $\eta = -\ln \tan(\theta/2)$, because differences in pseudorapidity are Lorentz invariant under boosts along the beam axis. In the case of massive objects, the rapidity $y = 1/2 \ln[(E + p_z)/(E - p_z)]$ is used instead of pseudorapidity. The distance between two objects in the $\eta - \phi$ plane is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\ \phi^2}$.

Figure 3.5 shows the cross-section of one quarter of the ATLAS detector and illustrates the different signatures deposited by various charged and uncharged particles in the different detector systems, except for neutrinos which do not interact with any material in the detector.

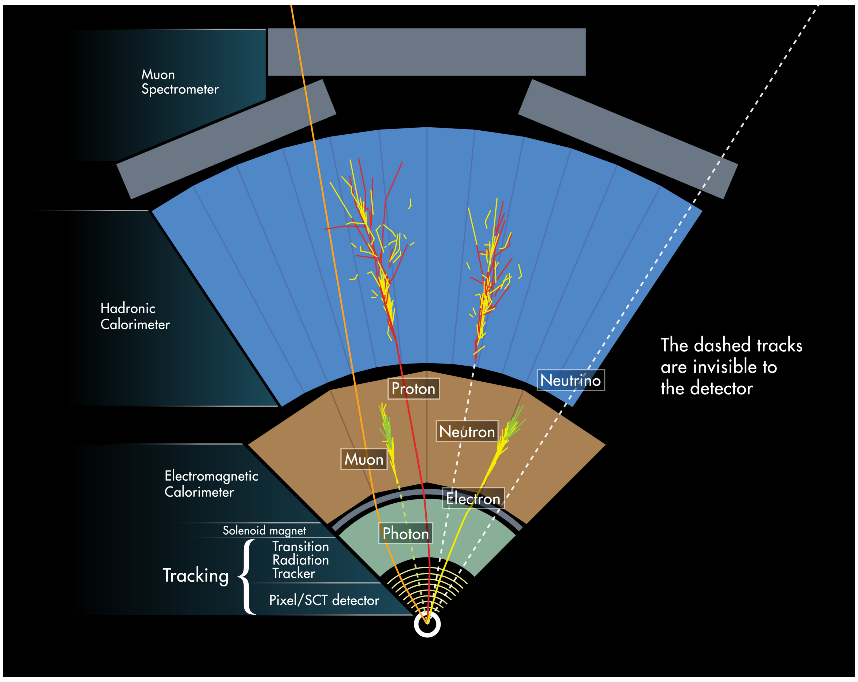

Muon Spectrometer
Hadronic Calorimeter
Electromagnetic Calorimeter
Solenoid magnet
Tracking
Transition Radiation Tracker
Pixel/SCT detector
Proton
Neutrino
Neutron
Muon
Electron
Photon
The dashed tracks are invisible to the detector

Figure 3.5: Cross-sectional view of one quarter of the ATLAS detector showing tracks and showers from different charged and uncharged particles [80].

3.2.1 Magnet system

The magnet system is a key ingredient for the detector design. As charged particles move along curved trajectories within a magnetic field, it is possible to determine the particle's momentum by:

$$p_{\mathrm{T}} = qBr \tag{3.4}$$

Here, p_{T} stands for the momentum orthogonal to the direction of the magnetic field B, q denotes the particle's charge and r is the radius of the curvature. In addition, the sign of the electric charge can be determined from the direction of the deflection (left - right) within the magnetic field.

The ATLAS magnet system consists of one superconducting central solenoid and three toriod magnets. The superconducting magnets are operated at a temperature of 4.5 K.

Central solenoid

The central solenoid has a total length of 5.8 m with an inner and outer diameter of 2.46 m and 2.56 m respectively. It provides an axial field of 2 T and has a stored energy of 40 MJ, provided by a single-layer coil which is wound with an Al-stabilised NbTi conductor. The radiative thickness of the solenoid had to be minimised to require sufficient performance of the calorimeter system, which is surrounded by the solenoid. Thus, the solenoid share a common vacuum vessel and cryostat with the LAr calorimeter. The flux of the magnetic field is returned by the steel of the hadronic calorimeter and its girder structure.

Toroid system

The muon spectrometer system is immersed by one barrel and two end-cap toroid magnet systems to achieve sufficient muon momentum performance. The provided magnetic field strength is 0.5 T and 1 T for the barrel and the end-caps, respectively. The overall size of the barrel toroid is 25.3 m in length with an inner and outer diameter of 9.4 m and 20.1 m, respectively. It consists of 8 coils in stainless steel vacuum vessels which surround the calorimeters as well as the two end-cap toroids. The total stored energy is 1.1 GJ.

The end-cap toroids provide the magnetic field for the end-cap regions of the muon spectrometer system. Each end-cap consists of a single cold mass built up from eight flat, square coil units and eight keystone wedges.

3.2.2 Inner detector

Tracking detectors are essential in particle physics experiments in order to detect charged particles and measure their momentum. The basic principle of tracking detectors is that a charged particle leaves a trail of ionised atoms and electrons if the charged particle traverses the detector medium. The trajectory of the charged particle can be reconstructed by detecting the ionised particles. In the case of semiconductor-based tracking detectors the charged particle leaves a trail of electron-hole pairs when it traverses the semiconductor material, and the trajectory is reconstructed based on the detection of the electrons and holes which drift through the semiconductor material.

The conceptual layout of the inner detector system is illustrated in Figure 3.6 (only the barrel region is shown, while the end-caps are omitted in this view). Closest to the interaction point, the particle flux is the highest and thereby a high track density is present. Thus, good spatial resolution in this part of the detector is vital. The tracking system in the ATLAS experiment is built up from pixel and silicon microstrip (SCT) trackers surrounded by a transitions radiation tracker (TRT). Robust pattern recognition and high precision measurements of charged particle tracks in the $R-\phi$ plane as well as along the z-axis are granted by the combination of those three independent detector techniques, which take up a total length of 5.3 m with a diameter of about 2.5 m.

The high precision pixel and SCT systems cover the region $|\eta| < 2.5$ and allow precise impact parameter measurements, vertexing for heavy-flavour objects and $\tau-$lepton tagging.

The pixel detector consists of three layers in the barrel region, while the pixel end-caps consists of three disks on either side of the interaction point. The minimum size of a pixel is 50×400 μm^2 $(R-\phi \times z)$ and the pixel sensors are identical in the whole pixel detector. The intrinsic accuracies in the $R-\phi$ plane are 10 μm and 115 μm along the z-axis in the barrel as well as both end-caps. The total number of readout channels of the pixel detector is approximately 80.4 million.

During the shut down period between Run 1 and Run 2 major upgrades to the various detector systems had been accomplished, also to the inner detector. A fourth layer with another 6.02 million pixels had been added to the pixel detector between a new, smaller beam pipe and the innermost pixel layer at a radius of about 33 mm, referred to as the insertable B-Layer (IBL). This was necessary to restore and further improve the quality of the tracking precision and thereby enhance the vertexing and b tagging efficiencies, because irreparable deficiencies in the pixel detector will appear with time for example due to radiation effects. Furthermore, readout inefficiencies due to high occupancy especially in the first pixel layer, which are caused by higher luminosity and larger event pileup are compensated by the IBL [81].

The SCT barrel consists of four layers of 6.4 cm long double-sided silicon strip modules with a rectangular shape. The strips of the double-sided sensors are slightly twisted to each other and form a small angle of 40 mrad, with one set of strips parallel to the beam axis to measure all coordinates. The strip pitch of the sensors in the barrel region is 80 μm and the intrinsic accuracies per module are 17 μm ($R-\phi$) and 580 μm (z). The SCT end-caps comprise nine disks on either side located at a distance of 810 mm < |z| < 2797 mm to the nominal interaction point. The sensors applied in the end-cap region have a trapezoidal shape with a set of strips arranged radially and the other set of strips at a small angle ($\sim$ 40 mrad). The mean strip pitch is also 80 μm and the intrinsic accuracies of the modules are approximately 17 × 5800 μm ($R-\phi \times z$). The SCT system has a total of 6.3 million readout channels.

The TRT system is built up from 4 mm wide straw tubes filled with a Xe based gas mixture. Transition-radiation photons enhance the capability of electron identification. An estimated number of $\sim$ 36 hits per track, provided by the straw tubes, in combination with a longer measured track length compensates its lower precision, compared to the pixel and SCT systems. This system covers a region up to $|\eta| < 2.0$ and only provides $R-\phi$ information with an intrinsic accuracy of 130 μm per straw. Alike the other two systems in the inner detector, the TRT can be divided into a barrel part with 144 cm long straws, which are aligned parallel to the beam axis, and an end-cap region where the straws with a length of 37 cm are radially arranged in wheels. The total number of readout channels in the TRT is roughly 351000.

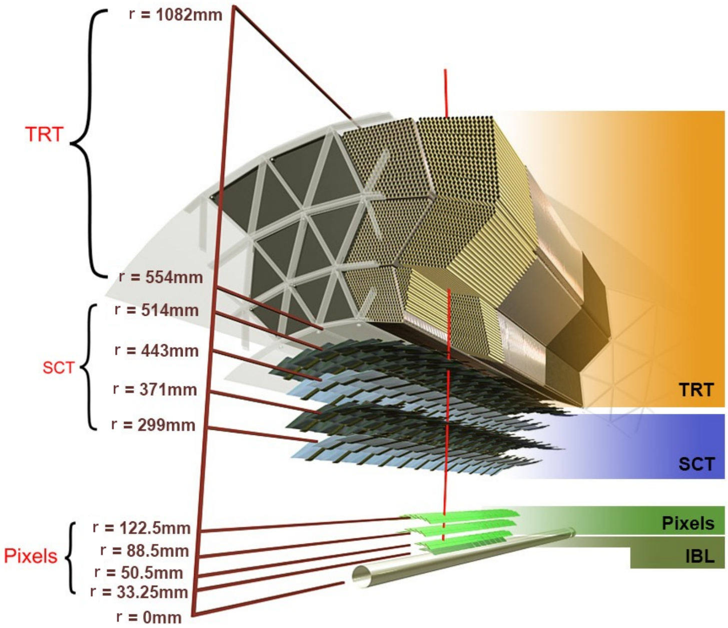


Figure 3.6: Layout of the inner detector system including the insertable B-Layer (IBL) implemented during the first long shut down, pixel and silicon microstrip tracker and the transition radiation tracker [82].

3.2.3 Calorimetry

In general, particles initiate particle showers when traversing characteristic material, thereby secondary particles deposit energy and produce further particles until the total energy is absorbed. The composition and the dimensions of such particle showers depend on the type and the energy of the primary particle. Calorimeters aim to fully absorb particles and hence the signals produced in a calorimeter are a measure for the particle's energy. In principle, they can be distinguished into electromagnetic (EM) and hadronic calorimeters to reconstruct either the energies of electrons, positrons and photons or the energies of hadrons. Two essential characteristics for calorimeters are the energy resolution and the thickness given in radiation length X_0 or interaction length λ, for EM or hadronic calorimeters, respectively. The latter reflects the distance at which the particle looses a certain amount of energy and the shower development through the detector material.

Figure 3.7 illustrates the various EM and hadronic calorimeter systems in the ATLAS detector covering a total of $|\eta| < 4.9$. Both sub-systems are sampling calorimeters, meaning that the active material where the signals are measured is intersected by multiple passive absorber layers. Such construction type is necessary to keep the detector design compact and reduce material costs, although it has a negative effect on the energy resolution, because only the amount of energy can be measured, that is deposited in the active material.

Electromagnetic calorimeter

Similar to the other sub-systems, the EM calorimeter can be sub-divided into a barrel region ($|\eta| < 1.475$) and two end-caps (EMEC; $1.375 < |\eta| < 3.2$), each housed in its own cryostat. As mentioned in Section 3.2.1, the barrel unit shares a common cryogenic system with the central solenoid to reduce material in front of the calorimeters and provide optimal performance, while the cryostats of the end-caps include also the hadronic end-cap and the forward calorimeter systems described later in this section. Liquid-argon (LAr) acts as an active material, interleaved by lead absorber plates. LAr has been chosen because of its intrinsic linear behaviour and radiation hardness. The electrodes and the passive plates are arranged in an accordion geometry to guarantee a complete ϕ symmetry without azimuthal cracks. The barrel consists of two identical half-barrels with a tiny gap of 4 mm at $z = 0$ and each end-cap unit includes a coaxial inner wheel ($1.375 < |\eta| < 2.5$) and outer wheel ($2.5 < |\eta| < 3.2$). The η region that matches with the ID system (Section 3.2.2) has a finer granularity to allow precise measurements of electrons, positrons and photons, while the residual part of the calorimeter system has a slightly coarser granularity. Nevertheless sufficient enough for accurate jet reconstruction (see Section 4.3.4) and measurements of the momentum imbalance in the transverse plane (see Section 4.3.7), which is a vital observable for searches for supersymmetric particles. At the innermost section of the EM calorimeter is an active LAr layer covering an $|\eta|$ region up to 1.8 with a depth of 1.1 cm (0.5 cm) in the barrel (end-cap) to correct energy losses by electromagnetic interacting particles. The EM calorimeter has thickness of at least 22 X_0 in the barrel and at least 24 X_0 in the end-caps. Where necessary, the depth increases up to 33 X_0 and 38 X_0 for the barrel and the end-caps, respectively. The required energy resolution of the electromagnetic calorimeter in the ATLAS detector which is based on the performance goals of the experiment is [69]:

$$\frac{\sigma(E)}{E} \sim \frac{10\%}{\sqrt{E}} + 0.007 \tag{3.5}$$

Hadronic calorimeter

Directly on top of the EM calorimeter is the hadronic tile calorimeter located, consisting of a central barrel ($|\eta| < 1.0$) and two extended barrels ranging from $|\eta| = 0.8$ to $|\eta| = 1.7$. Steel is used as the absorber material and scintillating tiles are used as the active detector medium. The signal produced in the scintillating tiles is guided by wavelength shifters into photomultiplier tubes. The front-end electronics are also able to combine sub-sets of readout channels to form trigger elements for the Level-1 trigger (see Section 3.2.5). The whole tile calorimeter is segmented into three layers which sum up to a total depth of 7.4 λ.

Larger pseudorapidities are covered by the hadronic end-cap calorimeters (HEC) in each direction and consist of two independent wheels per end-cap. They are located directly behind the EMEC and share a common cryogenic system. The coverage of one HEC ranges from $|\eta| = 1.5$ to $|\eta| = 3.2$ and thereby slightly overlaps with the tile calorimeter in the lower η range and with the forward calorimeter system in the upper η range. The front wheel is built up from 25 mm thick copper plates (absorber) interleaved by 8.5 mm wide gaps which are filled by LAr as the active medium. The layout of the rear wheel is similar but has a coarser sampling fraction using 50 mm thick copper plates. Front and rear wheel of the HEC are both composed of two segments which sum up to a total of four layers for each HEC system and results in a total thickness of approximately 10 λ.

Lastly, the forward calorimeter (FCal) system reaches a coverage of $3.1 < |\eta| < 4.9$. As mentioned above it is also integrated in the end-cap cryostats and significantly reduces the radiation background that reach the muon spectrometer. It is also based on a LAr technology and consists of three modules per end-cap. The first module is made of copper (absorber) and optimised for EM measurements, while the second and third modules are made of tungsten absorber plates suitable for hadronic measurements. All three modules together lead to a total depth of about 10 λ for each FCal end-cap, respectively.

The required energy resolution of the hadronic calorimeter system in the ATLAS detector which is based on the performance goals of the experiment is [69]:

$$\frac{\sigma(E)}{E} \sim \frac{50\%}{\sqrt{E}} + 0.03 \tag{3.6}$$

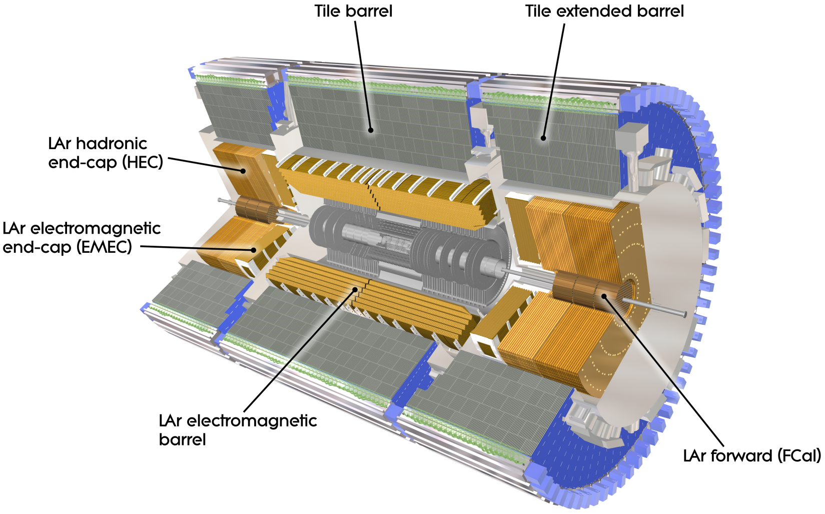

Tile barrel
Tile extended barrel
LAr hadronic end-cap (HEC)
LAr electromagnetic end-cap (EMEC)
LAr electromagnetic barrel
LAr forward (FCal)

Figure 3.7: Assembly of the electromagnetic and hadronic calorimeter system in the ATLAS experiment [83].

3.2.4 Muon spectrometer

Unlike other particles, muons are not stopped by any of the calorimeters in the ATLAS detectors, because they are able to penetrate several metres of material without showering.

Hence, the muon spectrometer system builds the outermost part in the detector where they are the only particles likely to register a signal.

Due to the separate toroidal magnetic field, mentioned in Section 3.2.1, a standalone reconstruction of muons is possible with high resolution. The schematic layout of the ATLAS muon system is illustrated in Figure 3.8. High precision chambers arranged in three concentric cylindrical layers are located in between and on the eight coils of the superconducting toroid magnet in the barrel region. In the end-cap region they are accommodated in three layers in front of and behind the end-cap toroid magnets and arranged as large wheels perpendicular to the beam axis. The total coverage of the muon system goes up to $|\eta| < 2.7$ while triggering on muons is also possible with separate trigger chambers up to $|\eta| < 2.4$. Magnetic deflection in the central part ($|\eta| < 1.4$) is provided by the barrel toriod, while the end-cap magnets provide magnetic bending from $|\eta| = 1.6$ up to $|\eta| = 2.7$. In the small transition region between 1.4 and 1.6 the magnetic deflection is provided by a combination of both magnet systems. This configuration provides a magnetic field mostly orthogonal to the muon trajectories.

Most of the η range is covered by Monitored Drift Tubes (MDT) for high precision measurements of muon tracks. The MDTs are operated with an $\mathrm{Ar/CO_{(2)}}$ gas mixture under pressure (3 bar). Depending on their location in the detector they consist of three to eight layers of aluminium drift tubes with a 50 μm thin tungsten-rhenium wire in the centre of the tube and achieve an average resolution of approximately 35 μm per chamber.

In addition, multi-wire proportional chambers with cathodes segmented into strips, referred to as Cathode Strip Chambers (CSC) are mounted on the innermost end-cap plane covering the region between $2.0 < |\eta| < 2.7$. CSC chambers have a higher granularity and are able to withstand the demanding rate and background conditions. The whole CSC system consists of one small and one large wheel with eight chambers each, their wires are radially oriented, while the two cathodes are segmented with the strips either oriented parallel or perpendicular to the wires to provide measurements of both coordinates. The chambers are operated with the same gas mixture as the MDTs. In the bending plane the CSCs achieve a resolution of about 40 μm and 5 mm in the traverse plane.

The muon triggering system is based on different technologies which provide faster readout. The purpose of the system is threefold. Detectors used for triggering should be precise enough to trigger on well defined p_{T} thresholds, identification of bunch crossing and they should also be able to measure the coordinate in the direction orthogonal to the coordinate determined by the precision tracking chambers. For those reasons, three concentric cylindrical layers of Resistive Plate Chambers (RPC) are used in the barrel region and a total of nine layers of Thin Gap Chambers (TGC) are applied either in the middle or the inner layer of the end-caps. Both types of detectors deliver signals with a spread of roughly $15 - 25$ ns.

3.2.5 Trigger and data acquisition

Another essential component of any collider experiment is the trigger logic, because this system builds up the decisions whether events will be kept and stored or discarded. With a zero-suppressed event size of about $1 - 2$ MB, together with the short bunch crossing of 25 ns, the produced data rate would be of the order of multiple TB/s, which is in general not desirable to record. Additionally, a large fraction of all the interactions is not of particular interest. An appropriate trigger strategy reduces the interaction rate from 40 MHz down to 1 kHz. This reduction is usually achieved in multiple stages by a hardware based first-level trigger and a subsequent software based high-level trigger.

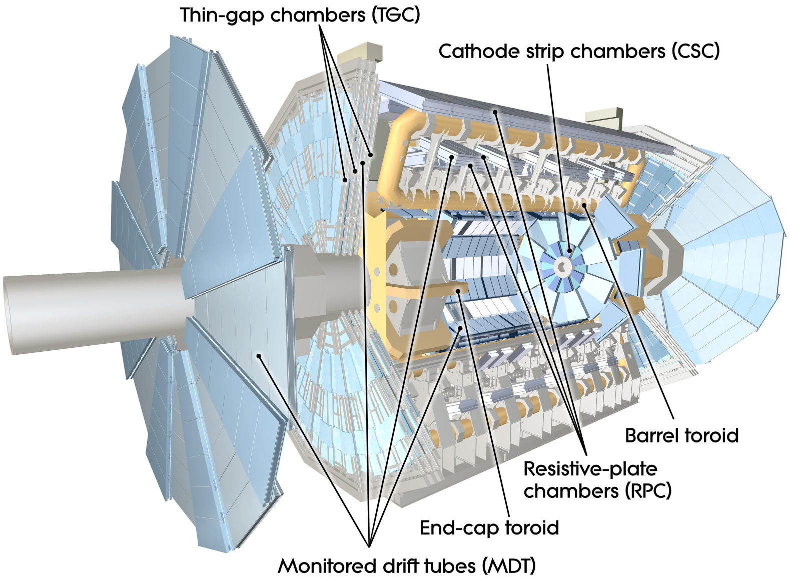

Thin-gap chambers (TGC)
Cathode strip chambers (CSC)
Barrel toroid
Resistive-plate chambers (RPC)
End-cap toroid
Monitored drift tubes (MDT)

Figure 3.8: Schematic layout of the ATLAS muon spectrometer consisting of high precision MDT and CSC chambers as well as RPC and TGC trigger chambers [84].

The ATLAS collaboration adapted a major change to the trigger structure during the first long shut down period, otherwise the maximum allowed trigger rate would have been exceeded in the following data taking periods. The optimised two-stage trigger logic is shown in Figure 3.9 and reported in [78, 85].

The first stage of the ATLAS trigger system is referred to as the Level-1 trigger (L1) and is based on hardware processors. The L1 trigger decision, whether an event will be recorded, is formed by the Central Trigger Processor (CTP) after receiving inputs from the L1 Calo, L1 Muon trigger units and many other sub-systems such as luminosity counter, etc. The decision from L1 Calo and L1 Muon in turn depend on the detected signals in the dedicated barrel and end-cap sectors of the calorimeter system and the muon spectrometer respectively. The CTP is also responsible for applying preventive dead-time to the readout. After an event is accepted by the L1, it is buffered in the readout system and further processed by the final stage, denoted as high-level trigger (HLT).

The software based HLT logic receives information in which part of the detector certain signals had been registered (Region-of-Interest) from the Level-1 trigger. This information is then further processed for regional reconstruction in the various trigger algorithms. After HLT acceptance, the event is transferred to the local storage for further offline reconstruction, at the experimental site as well as exported to the TIER-0 facility at CERN's computing centre. During the whole process the maximum output rate of the L1 which is roughly 100 kHz is reduced to an average HLT physics output rate of 1 kHz.

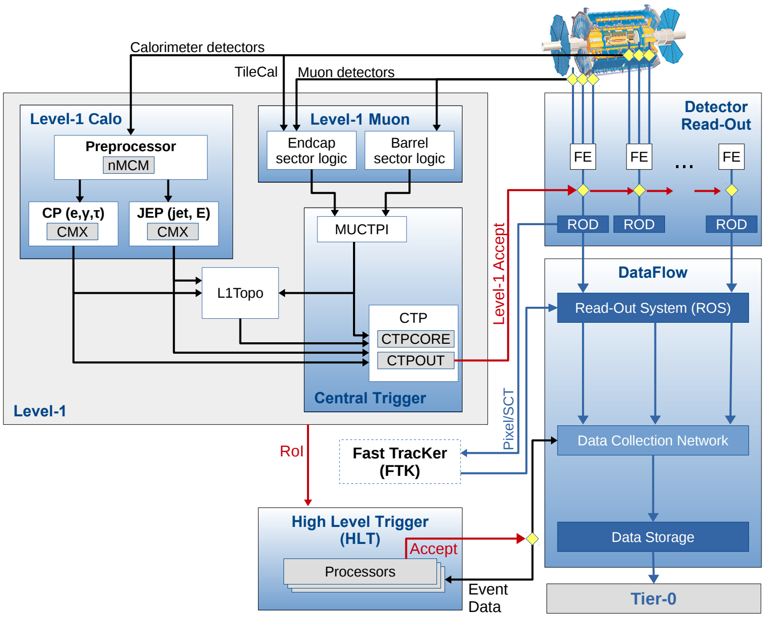

Calorimeter detectors
TileCal
Muon detectors
Level-1 Calo
Preprocessor
nMCM
CP (e,γ,τ)
CMX
JEP (jet, E)
CMX
Level-1 Muon
Endcap sector logic
Barrel sector logic
MUCTPI
L1Topo
CTP
CTPCORE
CTPOUT
Central Trigger
Level-1
Level-1 Accept
Detector Read-Out
FE
FE
...
FE
ROD
ROD
ROD
DataFlow
Read-Out System (ROS)
Data Collection Network
Data Storage
Tier-0
Pixel/SCT
Fast TracKer (FTK)
RoI
High Level Trigger (HLT)
Accept
Processors
Event Data

Figure 3.9: Conceptual structure of the ATLAS trigger system showing the two distinct levels, hardware based L1 and software based HLT triggers [85].

Chapter 4

Common aspects of the searches for top squark pair production

The purpose of the following chapter is to give an overview of the basic analysis strategy used in the searches reported in Chapter 7 and Chapter 8. Afterwards the recorded data samples and the simulated event samples utilised in the searches are described. The methods for the simulation of collision events will be briefly summarised as well.

In order to identify and reconstruct physics objects from the recorded signals, measured in the detector, particular algorithms are used. Dedicated procedures to obtain candidates for the major objects used in the analyses later on are described in the following.

4.1 Analysis strategy

The analysis strategies of the two searches for supersymmetry which are reported later on are similar. However, the discriminating variables used to isolate SUSY signal events from SM background events are different in the two analyses.

First, a so-called signal region (SR) is defined in which the expected number of SUSY signal events is significantly enhanced compared to the contribution from SM background processes in this region. After a coarse event preselection, a series of additional selection criteria is defined based on variables that emphasise the discriminative properties between the SUSY signal and the dominant background events. The SR is chosen such that the discovery sensitivity is maximised.

The contribution from the dominant SM background in the SR is constrained in a disjunct region of phase space, referred to as control region (CR), in which it is normalised to the observed number of events and further on extrapolated to the SR. This way, the analysis is more robust against mis-modelling effects of the simulation and background uncertainties can be reduced. Criteria for the definition of the CR are that it must be kinematically compatible with the signal region in order to yield a reasonable extrapolation, the associated background processes should be enhanced and a potential signal contribution must be negligible. To investigate the background estimation method in a data-driven manner, an additional validation region (VR) is defined. Hence, the SR targeting the particular three-body decay scenario is associated with a CR and VR for the estimation and validation of the dominant SM background processes.

A simultaneous fit (see Chapter 6) is performed in the signal region itself and the associated control region, while the validation region is only used to crosscheck the background prediction, but not to constrain the SM background events in the fit technique.

Two types of SRs are used in the analyses. First, a *discovery scenario* based on counting events in a single region is defined (“single-bin”) to quantify the existence and extent of a potential excess. Another analysis technique, referred to as *exclusion scenario* is used in the absence of any evidence for new physics beyond the SM, in which the SR is split into multiple bins in a specific discriminating variable. Afterwards, a maximum-likelihood fit is performed including all those bins. The reason for this procedure is that utilising the signal-to-background ratios in the different bins lead to an increased sensitivity for a potential signal exclusion.

4.2 Data collection and simulated event samples

4.2.1 Data sample

The two analyses which make up the central part of this work target the same signal model as described in Section 2.3.4. The recorded datasets on which the analyses are performed are distinct.

The analysis described in Chapter 7 will be referred to as `bWN-subSet` in the following. The abbreviation indicates the 3-body decay scenario of the top squark, comprising a bottom quark (b), a W boson and the neutralino (N), and in addition implies the dataset on which the analysis is performed. `bWN-subSet` is based on pp collision data at a centre-of-mass energy of $\sqrt{s} = 13$ TeV recorded in the years 2015 and 2016, corresponding to a total integrated luminosity of 36.1 fb^{-1}. The uncertainty of the integrated luminosity is 3.2 % and is determined based on the methods described in [86].

The subsequent analysis, documented in Chapter 8 and referred to as `bWN-fullSet`, is performed on the total recorded dataset of pp collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV from the period between 2015 and 2018 and corresponds to a total integrated luminosity of 139 fb^{-1} with an uncertainty of 1.7 %. The uncertainty on the integrated luminosity is derived from the calibration of the luminosity scale using x-y beam-separation scans, following the methodology described in [87], and using the LUCID-2 detector for the baseline luminosity measurements [88].

In both analyses, the dataset was recorded with one particular type of triggers requiring a large amount of missing transverse momentum, denoted as $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers. At the L1 trigger stage, the trigger algorithm is based on the vectorial sum of energy deposits in the calorimeters, while a more refined computation is done at the HLT, which is based on the vector sum of all calorimeter cells above a particular noise level. The $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger threshold at the HLT has been raised during the data-taking period from 70 GeV in 2015 to 110 GeV in 2018 in order to keep the trigger rate stable despite the increasing instantaneous luminosity. Nevertheless, it is fully efficient for an offline calibrated transverse momentum imbalance above 230 GeV, which is applied on all events later on.

4.2.2 Simulated event samples

Monte-Carlo event generation

In order to make reliable interpretations from the recorded data, comparisons to simulations are essential. Therefore, all SM processes which are expected from pp collisions at the LHC as well as possible BSM signatures that may be hidden in the vast majority of SM background events are generated using the so-called Monte-Carlo technique. With this numerical method, probabilities and their related quantities are estimated by using sequences of random numbers. Before discriminative quantities can be determined from the final states of scattering events in order to isolate SUSY signals from SM background processes, multiple consecutive steps are required during the event generation. The individual components of an event generation are the modelling of parton density functions inside the proton, the actual pp hard scattering process, the parton showering and hadronisation, and finally the detector response to the generated final state particles.

A comprehensive summary of current event generators for simulations of pp collisions at the LHC as well as an in-depth description of the generation of scattering processes may be found in [89].

It is a well known fact that the proton is a composite object and consists of partons, which is the collective term for particle constituents within hadrons, namely the gluons and quarks. In order to describe the momentum fractions of the partons within the proton, parton distribution functions (PDF) are used. A PDF describes the probability to find a respective parton with a momentum fraction x within a proton which is probed at a certain energy scale μ^2. Thus, PDFs are important for the simulation of the hard process and the parton showering. Parton distribution functions cannot be determined analytically because they depend on non-perturbative physics. However, they are extracted from measurements of deep inelastic scattering processes as well as from hadron colliders by using evolution equations [90–92]. There are a large variety of PDF sets provided by different groups. As an example, Figure 4.1 illustrates the parton distribution functions of the gluon, the valence quarks, and sea quarks of the proton at two different energy scales provided by the NNPDF group [93]. The calculations correspond to a precision at next-to-next-to-leading order (NNLO) of perturbation theory. As can be seen from these figures, PDFs depend on the energy scale and the momentum fraction x of the partons which are contributing to the total momentum. The choice of the PDF set also influences the cross section as well as the event shape.

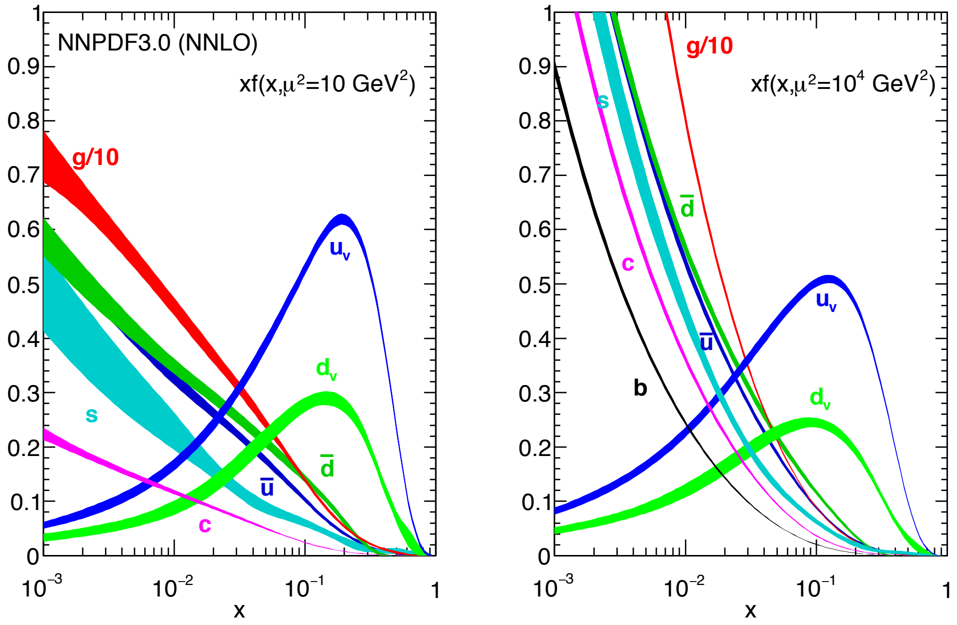


Figure 4.1: Parton distribution functions of the gluon, valence quarks and various sea quarks of the proton obtained in the NNLO NNPDF3.0 global analysis [93]. The left plot corresponds to hadronic scales ($\mu^2 = 10$ GeV^2) and the right plot corresponds to higher scales as present at the LHC ($\mu^2 = 10^4$ GeV^2) [5].

Due to the large momentum transfer, the partons rather than the proton participate in the hard process, which can be calculated at fixed order of perturbation theory. The cross section of a parton-parton collision with a particular final state can be computed using the factorisation theorem [94] and depends on the PDFs of the initiating partons, the squared parton-level matrix element and the integration over the corresponding phase space. Calculations beyond leading order (LO) increase the accuracy and tend to describe the experimental data better. For example, in order to determine the cross section of a particular final state at next-to-leading order (NLO) accuracy, extra corrections for additional real emissions and virtual quantum loops are taken into account in addition to the LO calculation. However, these emissions and loops may introduce divergences in the integrals and must cancel out, and as a result the combination afterwards with the parton shower algorithm becomes more complex.

Partons emerging from a collision carry colour-charge and thus may emit gluons, which in turn further give rise to quark-antiquark pairs. This process continues from the initial high energy scales down to low energy scales ($\mathcal{O}(\sim 1$ GeV$)$) at which non-perturbative effects become sizeable and the partons recombine to hadrons. Hence phenomenological models are required. In a simulation, this process is realised by the parton shower algorithm. Parton showers represent higher-order corrections to the hard process, such as collinear parton splitting and soft gluon emission. The parton splitting describes how the energy of the initial parton is distributed when it splits into two other partons by using dedicated splitting functions [92]. Additionally, a cut-off has to be introduced due to infinities which may occur in the integrals. Further on, the process of parton showering is applied to any new parton until the initial energy is distributed and reaches the hadronisation scale, where the partons form to hadrons. In principle, the same procedure is applied as a backwards evolution, at which the showering develops backwards and increases energy at each emission, in order to account for initial state radiation effects from incoming partons.

As already mentioned, the combination of the perturbatively computed hard-scatter process and the parton shower is complex. Matching techniques are required in order to differentiate between hard emissions from soft and collinear radiations, otherwise double counting of various effects would occur.

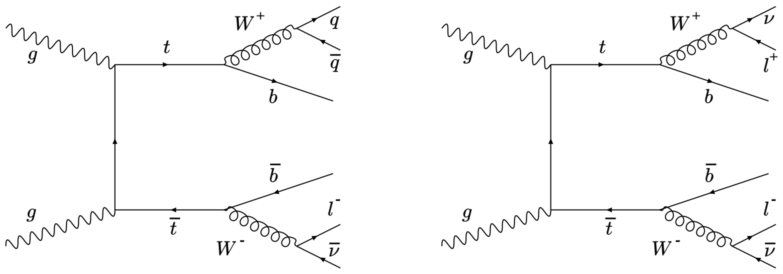
Finally, the hadronisation process, at which no further shower development of partons takes place, needs to be simulated. Hadronisation cannot be calculated from first principles because of the non-perturbative character at energy scales at the order of $\mathcal{O}(\sim 1$ GeV$)$, and thus phenomenological models are essential. The most commonly used phenomenological models are the string model [95, 96] and the cluster model [97].

The final result from these generation phases is a particle-level event, representing the true physics of a scattering process. Later on in this thesis, such events are referred to as *Truth* events. However, to provide reliable comparisons to the recorded data from the detector, interactions of the final-state particles from the generated events with the detector material are precisely simulated by the GEANT 4 simulation toolkit [98] and an accurate model of the ATLAS detector [99]. Moreover, the simulated detector response is provided in the same format as the recorded data. This kind of reconstructed events are denoted as *Reco* events later on.

Dominant SM backgrounds

The analyses performed in this thesis target final states containing exactly one electron or muon, which should be isolated from surrounding hadronic activity. Such prompt leptons are typically produced from the decay of a W boson. The other decay product from the leptonic W decay is the neutrino, which produces a significant amount of momentum imbalance in the detector, because it escapes the detector undetected. As a consequence, SM backgrounds populate the same phase space as potential supersymmetric decays.

The dominant SM background source in both analyses is top quark pair production ($t\bar{t}$). $t\bar{t}$ events, where the W boson from the top quark decay subsequently decays into a lepton-neutrino pair and the other W boson decays into a quark-antiquark pair are referred to as semi-leptonic $t\bar{t}$ events (Figure 4.2(a)). This type of $t\bar{t}$ processes can be suppressed by kinematic observables, due to the angular relation of the lepton-neutrino pair. The only remnants are therefore dileptonic $t\bar{t}$ events (Figure 4.2(b)), where one of the leptons is either not reconstructed, mis-identified as a different object or removed by the overlap removal procedure. Contributions from all-hadronic $t\bar{t}$ events as well as multi-jet production are found to be negligible in both analyses.

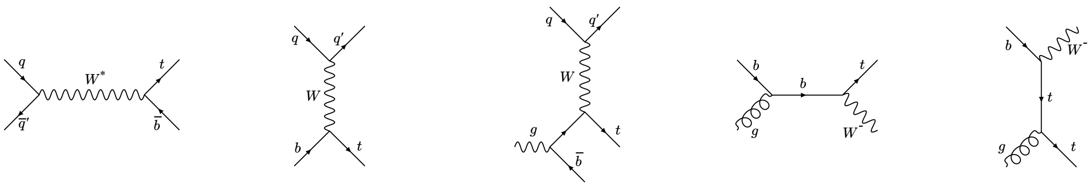


(a) Semi-leptonic top quark pair production

(b) Dileptonic top quark pair production

Figure 4.2: Feynman diagrams of top quark pair production and decay.

Another SM contribution comes from singly produced top quarks. Figure 4.3 shows Feynman diagrams of possible production channels of single top quark processes. From these scenarios, the largest contribution to the signal phase space is mainly from W boson associated production (Figure 4.3(d) and 4.3(e)), which is difficult to separate from a $t\bar{t}$ process because of the subsequent top quark decay into a b quark and a W boson. Contributions from single top quark production in s-channel (Figure 4.3(a)) and t-channel (Figure 4.3(b) and Figure 4.3(c)) are negligibly small.



(a) s-channel production

(b) t-channel production

(c) Dominant NLO t-channel production

(d) W boson associated production

(e) W boson associated production

Figure 4.3: Feynman diagrams for the production of single top quarks.

W boson production in association with additional jets, is also considered in the signal phase space if the W boson decays leptonically. This type of SM background is referred to as W+jets.

Other SM processes, such as di- or multi-boson production and top quark pair production in association with bosons, are also considered in the analyses but contribute only to a minor extent to the targeted phase space.

Signal

The phenomenology of the targeted signal model was already discussed in Section 2.3.4. In both analyses, a simplified model is considered for the scenario where the only light supersymmetric particles are the lightest top squark ($\tilde{t}_1$) and the lightest neutralino ($\tilde{\chi}_1^0$). The decay of the top squark is determined by the masses of the top squark and the neutralino, $\Delta m = m(\tilde{t}_1) - m(\tilde{\chi}_1^0)$. The main target model in this work is the 3-body decay mode of the top squark, $\tilde{t}_1 \rightarrow bW\tilde{\chi}_1^0$ (see Figure 2.4).

Simulated samples in `bWN-subSet`

Details of the simulated event samples used in analysis `bWN-subSet`, including the matrix element (ME) event generator, PDF set, the parton shower (PS) and hadronisation model, the set of tuned parameters (tune) for the underlying event (UE) and the order of the cross section calculation, are summarised in Table 4.1.

The nominal $t\bar{t}$ sample and the single top quark processes were generated with POWHEG-BOX (NLO) [100–104] interfaced to PYTHIA6 for parton showering and hadronisation. For both samples, cross sections were calculated to next-to-next-to-leading order with the resummation of soft gluon emission at next-to-next-to-leading-logarithm (NNLL) accuracy.

W+jets and diboson production were generated with SHERPA 2.2.0 [105] (and SHERPA 2.1.1 – 2.2.1 for the latter) using Comix [106] and OpenLoops [107]. The samples were merged with the SHERPA parton shower [108] using the ME+PS@NLO prescription [109]. Cross sections of the W+jets events were further normalised to NNLO accuracy.

Furthermore, $t\bar{t}+V$ processes were generated with MG5_aMC@NLO [110] and interfaced to PYTHIA8 for parton showering and hadronisation. The cross section of the $t\bar{t}+V$ samples are normalised to the theoretical predictions at NLO accuracy.

All SM background samples were simulated with the full simulation of the ATLAS detector based on GEANT4. More details of the $t\bar{t}$, W+jets, diboson and $t\bar{t}+V$ samples can be found in [111–114].

The SUSY signal samples were generated at leading order with MG5_aMC@NLO. The top squark decay was modelled with MadSpin [115] and interfaced with PYTHIA8 for the parton showering and hadronisation. In order to prove the sensitivity of the analysis in the adjacent 4-body phase space, a set of $\tilde{t}_1 \to bff'\tilde{\chi}_1^0$ samples using the exact same methodology were produced. The sensitivity in the 2-body regime was also tested. For additional $\tilde{t}_1 \to t\tilde{\chi}_1^0$ samples, the top squark decay was generated with PYTHIA8. Since the decay products of the 2-body samples do not preserve spin information, a polarisation reweighting was applied following the methodology described in [116, 117]. For the MadSpin samples, top squarks were assumed to be composed mainly of $\tilde{t}_\mathrm{R}$(∼70%), which is consistent with the samples in the 2-body decay mode. The cross sections for top squark pair production were calculated to next-to-leading order in the strong coupling constant, adding the resummation of soft gluon emission at next-to-leading-logarithm accuracy (NLO+NLL) [118–120]. The nominal cross section and the uncertainty were taken from an envelope of cross section predictions using different PDF sets and factorisation and renormalisation scales as described in [121]. The signal samples were all simulated with a fast detector simulation [122].

In addition, all samples which were produced with MG5_aMC@NLO and POWHEG-BOX used EVTGEN v1.2.0 [123] to model B-hadron decays. In order to simulate pileup effects from the same or nearby bunch crossings, a varying number of minimum-bias events, generated with PYTHIA8, are overlaid on the hard-scattering event for all samples. Further on, the number of interactions per bunch crossing in the simulated events was reweighted to match the pileup distribution measured in data.

Simulated samples in `bWN-fullSet`

Similar as above, Table 4.2 summarises the details of the simulated events used in analysis `bWN-fullSet`. For most of the samples only minor changes were implemented such as updates to the most recent releases of the various generator frameworks.

One major change for the $t\bar{t}$ sample and single top sample was to switch to PYTHIA8 for the modelling of the parton shower and hadronisation, while POWHEG-BOX has been

Process	ME generator	ME PDF	PS and hadronisation	UE tune	Cross section calculation
$t\bar{t}$	Powheg-Box v2 [100]	CT10 [124]	Pythia 6 [125]	P2012 [126]	NNLO+NNLL [127–132]
Single-top					
t-channel	Powheg-Box v1	CT104f	Pythia 6	P2012	NNLO+NNLL [133]
s and Wt-channel	Powheg-Box v2	CT10	Pythia 6	P2012	NNLO+NNLL [134, 135]
V+jets ($V = W/Z$)	Sherpa 2.2.0 [105]	NNPDF3.0 [136]	Sherpa	Default	NNLO [137]
Diboson	Sherpa 2.2.1	CT10	Sherpa	Default	NLO
$t\bar{t} + V$	MG5_aMC@NLO 2.2.2 [110]	NNPDF3.0	Pythia 8 [138]	A14 [139]	NLO [110]
SUSY signal	MG5_aMC@NLO 2.2 – 2.4	NNPDF2.3 [140]	Pythia 8	A14	NLO+NLL [121]

Table 4.1: Summary of the simulated samples used in analysis `bWN-subSet`.

kept for the generation of the matrix element. The cross sections for both samples were calculated at NNLO+NNLL accuracy as before.

For samples generated with Sherpa, the release version was changed. Thus, W+jets production was generated using Sherpa v2.2.1. Multi-boson events were generated with Sherpa 2.2.1 – 2.2.2. Parton showering and the merging scheme are identical to analysis `bWN-subSet`. The cross sections of W+jets processes were normalised with theoretical predictions at NNLO accuracy.

For $t\bar{t}+V$ samples, the same generator setup was used as above, only the release version of MG5_aMC@NLO was updated to 2.3.3.

The full simulation model of the ATLAS detector was used for the SM background processes.

For the generation and simulation of the signal samples, the exact same methodology was used as in analysis `bWN-subSet`. Only the release version of MG5_aMC@NLO was updated to 2.6.2. Again, the signal samples were all processed with the fast detector simulation. Signal cross sections were also updated and were calculated to NNLO in the strong coupling constant, adding the resummation of soft gluon emission at NNLL accuracy [118, 119, 141, 142]. The nominal cross section and its uncertainty were derived using the PDF4LHC15_mc PDF set, following the recommendations of [143].

B-hadron decays are modelled using EvtGen v1.6.0 for all samples which were produced with MG5_aMC@NLO and Powheg-Box. The effects of multiple pp interactions per bunch crossing are taken into account in the simulation, hence the number of interactions per bunch crossing was subsequently reweighted to the observed pileup distribution in data.

Process	ME generator	ME PDF	PS and hadronisation	UE tune	Cross section calculation
$t\bar{t}$	Powheg-Box v2	NNPDF3.0	Pythia 8	A14	NNLO+NNLL
Single top					
t-channel	Powheg-Box v1	NNPDF3.0	Pythia 8	A14	NNLO+NNLL
s and Wt-channel	Powheg-Box v2	NNPDF3.0	Pythia 8	A14	NNLO+NNLL
V+jets ($V = W/Z$)	Sherpa 2.2.1	NNPDF3.0	Sherpa	Default	NNLO
Multiboson	Sherpa 2.2.2 - 2.2.1	NNPDF3.0	Sherpa	Default	NLO
$t\bar{t} + V$	MG5_aMC@NLO 2.3.3	NNPDF2.3	Pythia 8	A14	NLO
SUSY signal	MG5_aMC@NLO 2.6.2	NNPDF2.3	Pythia 8	A14	NNLO+NNLL [118, 119, 141, 142]

Table 4.2: Summary of the simulated samples in analysis `bWN-fullSet`.

4.3 Object definition, event reconstruction and discriminating variables

In general, two definitions of physics objects are used. *Baseline* objects are used to compute the momentum imbalance in the transverse plane of an event and possible overlaps of different physics objects. Since the analyses in this work only select final states with exactly one electron or muon, baseline electrons or muons are necessary to veto events with more than one of those candidates. *Signal* physics objects are a sub-set of the baseline candidates that fulfil tighter identification and reconstruction criteria. The kinematic variables which are used in the analyses are computed from this type of objects.

4.3.1 Track and primary vertex reconstruction

Charged particle trajectories, referred to as tracks, are bent in the inner detector by a solenoidal magnetic field in order to determine the particle's momentum. The track reconstruction algorithm is based on fitting a set of measurements from the ID system to a trajectory model [144, 145]. Via an inside-out approach track seeds are determined by measurements in the pixel and SCT detectors. Further on, additional detector hits further apart from the interaction point are associated to the track using a combinatorial Kalman-filter algorithm [146, 147]. A χ^2-fit is performed repeatedly each time a hit is added to a track candidate. In addition, if for example more than one track shares the same hit, the tracks are resolved by an ambiguity-solving algorithm based on the track quality. Such hits are not taken into account during fit procedure of the track segment. Track candidates of good quality are propagated to the TRT system and measurements from this system are included in the reconstruction process.

The minimum requirement for a reconstructed track used in the following analyses is defined by $p_{\mathrm{T}} > 400$ MeV with at least six hits in the pixel and SCT system. Further on, track parameters are defined as the fraction of charge over momentum, $\frac{q}{p}$, the azimuthal and polar angles, ϕ and θ, and the transverse and longitudinal impact parameters, d_0 and z_0, which are the respective minimum distances to the centre of the detector. $\sigma(d_0)$ and $\sigma(z_0)$ are their respective uncertainties.

Afterwards, a collection of reconstructed tracks is associated to a vertex. The vertex reconstruction method is separated into two stages, first finding the vertex seed and second fitting the optimal vertex position [148,149]. At first, all tracks that satisfy the corresponding selection criteria are selected, and the vertex seed is initially determined based on the points of closest approach of the respective track to the reconstructed centre of the beam spot. Afterwards, an iterative fitting procedure is performed to find the optimal vertex position. An initial weight is associated to each track and the vertex position is recomputed using weighted tracks. This procedure is repeated and the track weights, which reflect the compatibility of the track with the estimated vertex position, are recalculated. After the final iteration, all tracks that are displaced from the estimated vertex position by more than seven standard deviations are removed from the vertex and thus can be used in the reconstruction of further vertices. The fitting procedure is then repeated until all tracks in the event are associated with a vertex or no further vertex is found in the remaining set of tracks.

The reconstructed vertex that has at least two associated tracks with $p_{\mathrm{T}} > 400$ MeV and with the largest $\sum_{\mathrm{tracks}} p_{\mathrm{T}}^2$ is defined as the primary vertex, and is assumed to be the location of the hard interaction process in analysis `bWN-subSet`. The threshold on the transverse momentum of the two associated tracks of the reconstructed vertex is raised to

$p_\mathrm{T} > 500$ MeV in analysis `bWN-fullSet`, while other requirements for the primary vertex definition remain identical.

4.3.2 Electrons

Electron candidates are formed from energy deposits in the electromagnetic calorimeter, which are matched to charged-particle tracks from the inner detector. A comprehensive summary of the reconstruction procedure is given in [150]. At first, energy clusters are built from energy deposits in the electromagnetic calorimeter via a sliding-window algorithm [151] with a window size of 3×5 in units of the calorimeter granularity $(\Delta\eta \times \Delta\phi)$. The transverse energy of a cluster is summed up in each layer of the EM calorimeter system and required to be above 2.5 GeV. Afterwards, tracks from the ID are associated to the cluster candidates. To account for bremsstrahlung effects, a subsequent fitting procedure is applied to tracks that match a cluster. Clusters with an associated track are re-formed using an extended window in the barrel region (3×7) and the endcap region (5×5), respectively. If more than one track matches a cluster, further criteria such as the distance between extrapolated tracks and cluster barycentres, track hits in the silicon detectors and the number of hits in the innermost layer of the silicon detectors are taken into account. If no track fulfils the matching criteria, the cluster is considered as a photon candidate. However, photons are not further considered in none of the analyses in this thesis.

In order to account for corrections due to the material in front of the EM calorimeter and the predefined cluster size, an energy calibration of the electrons is performed based on simulated events [152]. The electron four-momentum is determined from the energy measured in the calorimeter and the η and ϕ coordinates of the track.

Electron identification is based on a likelihood discriminant which contains various track-based as well as calorimeter-based measurements. A complete list of all the used quantities as well as a detailed description of the likelihood method is given in [150]. Based on well defined thresholds of the likelihood discriminant, four identification operating points are defined in order to meet the various electron requirements of different physics analyses. The four working points are denoted as VeryLoose, Loose, Medium and Tight and differ in the efficiency for identifying an electron. Additionally, a fifth working point will be used in the following analyses referred to as LooseAndBLayer, which is essentially identical to the Loose working point but requires at least one hit in the innermost layer of the pixel detector.

To differentiate the prompt production of an electron from a semi-leptonic decay of a heavy quark, hadrons mis-identified as leptons and photons or photon conversion into an electron-positron pair, an additional isolation requirement is needed. This means that no significant detector activity is measured within a cone around the electron candidate, defined by the quantity $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. Two different isolation criteria for signal electrons are defined in this work. First, only a track-based isolation criterion defined as the scalar sum of the transverse momenta of all tracks within a cone of variable size defined as $\Delta R \leq \min(0.2, 10\ \mathrm{GeV}/E_\mathrm{T})$ around the electron track and excluding the track associated to the electron is applied in analysis `bWN-subSet` [153]. Signal electrons are required to be isolated with a surrounding activity of less than 6% of the electron E_T. Since the efficiency measurements of the electron reconstruction algorithms have been continued during the data taking period, the isolation requirement for the latter analysis, `bWN-fullSet`, has been optimised. The requirement on the surrounding activity of the track-based isolation is changed to less than 15% of the electron E_T and an additional calorimeter-based isolation criterion is added. The calorimeter-based isolation requires an activity of less than 20% of the electron E_{T}. With the activity in the cone defined as the sum of transverse energies in calorimeter clusters within a fixed cone of $\Delta R = 0.2$ around the electron, after the contribution from the electron has been subtracted.

In addition, simulated events are corrected by E_{T}- and η- dependent efficiency scale factors. The electron reconstruction, identification and isolation efficiency is based on measurements in $Z \to ee$ and $J/\psi \to ee$ events.

In addition to the isolation requirement described above, the baseline and signal selection of electrons in the analyses `bWN-subSet` and `bWN-fullSet` are as follows. In `bWN-subSet`, baseline electrons satisfy the VeryLoose identification working point with an $E_{\mathrm{T}} > 5$ GeV and $|\eta| < 2.47$, while signal electrons must additionally satisfy the LooseAndBLayer identification working point, the longitudinal distance of the electron with respect to the primary vertex is required to be $|z_0 \sin\theta| < 0.5$ mm, the significance of the transverse impact parameter has to be $|d_0|/\sigma_{d_0} < 5$ and the requirement on the electron E_{T} is raised to be above 25 GeV.

Except for the isolation criterion mentioned above, the requirements in `bWN-fullSet` are similar. Only the requirement on the E_{T} of baseline electrons is lowered to 4.5 GeV and the working point for baseline electrons has been tightened to satisfy the LooseAndBLayer identification selection.

As a matter of clarity, a comprehensive summary of the electron definition of both analyses is listed in Table 4.3.

Selection	`bWN-subSet`	`bWN-fullSet`
	Baseline	
Identification	VeryLoose	LooseAndBLayer
E_{T}	> 5 GeV	> 4.5 GeV
$\|\eta\|$	< 2.47	< 2.47
	Signal	
Identification	LooseAndBLayer	LooseAndBLayer
E_{T}	25 GeV	25 GeV
Isolation	track-based only	calorimeter- and track-based
$\|d_0\|/\sigma_{d_0}$	< 5	< 5
$\|z_0 \sin\theta\|$	< 0.5 mm	< 0.5 mm

Table 4.3: Electron selection requirements for analysis `bWN-subSet` and `bWN-fullSet`, respectively.

4.3.3 Muons

Muon candidates are reconstructed independently in the inner detector and the muon spectrometer and finally combined to form a muon track. The reconstruction procedure in the ID is identical to any other charged-particle track described in Section 4.3.1. In the MS, segments are formed from hit patterns in each muon chamber. Afterwards, muon track candidates are reconstructed by fitting the segments from different layers of the MS together. In order to combine the information provided by the ID and the MS, the following four algorithms are used [154].

In the combined muon track reconstruction, muon tracks from the muon spectrometer are extrapolated inward and matched to an inner detector track. Hits from the MS may be removed or added to the fitted track depending on the fit quality.

Segment-tagged muons are reconstructed from a track in the ID which matches at least one track segment in the MDT or CSC chambers. This may be helpful if the muon candidate passes only one layer of the muon spectrometer because of its low p_{T}, or if it is measured in a region with reduced acceptance.

Calorimeter-tagged muon candidates consist of an inner detector track which is matched to an energy deposit in the calorimeter that is compatible with a minimum-ionising particle. Calorimeter-tagged muons have lower purity compared to other muon reconstruction algorithms but may recover acceptance in some regions of the muon spectrometer.

The last type are extrapolated muons, which only use information from the muon spectrometer and a requirement on compatibility with originating from the collision point. Measurements from at least two layers in the MS are required to reconstruct the muon track. In the forward region, the muon must traverse at least three layers. With this algorithm the muon acceptance can be extended to the region $2.5 < |\eta| < 2.7$, which is not covered by the inner detector.

If an ID track is matched to more than one type of muons described above, preference is given to combined muons first, then segment-tagged muons and calorimeter-tagged muons. Possible overlap with extrapolated muons is also resolved by criteria such as the fit quality or the number of hits.

Muon identification requirements are based on the number of hits in the ID and the MS layers and the compatibility between the measurements in the ID and MS. Five muon identification selections are defined in the ATLAS collaboration to meet the needs of various physics analyses. They are denoted Loose, Medium, Tight, Low-p_{T} and High-p_{T}. Only the Loose and Medium identification working points are used in the analyses later on. The Loose selection criterion maximises the reconstruction efficiency at good muon track quality and uses all aforementioned muon types, while the Medium selection only uses combined muon types and extrapolated muon types.

Signal muon isolation requirements are similar to those used for electrons based on the activity within a cone around the object. In `bWN-subSet`, the muon isolation is also only track-based and the definition is the same as described in Section 4.3.2, except the requirement on the variable cone size is changed to $\Delta R < \min(0.3, 10\ \mathrm{GeV}/p_{\mathrm{T}})$. Following continued efficiency measurements during the data taking period, the muon isolation selection in `bWN-fullSet` has been slightly changed. Similar to the electrons, the requirement on the activity of the track-based isolation is changed to less be than 15% of the muon p_{T} and an additional calorimeter-based isolation criterion is added. The calorimeter-based isolation requires an activity within a cone with fixed size of $\Delta R = 0.2$ to be less than 30% of the muon p_{T}. The activity in the cone is defined as the sum of transverse energies in calorimeter clusters around the muon candidate, after excluding the contribution from the muon.

Similarly as for electrons, the muon reconstruction, identification and isolation efficiencies are based on measurements in $Z \to ee$ and $J/\psi \to ee$ events and simulated events are corrected by corresponding p_{T}- and η- dependent scale factors.

The baseline and signal selection for muons in `bWN-subSet` and `bWN-fullSet` are as follows. In `bWN-subSet`, baseline muons are selected by the VeryLoose identification working point with $p_{\mathrm{T}} > 4$ GeV and $|\eta| < 2.7$. Signal muons additionally satisfy the Medium identification working point and the impact parameters must satisfy $|z_0 \sin\theta| < 0.5$ mm and $|d_0|/\sigma_{d_0} < 3$. The signal muon p_{T} is required to be above 25 GeV. Additionally, the already mentioned isolation requirements are applied for signal muons as well.

The requirements in `bWN-fullSet` are identical, except the working point for baseline muons must satisfy the Medium identification working point. The muon isolation condition is changed to a calorimeter- and track-based method as already described.

A summary of the muon selection criteria of both analyses is listed in Table 4.4.

Selection	`bWN-subSet`	`bWN-fullSet`
Baseline		
Identification	Loose	Medium
p_{T}	> 4 GeV	> 4 GeV
$\|\eta\|$	< 2.7	< 2.7
Signal		
Identification	Medium	Medium
p_{T}	27 GeV	25 GeV
Isolation	track-based only	calorimeter- and track-based
$\|d_0\|/\sigma_{d_0}$	< 3	< 3
$\|z_0 \sin\theta\|$	< 0.5 mm	< 0.5 mm

Table 4.4: Muon selection requirements for analysis `bWN-subSet` and `bWN-fullSet`, respectively.

4.3.4 Jets

After a hard scattering process, a parton usually hadronises to a collimated shower of hadrons, denoted as a jet. The jet reconstruction procedure initiates in the calorimeters by a topological cluster formation [155]. First, cluster seeds are localised in the calorimeter as single calorimeter cells in which the ratio of the energy deposit over the average noise in the cell is above a certain threshold ($|\frac{S}{N}| = |\frac{E_{\mathrm{cell}}}{\sigma_{\mathrm{noise,cell}}}| > 4$). Afterwards, neighbouring cells from the same calorimeter layer or cells which are neighbours of the seed cell in the adjacent calorimeter layer are added to the seed cell if they exceed again a certain energy threshold. The cluster growth is controlled by the ratio $|\frac{S}{N}| > 2$. Finally, further cells are added to the cluster if the energy deposit exceeds $|\frac{S}{N}| > 0$. Depending on the respective energy deposits, cluster seeds may be merged into a single cluster, or clusters may be separated if multiple local energy maxima occur. At this reconstruction stage, the cell energy is measured at electromagnetic scale, which means it relates to the activity of an electromagnetic shower. Corrections for hadronic showers are taken into account later on.

The topological calorimeter clusters are then combined to build jet candidates using a sequential recombination jet algorithm referred to as the anti-k_t jet clustering algorithm [156]. A jet radius parameter of $R = 0.4$ is used in the clustering method. In addition, ID tracks are matched to the jet candidate via the so-called ghost-association [157] during the clustering procedure. In this method, the track momenta are set to an infinitesimal amount, essentially only the direction of the track is retained. Further on, this matching procedure utilises also jet areas [158], a data-driven event-by-event, jet-by-jet method, in order to determine accurate jet energies, which are strongly affected by pileup.

After the jet clustering, jet candidates must be calibrated due to contributions from additional soft pp collisions and to account for residual detector response effects. The jet energy scale calibration consists of multiple consecutive stages of corrections based on simulated events as well as in-situ techniques [159]. At first, the origin of the jet candidate is corrected to point to the direction of the primary vertex rather than the centre of the detector.

At the next stage, corrections accounting for pileup contributions are applied. These corrections are separated into two components, an area-based p_{T}-density subtraction [158], applied event-by-event, and a residual pileup correction depending on the average number of concurrent interactions and the number of primary vertices, derived from simulated events.

Afterwards, the jet four momenta are corrected to match that of particle-level jets from simulations. A particle-level jet in that sense means a jet candidate from a generated particle-level event. Both jet energy and direction are calibrated in this step. This calibration stage is essential in order to incorporate transitions between different calorimeter technologies and prompt changes in the calorimeter granularity.

Whether a jet is initiated by a quark or gluon strongly affects the particle composition as well as the shower shape of the jet itself. This also has a consequence for the detector response and jet reconstruction. Thus, a further calibration procedure referred to as the global sequential calibration is applied. Once more, the jet four momenta are corrected as a function of the particle-level jet p_{T} and their direction in the detector.

Finally, a residual in-situ calibration is applied on the recorded data sample in contrast to the simulated events, using well-measured reference objects such as photons, Z bosons, and calibrated jets. This correction is necessary due to the imperfect description of the detector response and material in the simulation, electromagnetic and hadronic interaction with the detector, but also the simulation of the hard scattering process, pileup and jet formation.

Due to pileup interactions, extracting precise jet information is challenging. In order to identify jets which are not associated with the primary vertex, a dedicated technique denoted as the jet vertex tagger (JVT) [160] is applied to the jet candidates. The JVT technique is a likelihood discriminant based on quantities describing the fraction of the scalar sum of the track p_{T} corresponding to the jet as well as associated to the primary vertex.

The following selections are applied in the analyses. For analysis `bWN-subSet` the thresholds for baseline jets are $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 4.7$, while for signal jets, these requirements are tightened to $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$. Additionally, signal jets with a transverse momentum less than 60 GeV and $|\eta| < 2.5$ must also satisfy track-based criteria in order to exclude jets which may be associated to pileup, by requiring a value of the JVT discriminant above 0.59. The selections are slightly different for analysis `bWN-fullSet`. In this analysis the jet baseline criteria are $p_{\mathrm{T}} > 20$ GeV, whereas the requirements for signal jets are $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$. In addition, signal jets with transverse momentum below 120 GeV have to satisfy a JVT threshold of 0.59 in order to remove jets initiated from pileup. Furthermore, after the overlap removal procedure (see Section 4.3.8) is performed, jets also need to pass loose jet cleaning requirements, otherwise the whole event is removed from the selection.

4.3.5 Flavour tagging

Jets which are initiated by a b quark can be distinguished from other jets which could be initiated by gluons, light quarks or even a charm quark or a hadronically decaying τ lepton. The reason is because jets originating from b quarks contain B-hadrons, which have a significant lifetime. Due to this fact, they decay apart from the primary vertex and initiate a so-called secondary vertex.

The main b-tagging algorithm combines the outcome of three independent tagging algorithms into a Boosted Decision Tree (BDT) classifier which is trained to distinguish b-tagged jets from c-tagged jets and light-flavour jets. The algorithm is referred to as the

MV2 algorithm [161,162]. One of the three independent tagging techniques relies on measurements of impact parameters of tracks which are associated to a jet. Thus, information of the track position relative to the primary vertex and the jet are extracted. Transverse impact parameters as well as longitudinal impact parameters are finally combined into a likelihood-based discriminant. Another likelihood-based tagging algorithm aims to reconstruct secondary vertices by building two-track vertices from all tracks which are matched to a jet. The third tagging method is referred to as the JetFitter algorithm which aims to reconstruct the complete B-hadron decay chain inside the jet, by taking also subsequent decays of B-hadrons and C-hadrons into account.

In order to tag a jet as initiated by a b-quark, the output of the MV2 discriminant must be above a certain threshold. In analysis `bWN-subSet`, the b-tagging algorithm is applied to jets that satisfy the following criteria. Jets passing the same JVT requirements as described in Section 4.3.4 and satisfying $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$, are tagged as a b-jet if the MV2c10 discriminant is above 0.6459. This working point corresponds to a b-tagging efficiency of $\epsilon_b = 76.97$ %. Moreover, rejection rates for light-flavour jets, c-tagged jets as well as τ-jets have been estimated and are found to be 134.34, 6.21 and 22.04, respectively [163]. In analysis `bWN-fullSet`, the b-tagging algorithm is applied to jets that also satisfy the dedicated JVT requirements as described in Section 4.3.4 and satisfying $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$. The threshold on the MV2c10 discriminant is above 0.64, which corresponds to a b-tagging efficiency of $\epsilon_b = 77.53$ %.

Similar as for other physics objects, p_{T}-dependent scale factors to correct for discrepancies between recorded data and simulations in tagging efficiencies are applied to simulated events.

A summary of the jet selection criteria, including the selection of b-tagged jets, for both analyses is listed in Table 4.5.

Selection	`bWN-subSet`	`bWN-fullSet`
	Baseline	
p_{T}	> 20 GeV	> 20 GeV
$\|\eta\|$	< 4.7	–
	Signal	
p_{T}	> 25 GeV	> 25 GeV
$\|\eta\|$	< 2.5	< 2.5
JVT	> 0.59 for $\|\eta\| < 2.4$ and $p_{\mathrm{T}} < 60$ GeV	> 0.59 or $p_{\mathrm{T}} > 120$ GeV
	b-jets	
p_{T}	> 20 GeV	
$\|\eta\|$	< 2.5	
JVT	> 0.59 for $\|\eta\| < 2.4$ and $p_{\mathrm{T}} < 60$ GeV	> 0.59 or $p_{\mathrm{T}} > 120$ GeV
MV2c10	> 0.6459 ($\epsilon_b = 76.97$ %)	> 0.64 ($\epsilon_b = 77.53$ %)

Table 4.5: Jet selection requirements, including requirements for b-tagged jets.

4.3.6 Hadronically decaying τ leptons

A peculiar feature of the τ lepton is that it decays via an electroweak process into a τ neutrino and leptons of the first and second generation (leptonic τ decay) or even in about two out of three τ decays into a τ neutrino and charged hadrons (hadronic τ decay, τ_{had}).

The selections of both analyses require exactly one prompt electron or muon from the decay of a real or a virtual W boson. Hadronically decaying τ leptons are not explicitly included in the event selection. However, in order to suppress background processes with τ_{had} candidates in the final state, a kinematic variable is defined based on τ_{had} see Section 4.3.9).

Detailed information about the reconstruction of hadronic τ leptons may be found in [164, 165]. Hadronic τ decay modes are distinguished by final states with either one charged hadron (one prong), or three charged hadrons (three prong). All other decay modes are of minor interest. τ_{had} candidates are reconstructed from jet seeds with $p_{\text{T}} >$ 10 GeV and $|\eta| < 2.5$. The transverse energy of the τ_{had} candidate is determined from jet constituents in the calorimeter within a cone of $\Delta R < 0.2$, and further on, tracks from the ID with $p_{\text{T}} > 1$ GeV and within a cone of $\Delta R < 0.2$ around the seed candidate are associated to the hadronic τ candidate. Hadronic τ identification is performed with a boosted decision tree classifier, which takes sensitive variables to the τ_{had} decay topology into account. Exactly one or three associated tracks with a total electric charge opposite to the selected electron or muon are required for hadronic τ candidates, and τ_{had} must satisfy $p_{\text{T}} > 20$ GeV and $|\eta| < 2.5$.

4.3.7 Missing transverse momentum

Momentum conservation implies that the total sum of transverse momenta should be zero in a pp collision.

Neutral, weakly interacting particles, such as the neutrino or the neutralino in the hypothetical supersymmetric model discussed in Section 2.3.4, do rarely interact with any detector material and hence escape the detector undetected. As a consequence, a momentum imbalance in the transverse plane originates in an event, referred to as the missing transverse momentum $\vec{E}_{\text{T}}^{\text{miss}}$. Its magnitude is denoted as $E_{\text{T}}^{\text{miss}}$. $\vec{E}_{\text{T}}^{\text{miss}}$ is defined as the negative vectorial sum of all reconstructed objects in an event. Moreover, the reconstruction of $\vec{E}_{\text{T}}^{\text{miss}}$ contains a *hard* component from calibrated objects and a *soft* component of charged-particle tracks from the ID which are associated with the primary vertex but not related to any reconstructed object [166, 167]:

$$\vec{E}_{\text{T}}^{\text{miss}} = -\sum_{\text{electrons}} \vec{p}_{\text{T}}^{\,e} - \sum_{\text{muons}} \vec{p}_{\text{T}}^{\,\mu} - \sum_{\text{jets}} \vec{p}_{\text{T}}^{\,\text{jet}} - \sum_{\text{unused tracks}} \vec{p}_{\text{T}}^{\,\text{track}} \tag{4.1}$$

In the following analyses, $E_{\text{T}}^{\text{miss}}$ is computed using baseline electrons, muons, and jets before the removal procedure of overlapping objects (see Section 4.3.8), although overlap between reconstructed objects is taken into account in the computation. For the soft term, $E_{\text{T}}^{\text{miss,soft}}$, badly measured tracks are removed. Photons and hadronically decaying τ leptons are not explicitly included in the computation of $\vec{E}_{\text{T}}^{\text{miss}}$, but enter either as jets, electrons, or via the soft term.

4.3.8 Overlap removal

It may happen that detector signatures are associated to multiple reconstructed objects. For example, calorimeter deposits may be related to electrons or jets, or charged-particle tracks may be associated to a muon or jet candidate. In order to avoid this kind of double-counting, various conditions are defined and reconstructed objects may be removed in favour of other objects. Only reconstructed objects which pass the overlap removal procedure are used later on for event selections and computations of kinematic observables.

Baseline objects are used in all the following requirements. The criteria in the overlap removal procedure are optimised with respect to the acceptance of prompt leptons to achieve good signal acceptance. The different criteria of the overlap removal procedure of `bWN-subSet` and `bWN-fullSet` are described in the following.

Overlap removal in analysis `bWN-subSet`

1. Electron/muon: If an electron and a muon overlap within $\Delta R < 0.01$, the muon is removed if it is a calorimeter-tagged muon, otherwise the electron is removed.

2. Electron/jet: If an electron and a jet, which is not b-tagged, overlap within $\Delta R < 0.2$, the jet is removed and the electron is kept.

3. Muon/jet: If a muon and a non b-tagged jet overlap within $\Delta R < 0.4$, the object is interpreted as a muon, if the jet has less than three tracks with $p_{\mathrm{T}} > 500$ MeV, or the p_{T} ratio of the muon and the jet satisfies $p_{\mathrm{T}}^{\mathrm{muon}}/p_{\mathrm{T}}^{\mathrm{jet}} > 0.7$. Otherwise both objects are kept.

4. Jet/lepton: If a jet, which satisfies the previously mentioned overlapping conditions, is close to an electron or muon within $\Delta R < \min(0.04 + 10\ \mathrm{GeV}/p_{\mathrm{T}}^{\ell}, 0.4)$, the jet will be kept in favour of the lepton.

5. Electron/τ: If an electron satisfying the previous conditions overlaps with a τ within $\Delta R = 0.1$, the τ is removed.

Overlap removal in analysis `bWN-fullSet`

Despite some similarities between the overlap removal criteria of the two analyses, a few changes have been adapted to the overlap removal procedure in analysis `bWN-fullSet`. One difference with respect to the former analysis is that the electron/muon overlap requirement is omitted. Another update concerns the muon/jet overlap criterion. In the former analysis, the muon candidates are retained in favour of non-b-tagged jet candidates, while in analysis `bWN-fullSet` the b-tag requirement is omitted and muon candidates can remove all jet candidates which are close to the muon.

1. Electron/jet: If a baseline electron and a baseline jet are found within $\Delta R < 0.2$ and the jet is not b-tagged, the electron is kept and the overlapping jet is discarded. Whether or not the jet is b-tagged is only checked up to jet $p_{\mathrm{T}} = 200$ GeV.

2. Muon/jet: If a baseline muon can be ghost-matched [157] to a baseline jet within $\Delta R < 0.2$, the muon is retained if the jet has less than 3 tracks with $p_{\mathrm{T}} > 500$ MeV. Otherwise the jet is kept.

3. Jet/lepton: If a jet, which satisfies the previously mentioned overlapping conditions, is close to an electron or muon within $\Delta R < \min(0.04 + 10\ \mathrm{GeV}/p_{\mathrm{T}}^{\ell}, 0.4)$, the jet will be kept in favour of the lepton.

4. Electron/τ: If an electron satisfying the previous conditions overlaps with a τ within $\Delta R = 0.1$, the τ is removed.

4.3.9 Discriminating variables

Azimuthal angle between the lepton and the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ directions

Selecting final states with exactly one lepton leads to a sample whose composition of SM processes is dominated by semi-leptonic $t\bar{t}$ events and W+jets processes. Further reduction of these two background sources can be achieved by selecting only events with moderate values of the azimuthal angle between the directions of the lepton ℓ and $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$, referred to as $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$. The reason is that for those two SM processes the momentum imbalance in the transverse plane is caused by the neutrino from the leptonic decay of the W boson, and thus has an angular relation to the lepton. This is in contrast to the SUSY signal, where the neutralinos also contribute to the momentum imbalance. As a result, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is arbitrarily distributed in the transverse plane and not correlated with the direction of the lepton.

Transverse mass of the W boson

Another observable which suppresses semi-leptonic $t\bar{t}$ events as well as W+jets events is the transverse mass m_{T}, which aims to reconstruct the mass of the W boson and is defined by:

$$m_{\mathrm{T}} = \sqrt{2 p_{\mathrm{T}}^{\ell} E_{\mathrm{T}}^{\mathrm{miss}} \left[1 - \cos\left(\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})\right)\right]} \tag{4.2}$$

In this equation, p_{T}^{ℓ} is the transverse momentum of the lepton and $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ is the azimuthal angle between lepton and $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ directions as already described. In this definition, the lepton mass is assumed to be negligible. As can be seen in the formula, $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ is correlated with m_{T}.

After applying either one of these two variables to the event selection, the remaining dominant SM background source are dileptonic $t\bar{t}$ events, where one lepton is either outside the detector acceptance, mis-identified as a jet, or a hadronically decaying τ lepton. The reason is that in all the aforementioned cases the decay products from the top quark pair include at least two neutrinos, resulting in an arbitrary value for $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ and large values for m_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$.

Invariant mass of the lepton and b-tagged jets

The invariant mass of the lepton and a b-tagged jet, referred to as m_{bl}, was also found to have discriminating power. For the reconstruction of this variable the leading b-tagged jet is used unless the event has a second b-tagged jet. In that case the minimum of the invariant mass computed with the first or the second b-tagged jet is used ($m_{\mathrm{bl}} = \min(m_{\mathrm{b_1l}}, m_{\mathrm{b_2l}})$).

Due to the compressed mass-splitting between the mass of the top squark and the neutralino in a potential top squark decay, the kinematics of the b-tagged jet and the lepton which emerge from the top squark decay tend to be softer, leading to small values of m_{bl}. This is in contrast to SM background events. For these processes values of m_{bl} are typically larger.

Stransverse mass

The stransverse mass m_{T2} [168] is not used as a requirement to any of the selections in the following analyses. However, variants of this observable are used, and thus its definition will be outlined briefly.

Assuming a decay scenario where a pair of particle and anti-particle decay in two separate branches, referred to as branch a and branch b. There may be well reconstructed particles as well as particles with unknown momenta (e.g. a neutrino which escapes the detector) in the two branches. The sum of the four-vectors of visible particles per branch is given by $p_i = (E_i, \vec{p}_{\mathrm{T}i}, p_{zi})$, while the sum of the four-vectors of invisible particles per branch is given by $q_i = (F_i, \vec{q}_{\mathrm{T}i}, q_{zi})$. The index i denotes the branch $i \in \{a, b\}$. In this decay topology, the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is splitted in two parts, each associated to a branch:

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = \vec{q}_{\mathrm{T}a} + \vec{q}_{\mathrm{T}b}. \tag{4.3}$$

The transverse mass m_{T} of each branch is defined by:

$$m_{\mathrm{T}i}^2 = \left(\sqrt{p_{\mathrm{T}i}^2 + m_{p_i}^2} + \sqrt{q_{\mathrm{T}i}^2 + m_{q_i}^2}\right)^2 - \left(\vec{p}_{\mathrm{T}i} + \vec{q}_{\mathrm{T}i}\right)^2, \tag{4.4}$$

with $m_{p_i}^2 = E_i^2 - \vec{p}_i^2$ and $m_{q_i}^2 = F_i^2 - \vec{q}_i^2$ being the invariant masses of the respective branch. m_{T2} is defined as the configuration where the corresponding transverse mass of the respective branch, $m_{\mathrm{T}a}$ or $m_{\mathrm{T}b}$ becomes smallest:

$$m_{\mathrm{T2}} \equiv \min_{\vec{q}_{\mathrm{T}a} + \vec{q}_{\mathrm{T}b} = \vec{E}_{\mathrm{T}}^{\mathrm{miss}}} \left\{ \max(m_{\mathrm{T}a}, m_{\mathrm{T}b}) \right\}. \tag{4.5}$$

The masses of the invisible particles in each branch m_{q_a} and m_{q_b} have to be predefined and the result of m_{T2} gives the minimum parent mass consistent with one of the branches. The following variants of m_{T2} only differ in the measured particles and the assumed undetected particles, as well as choices for the predefined masses of m_{q_a} and m_{q_b}.

Asymmetric stransverse mass

The first modification of m_{T2} described here is an asymmetric variant, referred to as am_{T2} [169–172]. am_{T2} aims to estimate the top quark mass in dileptonic $t\bar{t}$ events, where one lepton is not reconstructed. Since the mass difference between the top squark and the neutralino of the aforementioned supersymmetric signal is supposed to be smaller than the top quark mass, am_{T2} can be powerful to discriminate signal from the dileptonic $t\bar{t}$ background. Considering a dileptonic $t\bar{t}$ event as illustrated in Figure 4.2(b), one branch contains a well reconstructed b-tagged jet and the other branch contains a b-tagged jet and the lepton, which are also measured directly (Figure 4.4). Invisible particles are the neutrino in one branch and the neutrino as well as a mis-identified or not reconstructed lepton from the leptonically decaying W boson in the other branch. The calculation of am_{T2} is identical as above, but using the predefined mass assumptions for the undetected particles in the separate branches $m_{q_a} = m_W = 80$ GeV and $m_{q_b} = m_\nu = 0$ GeV. As a consequence of these predefined mass assumptions, am_{T2} has a lower bound at 80 GeV.

$\tau-$based stransverse mass

Another variant of m_{T2} is m_{T2}^{τ}. This variable targets $t\bar{t}$ events where one of the W bosons decays into a hadronic τ lepton. As indicated in Figure 4.5, the $t\bar{t}$ decay is again considered as two branches with visible and invisible particles. This time the visible particles in one branch are a b-tagged jet and a τ_{had}, and a b-tagged jet and the charged lepton in the other branch. The invisible particles are two neutrinos in one branch which are associated with the hadronic τ decay and a neutrino corresponding to the charged lepton in the other branch. Masses of the invisible particles in the two branches are assumed as $m_{q_a} = 0$ GeV

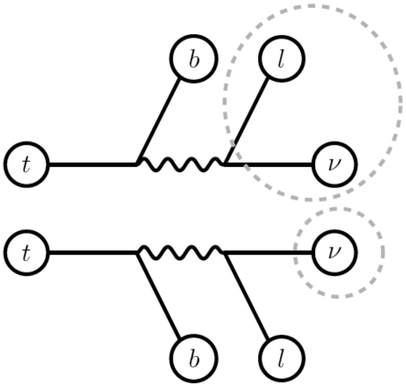


Figure 4.4: Illustration of the construction of am_{T2}, in order to discriminate dileptonic $t\bar{t}$ events with one lost lepton [173].

and $m_{q_b} = m_\nu = 0$ GeV. This variable is not used to isolate potential signal events from SM background events, but a requirement on m_{T2}^{τ} is applied at event preselection (see Chapter 7 and Chapter 8) to veto events containing hadronically decaying τ leptons.

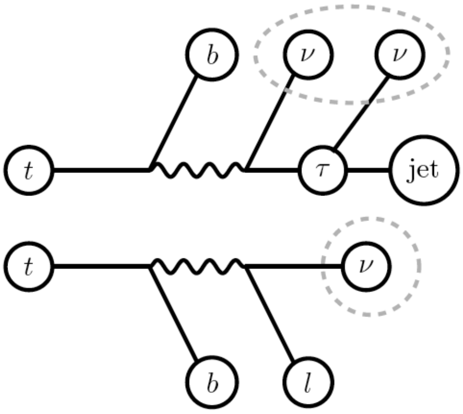
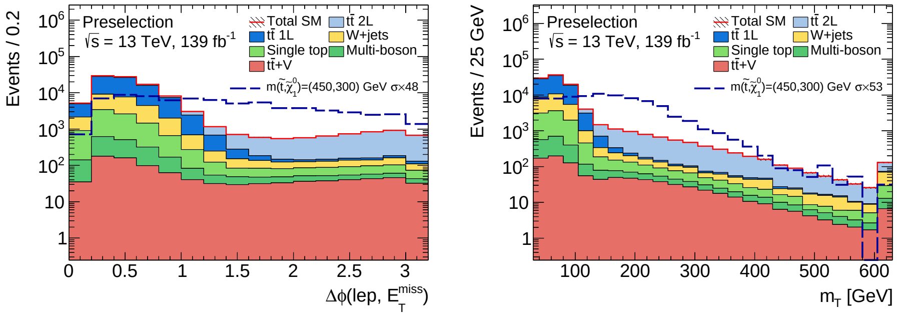
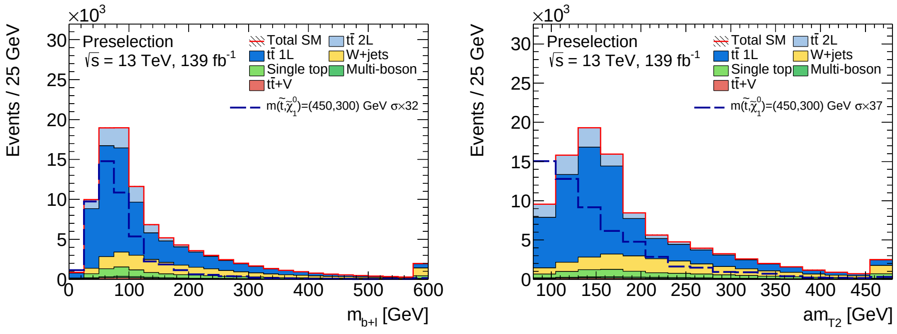


Figure 4.5: Illustration of the construction of m_{T2}^{τ}, used to discriminate dileptonic $t\bar{t}$ events with a hadronically decaying τ lepton [173].

Figure 4.6 shows the distributions of the discriminating variables from simulated SM background and SUSY signal events. The object definitions as described above for the `bWN-fullSet` analysis as used. The event preselection from analysis `bWN-fullSet` as described in Table 8.1 is utilised except the threshold on m_{T} is loosened to $m_{\mathrm{T}} \geq 30$ GeV. As indicated in the legend, the cross section of the SUSY signal is magnified in order to stress the discriminating character of the respective kinematic variable.



(a) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ (b) m_{T}



(c) m_{bl} (d) am_{T2}

Figure 4.6: Distributions of the discriminating variables after the preselection as described in Table 8.1 except for the threshold on m_{T} which is loosened to $m_{\mathrm{T}} \geq 30$ GeV. The object definitions from the `bWN-fullSet` analysis are used. The cross section of the SUSY signal is magnified in order to stress the discriminating power of the variable.

Chapter 5

Fundamentals of machine learning

This section is dedicated to the general concepts and principles of machine learning (ML). The basics of a supervised learning task will be outlined and model classification will be described. Key aspects of a neural network (NN) and a recurrent neural network (RNN) are presented, model parameters and parameter optimisation, which is essential for the general performance of a learning algorithm, will be discussed. The training procedure, at which a machine or algorithm learns attributes from a given dataset, will be explained in detail further on.

5.1 Supervised learning

The essential concept of machine learning algorithms is that they are able to derive estimations from multi-variate data itself, thus these algorithms "learn" from certain characteristics or discrepancies which are present in the data. In order to derive conclusive predictions from data, the model needs to be trained. The dataset is therefore split into a dedicated *training set* and *test set*. As the name implies, the training set contains events, also denoted as examples, on which the model is trained, while the events in the test set are used to assess the model's performance and eventually tune the algorithm. Typically, there exists a third set which is completely independent from the training and test set, referred to as the *validation set*. The validation set contains a collection of unseen events, from which the final model derives predictions. The ability to make successful predictions for events which are not included in the training set is called generalisation. In a supervised learning task, an algorithm attempts to model a function that relates input features to output features. In contrast to other tasks in machine learning, in a supervised learning algorithm the training set is labelled, which means the outputs in the training set are well defined and the learning is based on the relation of known input-output pairs.

Other domains in machine learning are unsupervised learning and reinforcement learning. Other than in supervised learning, in an unsupervised learning problem the dataset does not contain information of the output and thus the algorithm attempts to find patterns or clusters in unlabelled data. The goal of reinforcement learning is that a machine learns to evolve in an environment by trial and error. Both, unsupervised and reinforcement learning are not of particular importance for this work and will not be discussed in detail here.

5.1.1 Classification

The data set used to train the model contains m training events of the form $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}\}$. $\mathbf{x}^{(i)}$ denotes the feature vector or input vector of the i-th event. In this work, features are usually kinematic properties, such as the p_{T} of particular physics objects, $E_{\mathrm{T}}^{\mathrm{miss}}$, m_{T}, *etc.* The features are composed into a vector, whereas the size of the vector may depend on the corresponding purpose. The events in the dataset are associated to a set of outputs or targets, $\{y^{(1)}, \ldots, y^{(m)}\}$. Depending on the kind of task the algorithm is supposed to solve, the targets may be different. In a regression task, for example, the outputs are continuous, while in a classification task the targets are discrete and the algorithm is specified to differentiate to which class a certain event belongs to. Hence, the algorithm models a function of the form $f : \mathbb{R}^n \to \{0, \dots, k\}$, where n represents the number of input variables used to train the algorithm and k is the number of different classes or categories.

In the task of distinguishing between SM background processes and supersymmetric signal events, the purpose is to classify a set of input variables into two discrete classes, namely background and signal. Thus, this type of problem is a classification problem.

5.2 Neural networks

Neural networks (NN) are already known for a long time. Their origin dates back to the middle of the previous century, while they gained a lot of attention just in the recent decades as the computing era evolved and advanced computing resources become widely available. Neural networks have become an excellent tool for data representation, are applied in a huge variety of fields and are subject of modern research. NN can be subdivided by their intrinsic structure, for example feed-forward neural networks, recurrent neural networks (RNN) for sequential data structures or convolutional neural networks for image and video processing, just to name a few. While only the former two categories are described here, comprehensive information may be found in [174, 175].

5.2.1 Feed-forward neural networks

Figure 5.1 illustrates the basic structure of a feed-forward neural network, sometimes also referred to as multilayer perceptron. Depending on the application, individual networks may have a different layout, although some key aspects are common for all neural networks. A NN consists of an input layer, at least one hidden layer and an output layer. Each layer is composed of an individual number of neurons or nodes. The input layer contains a neuron for each input variable used to train the network. The number of hidden layers as well as the number of neurons per hidden layer is arbitrary and is usually optimised in regard to the application. It is common to add a bias node to the input and the hidden layers, which can be understood as an offset parameter and should not be confused with a bias in the statistical sense. The number of neurons in the output layer depends on the number of classes a network is required to distinguish. The final value of each neuron in the output layer corresponds to the network's prediction. The neurons in the different layers can be represented as vectors of dimensions $n_{\text{inputs}} + 1$, $n_{\text{hidden}} + 1$ and n_{classes}. Each connection in the network represents a weight, $w_{ji}^{(l)}$. The superscript l specifies the hidden layer, while the indices represent the connection of the j-th neuron in the layer with the i-th neuron of the previous layer. Due to these connections, information can be propagated through the network and by optimisation of the weights during the training procedure the network's performance can be enhanced. The weight parameters between the layers can be summarised to matrices and the neurons' activations, a_j, in the first hidden layer are constructed from linear combinations of the input variables:

$$a_j = w_{j0}^{(1)} + \sum_{i=1}^{k} w_{ji}^{(1)} x_i, \tag{5.1}$$

with $w_{j0}^{(1)}$, $w_{ji}^{(1)}$ and x_i being the bias, weights and input variables, respectively.

Afterwards, the hidden neurons are transformed by a differentiable activation function f:

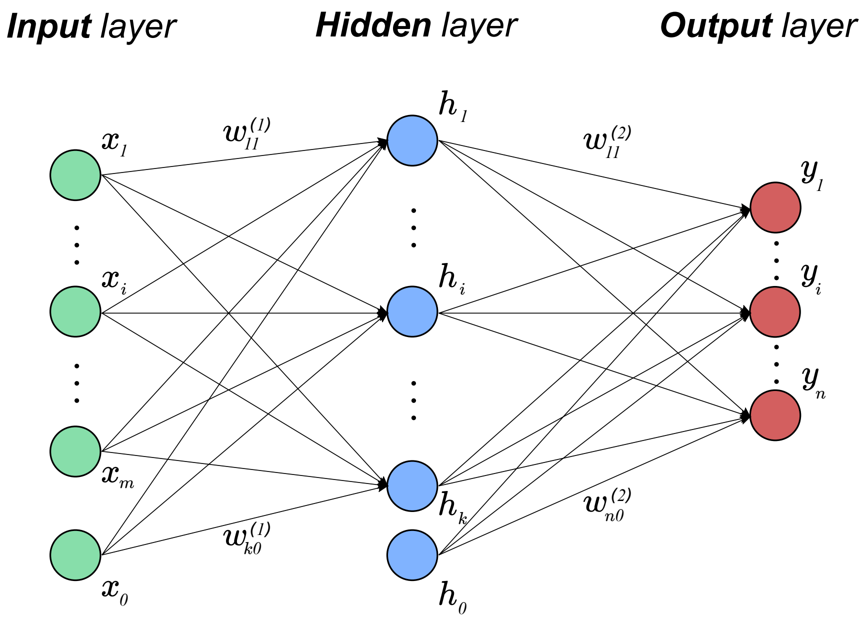
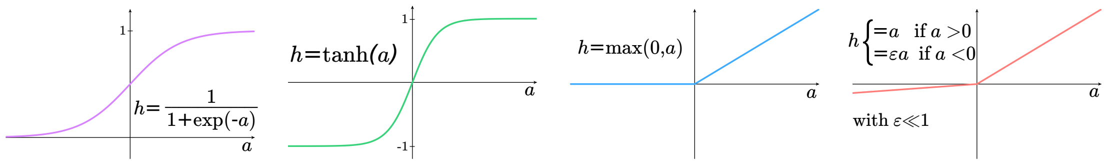


Figure 5.1: Basic structure of a multilayer perceptron, which consists of an input layer, an individual amount of hidden layers and an output layer. The various layers in turn are build up from single neurons, representing for example input variables of the data set or predicted outputs.

$$h_j = f(a_j). \tag{5.2}$$

The choice of activation function in the network is usually optimised based on the performance of the NN. Typical activation functions are the logistic sigmoid ($f(a) = \frac{1}{1+\exp{(-a)}}$) or the hyperbolic tangent function. Recently, the rectified linear unit (ReLU) defined as $f(a) = \max(0, a)$ and a modified version of it, denoted as leaky ReLU and defined as $f(a) = \max(\epsilon a, a)$, with $\epsilon \ll 1$, have become more popular. An advantage of the latter two functions is that they learn faster without affecting the performance, which may have a significant effect for very deep neural networks with a large number of hidden layers [176]. Figure 5.2 shows the four examples of activation functions.



(a) Sigmoid (b) Hyperbolic tangent (c) ReLU (d) Leaky ReLU

Figure 5.2: Examples of activation functions h used to transform activations a in a dedicated layer.

The procedure described by Equation 5.1 and Equation 5.2 needs to be repeated for each subsequent hidden layer until the last hidden layer is reached. In order to determine the final network outputs, denoted as y_i in Figure 5.1, the output neuron activation function σ needs to be applied on each neuron activation in the output layer as follows:

$$y_i = \sigma(a_i), \tag{5.3}$$

Since σ should map the multi-dimensional input space into a finite interval from 0 to 1, the logistic sigmoid function is a good candidate for the output neuron activation function in a binary classification problem. For problems with multiple output classes, like it is indicated in Figure 5.1, a multiclass generalisation of the logistic sigmoid function, referred to as the softmax function, is commonly used. Like the logistic sigmoid function in the case of only two distinct classes, the softmax function has its origin from a probabilistic point of view of the classification problem. Using Bayes' theorem, the posterior probability of the i-th output in a classification problem with more than 2 classes can be expressed as [174]:

$$p(y_i|\mathbf{x}) = \frac{p(\mathbf{x}|y_i)p(y_i)}{\sum_j p(\mathbf{x}|y_j)p(y_j)} = \frac{\exp(a_i)}{\sum_j \exp(a_j)} = \sigma(a_i), \tag{5.4}$$

and represents the probability of a certain class for the output neuron y_i given the input vector of a particular event. In the above equation, the quantities a_i are defined by:

$$a_i = \ln\big(p(\mathbf{x}|y_i)p(y_i)\big). \tag{5.5}$$

The softmax function maps the input variables to the finite interval for each output neuron between $0 \leq y_i \leq 1$ and additionally also satisfies $\sum_i y_i = 1$. The name "softmax" originates from the fact that if $a_i \gg a_j$ for all $j \neq i$, then $p(y_i|\mathbf{x}) \simeq 1$, while for all other output neurons $p(y_j|\mathbf{x}) \simeq 0$, and thus it represents some sort of "max" function.

Finally, the overall output function of the i-th output neuron y_i, which represents the forward propagation of information in a NN with a single hidden layer, like it is shown in Figure 5.1, can be determined as a composition of activation functions:

$$\begin{aligned} y_i(\mathbf{x},\mathbf{w}) &= (\sigma \circ h^{(1)})(\mathbf{x},\mathbf{w}) \\ &= \sigma\bigg(\sum_{j=1}^{k} w_{ij}^{(2)} h\bigg(\sum_{l=1}^{m} w_{jl}^{(1)} x_l + w_{l0}^{(1)}\bigg) + w_{i0}^{(2)}\bigg). \end{aligned} \tag{5.6}$$

With σ and h being the output activation function and the neuron activation function of the hidden layer, respectively. $\mathbf{x}$ and $\mathbf{w}$ representing the input vector and the weight matrix of a single event. In general, the output function of any network with n hidden layers is defined similarly:

$$y_i(\mathbf{x},\mathbf{w}) = (\sigma \circ h^{(n)} \circ \ldots \circ h^{(2)} \circ h^{(1)})(\mathbf{x},\mathbf{w}). \tag{5.7}$$

5.2.2 Recurrent neural networks

The flow of information in neural networks is illustrated by the connections of the neurons. Neural networks can only process information forward and backward between the input, via the hidden layers to the output layer. In addition, the dimension of the input layer, meaning the number of input variables to train the network, is predefined. Figure 5.3 illustrates the basic structure of a specific type of neural network that allows cyclic connections on neurons in hidden units. This kind of network is referred to as recurrent neural network (RNN) [177]. In principle, many different types of RNNs can be realised, based on how the cyclic connections are interpreted.

The essential feature of the RNN is that it can map not only the actual input to the output, but also information from previous inputs. As a result, it is possible to process sequences of data as well as inputs of variable length with these networks. Typical examples of applications for RNNs are audio or text processing, speech recognition, or as it will be shown later, processing of a collection of hadronic jets with arbitrary length. In order to

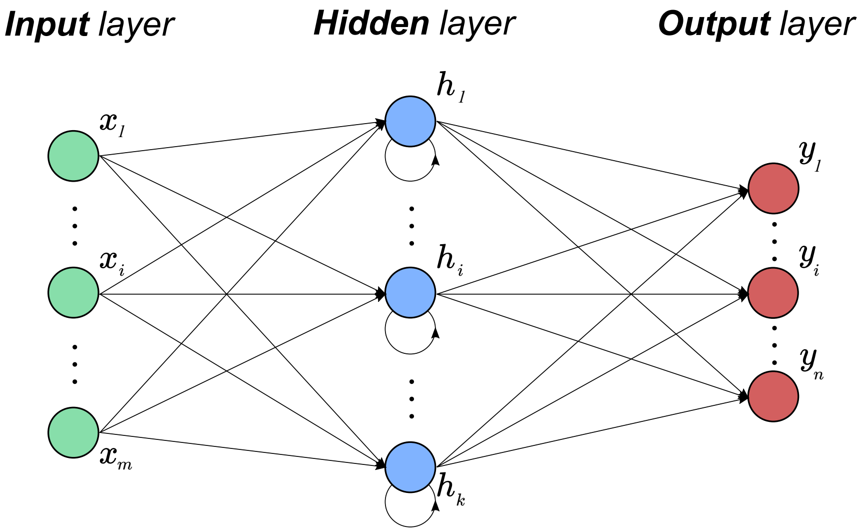


Figure 5.3: Basic concept of a recurrent neural network. In addition to standard connections between various layers, cyclic connections in hidden layers are allowed. As a consequence, a RNN is able to process not only the information from the current input neuron, but also from previous inputs. This means that sequential data as well as inputs of variable length can be handled. For simplicity the bias neurons are omitted.

understand how sequential information is passed through the RNN, Figure 5.4 visualises the unfolded cyclic connection of a single neuron. The superscript τ represents the total length of the sequence. The fundamental part of a neuron in a recurrent neural network is that it gathers the information of all input variables at the current step in the sequence as well as information from the input variables of the previous step in the sequence. Afterwards, both informations are merged, based on a set of rules which can be optimised during the training procedure, and passed on to the next step of the sequence and the output neuron. Comprehensive insights on recurrent neural networks may be found in [175, 178].

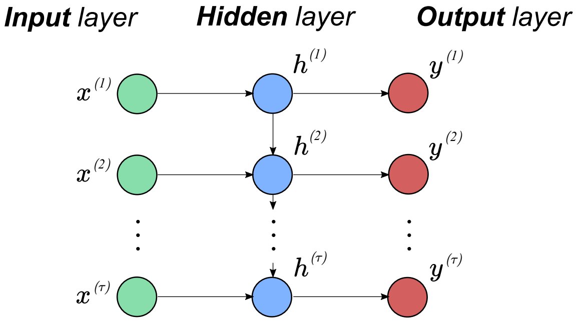


Figure 5.4: Unfolding of a single neuron of a hidden layer in a RNN. The unfolded neuron does not necessarily need a connection to the output neuron at each sequence. Instead, only a single output neuron at the end of the sequence is also viable.

Formally, the calculations to determine the neuron activation a_j^τ, at the current step in the sequence, are similar as described in Section 5.2.1, with the difference that also the activation of the previous step in the sequence needs to be taken into account:

$$a_j^\tau = \sum_{i=0}^{k} w_{ji} x_i^\tau + \sum_{n=1}^{l} w_{nj} h_n^{\tau-1}, \tag{5.8}$$

where τ denotes the length of the sequence, k represents the number of input variables,

l represents the number of neurons in the layer and $h_n^{\tau-1}$ is the neuron activation from the previous step in the sequence, transformed by the dedicated activation function $h_n^{\tau-1} = h(a_n^{\tau-1})$.

Long short-term memory

A major drawback of traditional RNNs is that the input information may not be conserved or is overwritten over long sequences in an unfolded neuron. In other words, the neuron forgets former inputs and cannot remember long term dependencies. This effect is referred to as the vanishing gradient problem [179, 180] and in short, it means that the gradients which need to be propagated through the network during the error backpropagation (see Section 5.3.4) may vanish.

However, there are algorithms that attempt to prevent these issues, such as the long short-term memory (LSTM) [181] unit. Due to its advanced internal architecture, LSTMs can handle long-term dependencies well. The main difference is that an ordinary RNN unit usually applies a single activation function to all the information that arrives at the unit, while the long-term, short-term and current information in a LSTM unit is treated differently. The key element of a LSTM is a so-called cell state, where the information from long-term, short-term and current input neurons is encapsulated and regulated by three types of so-called gates. Gates collect neuron activations from inside and outside the LSTM neuron and transform them with different activation functions. The three types of gates are the forget gate, the input and output gate, which are responsible for which information is discarded and which information is kept, which new information will be updated and finally which information is transferred to the output. Further detailed information on LSTMs is documented in [175, 178].

5.3 Network training

Until now, the considered networks as described in Section 5.2.1 and Section 5.2.2 are not able to make appropriate estimations, nor have they learned anything from the given dataset at all. Only the information flow from the input layer through the networks to the output layer has been outlined. As already mentioned, the networks need to be trained.

5.3.1 Loss function

In order to derive proper estimations from the network output, the weights, which describe the connections between the layers, need to be optimised accordingly. The basic concept of the training procedure is to find a set of weight parameters that minimise a particular function. This function is commonly referred to as error function, loss function or cost function, and usually gives a relation between the estimated output function $y_i(\mathbf{x}, \mathbf{w})$ and the corresponding true target values t_i.

In a binary classification task, where only two different classes are present, the output of a ML algorithm may predict values from 0 to 1 and represents conditional probabilities for both classes $p(0|\mathbf{x}) = y(\mathbf{x}, \mathbf{w})$ and $p(1|\mathbf{x}) = 1 - y(\mathbf{x}, \mathbf{w})$. Since the true target values can be only 0 or 1, they are interpreted by a Bernoulli distribution of the form:

$$p(t|\mathbf{x}, \mathbf{w}) = y(\mathbf{x}, \mathbf{w})^t \big(1 - y(\mathbf{x}, \mathbf{w})\big)^{1-t} \tag{5.9}$$

The loss function can be derived from this equation by building the negative log likelihood which results in:

$$L(\mathbf{w}) = -\sum_{i=1}^{m} \left[t_i \ln y_i(\mathbf{x}_i, \mathbf{w}) + (1 - t_i) \ln(1 - y_i(\mathbf{x}_i, \mathbf{w})) \right], \tag{5.10}$$

where the sum runs over m training events. The loss function of Equation 5.10 is called *cross-entropy*, and has the advantage that it yields faster training results and improved generalisation results for classification problems [182].

Going further to multi-classification tasks that apply a softmax output activation function (see Equation 5.4), the loss function is slightly modified to the form:

$$L(\mathbf{w}) = -\sum_{i=1}^{m} \sum_{j=1}^{n} t_{ij} \ln y_j(\mathbf{x}_i, \mathbf{w}), \tag{5.11}$$

where the output consists of n mutually exclusive classes and the true targets are hence interpreted as a $m \times n$ identity matrix.

One particular feature of the two presented loss functions, which is essential for the training, is that the derivative of the loss with respect to the output activation leads to the following form:

$$\frac{\partial L}{\partial a_i} = y_i - t_i. \tag{5.12}$$

The index i correspond to the dedicated output unit, while for the multi-classification problem the target gets a second index since it involves a matrix.

5.3.2 Parameter optimisation

The loss function $L(\mathbf{w})$ can be imagined as a surface in the hyperspace spanned over each weight parameter. Thus, the goal is the find the set of weights at which the gradient of the loss function vanishes:

$$\nabla L(\mathbf{w}) = 0 \tag{5.13}$$

Additionally, it needs to be verified whether such stationary point is a minimum, maximum or saddle point. However, there may be many points in the weight space at which the condition of Equation 5.13 is satisfied because the loss function usually has a highly nonlinear dependence on the weights. Moreover, large networks contain a huge amount of weight parameters, hence there might be also local minima which might be equivalent to other points in the weight space as well. Due to these facts, finding a global minimum analytically is nearly impossible and most algorithms perform an iterative procedure to determine the weight parameters. Most algorithms are defined such that the weights of the following iteration, $\mathbf{w}^{(i+1)}$, depend on the current weight parameters, $\mathbf{w}^{(i)}$, starting with an arbitrary choice of initial weights $\mathbf{w}^{(0)}$:

$$\nabla \mathbf{w}^{(i+1)} = \mathbf{w}^{(i)} + \Delta \mathbf{w}^{(i)}, \tag{5.14}$$

where the superscript i denotes the number of the actual training cycle. After each training cycle, $\nabla L(\mathbf{w})$ is updated using the new weight parameters.

It is also worth mentioning that it is not even necessary to find the global minimum of the loss function, but various observed local minima should be compared in order to reach a sufficient and generalised performance.

5.3.3 Stochastic gradient descent

The most common approach to find the optimal set of weights uses the gradient on the loss during the update of the weights and hence is referred to as gradient descent:

$$\nabla \mathbf{w}^{(i+1)} = \mathbf{w}^{(i)} - \eta \nabla L(\mathbf{w}^{(i)}). \tag{5.15}$$

The parameter η should be positive and is called the learning rate. At each iteration the gradient on the loss function is updated and the method is repeated. An illustration of the optimisation of weight parameters is shown in Figure 5.5.

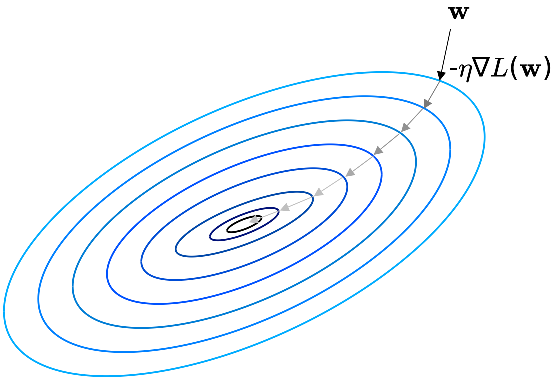


Figure 5.5: Stochastic gradient descent. The ellipses represent contours in the weight space. By iteratively updating the weights with respect to the actual gradient of the loss function, the trajectory is moving towards the minimum.

A potential drawback is that the gradient descent method needs to process the complete training set in order to compute gradient of the loss.

However, there also exist modified versions of the gradient descent method, which tend to be useful especially on large datasets. The stochastic gradient descent (SGD) method only computes the gradient of one event at a time:

$$\nabla \mathbf{w}^{(i+1)} = \mathbf{w}^{(i)} - \eta \nabla L_n(\mathbf{w}^{(i)}). \tag{5.16}$$

Where the index n refers to the dedicated event. Repeating the SGD method multiple times with randomly chosen starting points tend to result in a sufficient network performance. Another common modification is to gather a small amount of events to so-called mini-batches and determine the average of the gradient of the loss function for those batches at a time, instead of using single events. Especially when training on large datasets, this method tends to train faster, while resulting in a sufficient performance.

Usually, the trajectory that indicates the iteratively updated weight parameters in Figure 5.5 is not a smooth line. Due to small or noisy gradients, the trajectory can be very shaky and the training procedure can become slow and ineffective. Thus, there are more advanced gradient-based algorithms that include additional parameters mainly to accelerate the learning procedure. Such parameters are for example the learning rate decay, which decreases the learning rate after each iteration in order to get closer to the minimum in the weight space and thus achieve better performance. The momentum parameter [183], for example, introduces slight changes on the direction of the gradient which may also help to speed up the training in some cases. Popular algorithms that incorporate adaptive learning rates are, for example, the AdaGrad [184] and Adam [185] optimisation methods.

5.3.4 Error backpropagation

As noted in the previous sections, the calculation of the gradient of the loss function plays a central role in the training procedure. A breakthrough for the success of neural networks was the development of the error backpropagation technique [177], which provides a computationally efficient method in order to determine the derivatives of the loss function with respect to the weight parameters. The basic concept of the method is to propagate the error backwards through the network.

The loss as given in Equation 5.10 or Equation 5.11 can be expressed as a sum over all considered events:

$$L(\mathbf{w}) = \sum_n L_n(\mathbf{w}). \tag{5.17}$$

Considering a feed-forward neural network as it is outlined in Section 5.2.1, the activation of a particular neuron is defined by the sum over all connections that point to the neuron:

$$a_j = \sum_i w_{ji} h_i, \tag{5.18}$$

which looks similar to Equation 5.1, although the bias neuron has been included in the sum and the index i starts from 0. h_i represents the activation of a neuron that points to the actual neuron, which is associated by the weight parameter w_{ji} and is already transformed by the activation function $h_i = h(a_i)$. In principle, Equation 5.18 holds for any neuron in the hidden layers or the output layer in which the index i stands for any neuron from the input or the hidden layer if it is connected to neuron j. Since the loss L_n depends on the activation a_j, which is a function of the weight w_{ji}, the chain rule can be applied in order to determine the derivative of the loss with respect to the dedicated weight:

$$\frac{\partial L_n}{\partial w_{ji}} = \underbrace{\frac{\partial L_n}{\partial a_j}}_{\equiv \delta_j} \frac{\partial a_j}{\partial w_{ji}}, \tag{5.19}$$

where δ_j is referred to as the error of the neuron. From Equation 5.18, it can be seen that:

$$\frac{\partial a_j}{\partial w_{ji}} = h_i. \tag{5.20}$$

Thus, Equation 5.19 can be simplified to:

$$\frac{\partial L_n}{\partial w_{ji}} = \delta_j h_i. \tag{5.21}$$

If the weight is interpreted as a connection between two neurons, Equation 5.21 means that the derivative is expressed by the product of the error of the neuron δ_j to which the weight points to and the transformed activation of the neuron from which the connection emerges. In addition, for bias neurons h_i is simply 1. Therefore, in order to calculate the derivatives, the errors of each neuron in the hidden and output layers need to be determined. For all neurons in the output layer, Equation 5.12 already gives the solution for the corresponding error, while for the errors of neurons in the hidden layers, the chain rule is required:

$$\delta_j \equiv \frac{\partial L_n}{\partial a_j} = \sum_k \frac{\partial L_n}{\partial a_k}\frac{\partial a_k}{\partial a_j}, \tag{5.22}$$

The sum in the above formula runs over all subsequent neurons to which the neuron is connected. Finally, the definition of δ_j can be used to derive the error backpropagation formula [174]:

$$\delta_j = h'(a_j)\sum_k w_{kj}\delta_k, \tag{5.23}$$

which reflects the purpose of the error backpropagation technique. First, the errors in the output layer are determined and subsequently each error of the predecessor neurons is computed. Hence the information is propagated backwards through the network.

5.4 Performance measure

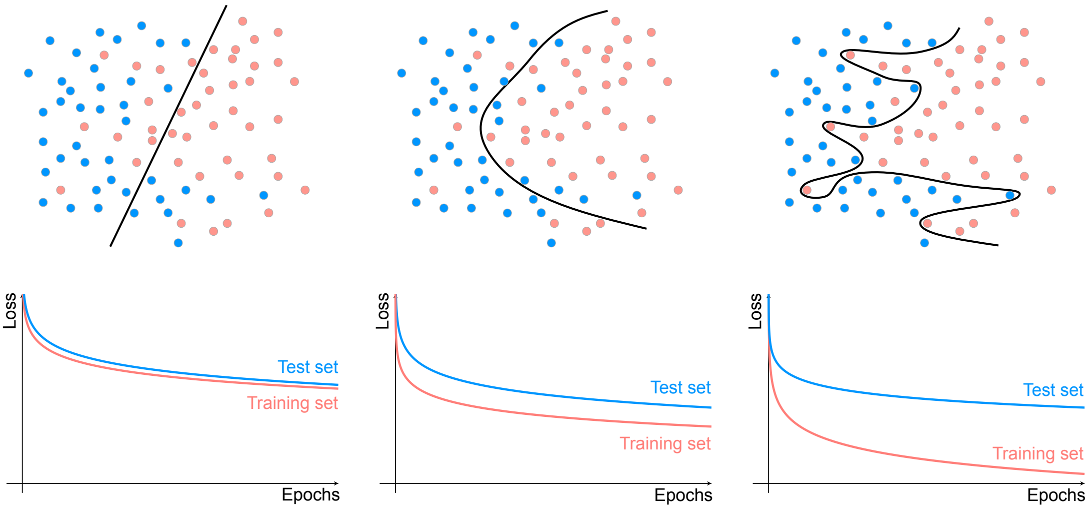
It has been mentioned that finding the global minimum for the loss function in the weight space is nearly impossible. In order to develop a generalised model that performs also well on unobserved events, it is not even necessary to find the global minimum. However, the weight parameters should be tuned to some extent and possible configurations should be compared.

Figure 5.6 illustrates three possible situations of a machine learning algorithm that has been trained to distinguish between two classes. In the first case, the model is not well advanced and did not learn from the given input variables (see Figure 5.6(a)). Models of this kind are denoted as underfitted or undertrained. Possible improvements would be to add further input variables, extend the network architecture by adding additional hidden layers or more neurons in the hidden layers or gathering more data. Such behaviour manifests itself in the distribution of the loss as a function of the number of training epochs, such that the loss for training and test set is large. The other extreme would be that the model learns each single property from each event in the training set (see Figure 5.6(c)). The loss for the training set is small, while the loss for the test set is large. This behaviour is referred to as overfitting or overtraining and is a crucial challenge in ML applications. There are many reasons for overtraining, such as a too small and too diverse dataset or the network structure may be too complex. Figure 5.6(b) indicates an optimally trained model. The parameters are sufficiently optimised and the loss for the training and test set is similar as well as small. In any case, more data is always favourable and leads to a positive effect on the model's performance, however, it is not always trivial to accomplish.

In the following, various performance metrics are described and have been used in order to evaluate the performance of the model used in analysis `bWN-fullSet`. A further discussion on the performance metrics follow in Section 8.2.1.

Confusion matrix

After the class of an event has been predicted by the model, it can be compared with the true target class of the event and categorised to different categories. In the case of binary classification there are two true targets, like signal and background, 0 or 1. Hence, a signal event can be classified correctly as a signal, denoted as true positive (TP) event, or it could be falsely identified as a background event. Events of this type are called false positives (FP). Similarly, a background event can be correctly estimated as background, referred to as true negative (TN) event, or mis-classified as a signal event, that is denoted as false

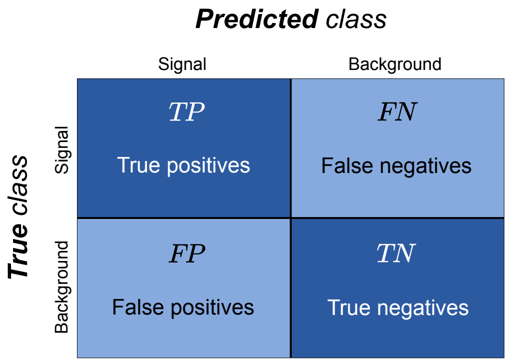


(a) Underfitting (b) Optimal loss (c) Overfitting

Figure 5.6: Loss diagnostics.

negative (FN) event. Those four categories can be visualised via the so-called confusion matrix as shown in Figure 5.7.

In general, the confusion matrix is not limited to binary classification tasks and the dimension of the matrix can be extended depending on the number of distinct classes.



***Predicted** class*

***True** class*	Signal	Background
Signal	TP True positives	FN False negatives
Background	FP False positives	TN True negatives

Figure 5.7: Confusion matrix. Signal and background events can be predicted correctly or misclassified, resulting in four categories, true positives, false positives, true negatives and false negatives.

Accuracy

Equipped with the categories from the confusion matrix the accuracy of the model is defined as:

$$\mathrm{acc} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{5.24}$$

which reflects the general ability of the model to classify events properly. In a typical search for supersymmetry the statistics of the simulated SM background is much larger than the statistics of the simulated SUSY signal events. Accuracy alone is not a good metric because of the large imbalance between the signal and background. A significant disparity between the number of signal and background events can lead to a large accuracy value even if only the background events are correctly classified.

Signal efficiency

Furthermore, the signal efficiency $\epsilon_{\mathrm{Sig.}}$ is defined as the fraction of correctly predicted signal events over all true signal events:

$$\epsilon_{\mathrm{Sig.}} = \frac{TP}{TP+FN}. \tag{5.25}$$

$\epsilon_{\mathrm{Sig.}}$ is also known as the true positive rate (TPR), recall or sensitivity. The signal efficiency metric is robust against an imbalance between the signal and background statistics.

Background rejection

The background rejection is defined as follows:

$$r_{\mathrm{Bkg.}} = 1 - \underbrace{\frac{TN}{TN+FP}}_{\text{False positive rate}} . \tag{5.26}$$

The background rejection is also known as the false positive rate and represents the fraction of correctly predicted background events over all background events and is also robust against an imbalance between the number of signal events and background events.

Receiver operating characteristic

The right plot in Figure 5.8 illustrates the distribution of the activation of the final output neuron within a range from 0 to 1, for signal and background events, respectively. If a threshold on the output score y is selected within 0 to 1, the area under the two distributions on the right of the threshold reflects the signal and background efficiencies. Hence, the background rejection can be plotted as function of the signal efficiency as shown in the left plot of Figure 5.8, with each point on the curve corresponding to a certain threshold on the output score. This distribution is referred to as the receiver operating characteristic (ROC) and its shape as well as the area under curve, denoted as AUC, also describes the performance of the model.

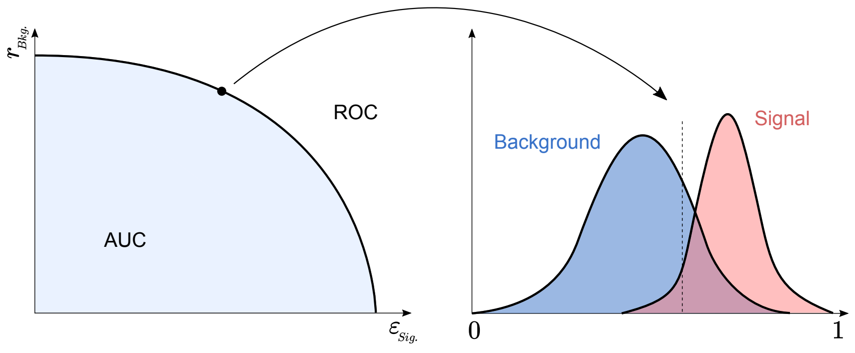


Figure 5.8: Receiver operating characteristic. The ROC can be derived from the distribution of the output activation of signal and background events.

Chapter 6

Statistical data analysis

This chapter is devoted to the statistical tools and methods used in order to obtain quantitative interpretations from the conducted analyses. First, the likelihood-based fit procedure, test statistics and the concept of statistical significance in the context of p-values is described. Moreover, the CL_s technique, necessary to determine exclusion intervals, is briefly discussed. Finally, systematic uncertainties, which are incorporated in the statistical evaluations as so-called nuisance parameters are explained in detail.

The data analysis, which is based on these concepts, is accomplished using the `HistFitter` software framework [186], which in turn adopts the functionality of the `HistFactory` package [187].

6.1 Likelihood

The statistical evaluation of the analyses in Chapter 7 and Chapter 8 follows the concepts and definitions described in [188]. In the analyses, a maximum likelihood fit is performed for the signal regions (see Section 4.1) in order to test the background-only and the background+signal hypotheses against the observed number of events. The expected event yield in each region is described by a Poisson probability density function with the sum of background events as the mean value. Signal events may be included to the background events, depending whether a background-only or background+signal hypothesis is considered. The estimated background events may be taken from simulation or are scaled to the observed events in the corresponding CR using a normalisation factor computed in the fit to the observed data. In the latter approach, the normalisation is controlled by independent free parameters. Depending on the considered hypothesis, signal processes may be included in the fit or omitted. Hence, the signal normalisation is also controlled by a free parameter, denoted as the signal strength μ.

Systematic effects, originating from experimental or theoretical nature, are taken into account in the fit procedure and are interpreted as so-called nuisance parameters. A nuisance parameter θ that follows a Gaussian probability density function is implemented as a smearing factor on the expected events in the likelihood. Further on, the mean (θ_0) of the Gaussian distribution is defined by the nominal prediction, whereas the standard deviation (σ) is determined by the size of the systematic uncertainty:

$$\mathcal{N}(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(\theta - \theta_0)^2}{2\sigma^2} \right). \tag{6.1}$$

Furthermore, by substituting $\alpha = (\theta - \theta_0)/\sigma$, the mean of the new nuisance parameter α becomes 0 and the standard deviation becomes 1.

Statistical uncertainties of the SM background processes are taken into account as additional nuisance parameters, if the relative statistical uncertainty of the expected number of events in a signal region is larger than 5%.

The number of signal events s_r and SM background events b_r in a corresponding region of phase space r are predicted by a model with a set of nuisance parameters $\boldsymbol{\theta}$, which are estimated in a maximum likelihood fit. The likelihood function is defined as the product of the Poisson probabilities for all regions:

$$L(\mu, \boldsymbol{\theta}) = \prod_{r \in \text{Regions}} \frac{(\mu s_r + b_r)^{n_r}}{n_r!} e^{-(\mu s_r + b_r)} \prod_{i \in \text{Syst.}} \mathcal{N}(\theta_i), \tag{6.2}$$

with the number of observed events n_r in the corresponding region of phase space. Depending on the fit procedure, the CR and/or SR are considered in the likelihood function. The expected number of background events also depends on the background normalisation factors, which are determined via a background-only fit. In this fit configuration, the likelihood is maximised for $\mu = 0$ and the signal regions are not included in the fit, hence the signal contribution is not taken into account.

6.2 Test statistics and p-values

The test statistic t_μ is defined in order to quantify hypotheses with different configurations for the signal strength μ, given an observed number of data events. Reasonable configurations for the signal strength are for example $\mu = 0$, which corresponds to the SM expectation. This hypothesis is usually referred to as the background-only hypothesis. A concurrent scenario would be if potential signal events are considered to the total expected number of events and $\mu \geq 0$. Since a signal region is typically not free of background, this hypothesis is also called background+signal hypothesis. Depending on the hypothesis in questions, either background-only or background+signal hypothesis can be considered as the null hypothesis and the alternative hypothesis. In order to compare competing hypotheses, the profile likelihood ratio is defined as [188]:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}, \tag{6.3}$$

with $\hat{\hat{\boldsymbol{\theta}}}$ being the estimator that maximises the likelihood for a given signal strength μ, while for the likelihood function in the denominator both $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are estimators that maximise the likelihood. Thus, the profile likelihood ratio assigns a number within the range of 0 and 1 to the hypothesis in question and describes the level of agreement of observed data with the hypothesis specified by μ.

It is common to define the test statistic t_μ as:

$$t_\mu = -2 \ln \lambda(\mu) \tag{6.4}$$

Here, larger values of t_μ reflect larger disagreement between the observed data events and the hypothesis in question. Further on, the level of incompatibility between data and the hypothesis can be determined by the p-value:

$$p_\mu = \int_{t_{\mu,\text{obs.}}}^{\infty} f(t_\mu|\mu) dt_\mu. \tag{6.5}$$

$t_{\mu,\text{obs.}}$ is the measured test statistic given the observed data and $f(t_\mu|\mu)$ is the probability density function of t_μ under the hypothesised value of μ. A p-value may be also interpreted as a statistical significance $Z = \Phi^{-1}(1 - p_\mu)$. In this context, Z reflects multiples of the standard deviation of a Gaussian distributed variable. Φ^{-1} is the inverse of the cumulative distribution of the Gaussian distribution. In the particle physics community, evidence for new physics is found if the background-only hypothesis is rejected with a significance of at least $Z = 3$ standard deviations, corresponding to a p-value of $p_0 = 1.35 \times 10^{-3}$, whereas a discovery is only claimed if $Z \geq 5$ ($p_0 \leq 2.87 \times 10^{-7}$).

The requirement to exclude a potential signal hypothesis is less stringent with a threshold of $p_1 = 0.05$ ($Z = 1.64$), corresponding to a 95 % confidence level.

6.3 The CL$_s$ technique

If the sensitivity for a potential signal model in a dedicated search region is very low or a downward fluctuation of data compared to the prediction is observed, results can be mis-interpreted and a potential signal could be falsely excluded. The CL$_s$ technique [189] aims to prevent excluding signal models with low sensitivity and is defined as:

$$\mathrm{CL}_s = \frac{p_\mu}{1 - p_0}. \tag{6.6}$$

Here, the denominator can only be less than or equal to unity, hence the exclusion criterion based on the CL$_s$ method is more stringent than the usual requirement which is based on p_μ. If the signal sensitivity is very low, then p_μ decreases, but $1 - p_0$ also decreases, hence the exclusion condition that CL$_s$ is below a certain threshold will not be satisfied.

The CL$_s$ technique is a widely acknowledged procedure in the high energy physics community and is used in the analyses documented in this work in order to derive an exclusion limit on the signal strength. A signal is excluded at 95% confidence level if the CL$_s$ value is smaller than 0.05 for the background+signal hypothesis.

6.4 Systematic uncertainties

Systematic effects play a crucial role in statistical data analysis, because they affect the profile-likelihood ratio as a function of the signal strength μ, which can be interpreted as a loss of information about μ. The systematic uncertainties on the signal as well as the background estimates originate from experimental effects, such as imperfect description of detector conditions and finite resolution of the object calibration, and from uncertainties in the theoretical predictions and modelling.

The same set of experimental and theoretical systematic uncertainties is considered in analysis `bWN-subSet` and `bWN-fullSet`. The various sources of systematic uncertainties are summarised in the following, while the impact of the particular effect on the final result is discussed later on in the dedicated section of the respective analysis.

6.4.1 Experimental uncertainties

Jet calibration

In analysis `bWN-subSet` as well as analysis `bWN-fullSet`, the dominant sources of experimental uncertainties are from imperfect knowledge of the jet energy scale and jet energy resolution [159, 190]. In both analyses, a strongly reduced set of jet energy scale uncertainties is considered. A large set of related representations is combined in a strongly reduced set of uncertainty sources, in which an adequate level of correlations is still preserved [191]. Two concurrent strongly reduced sets of jet energy scale uncertainties have been applied to the analyses individually and the obtained results were not significantly different, which indicated that both analyses are not sensitive to the correlations associated with the uncertainties. By this procedure the complexity of the analysis is reduced.

Flavour tagging

The efficiency measurements for the calibration of b-tagged jets provide also variations on the efficiencies, which are further used to determine dedicated systematic uncertainties [163]. In addition, systematic uncertainties associated with the mis-tagging rate of c-tagged jets and light-flavour jets are derived [192, 193].

Lepton reconstruction, isolation and identification

As already mentioned in Section 4.3.2 and Section 4.3.3, dedicated scale factors, based on efficiency measurements, are applied to electrons and muons to correct for imperfect detector simulations. Moreover, the efficiency measurements provide also independent variations of the reconstruction, identification and isolation efficiencies for electrons and muons. The effect of these variations has been evaluated and is found to be negligible in both analyses.

Missing transverse momentum

Systematic uncertainties associated with the scale as well as the resolution of the missing transverse momentum are also considered, leading to changes in both the magnitude and the direction of the missing transverse momentum. Since $E_{\mathrm{T}}^{\mathrm{miss}}$ is determined from the physics objects in an event, $E_{\mathrm{T}}^{\mathrm{miss}}$ is re-calculated for each systematic variation of the calibrated objects. Additionally, the soft term (see Section 4.3.7) also contributes to the uncertainties in the scale and the resolution. A detailed methodology for the scale and resolution uncertainties of $E_{\mathrm{T}}^{\mathrm{miss}}$ is given in [166].

Residual experimental uncertainties

Additionally, further experimental uncertainties such as the modelling of pileup, trigger efficiencies and the estimation of the integrated luminosity have been investigated. However, those uncertainties have a minor impact on the final results.

6.4.2 Theory uncertainties

$t\bar{t}$ modelling

The largest source of theoretical systematic uncertainty corresponds to the modelling of the $t\bar{t}$ background. Three independent sources of systematic effects in the context of the modelling of the top quark pair process have been investigated. Systematic uncertainties may be introduced due to the choice of the Monte-Carlo event generator used to generate the hard scattering process, the modelling of the fragmentation and hadronisation in the parton shower and the amount of initial- and final-state radiation.

To evaluate the size of the systematic effects, alternative $t\bar{t}$ samples are compared to the nominal sample. In analysis `bWN-subSet`, the POWHEG-BOX event generator interfaced with PYTHIA6 for parton showering is used for the generation of nominal $t\bar{t}$ events. Hence, the uncertainty from the hard scattering process is determined by comparing $t\bar{t}$ events generated with the POWHEG-BOX and the MG5_aMC@NLO event generators. Uncertainties due to fragmentation and hadronisation effects are computed from the comparison of nominal $t\bar{t}$ events and events generated with the alternative Herwig++ [194] software framework for parton showering. The impact of altering the amount of initial- and final-state radiation is estimated from comparisons of individual POWHEG-BOX+PYTHIA6 samples with different parton shower radiation, modified factorisation and renormalisation scales and NLO radiation.

The nominal $t\bar{t}$ sample in analysis `bWN-fullSet` is generated using the POWHEG-BOX event generator and PYTHIA8 parton shower. The systematic uncertainty on the hard scattering process is based on the comparison of $t\bar{t}$ events generated with the POWHEG-BOX and the MG5_aMC@NLO event generators. Systematic uncertainties from fragmentation and hadronisation effects are determined from the comparison of nominal $t\bar{t}$ events and with those generated using the alternative HERWIG7 [195] package for parton showering. Systematic uncertainties from initial- and final-state radiation in `bWN-fullSet` are computed from the comparison of the nominal $t\bar{t}$ events with $t\bar{t}$ events which are reweighted using internal event weights. The internal event weights include the variation of factorisation and renormalisation scales as well as variable shower radiation.

In both analyses, the expected event yield of the $t\bar{t}$ background is normalised to data in a dedicated control region. Hence, only the extrapolations from the control region into the signal regions are affected by the systematic uncertainties of the $t\bar{t}$ background process. Systematic uncertainties on the $t\bar{t}$ modelling does not affect the overall normalisation. In order to compute the effect of the systematic uncertainty on the extrapolation of the $t\bar{t}$ events, a so-called transfer factor is computed for the alternative as well as for the nominal sample:

$$TF = \frac{y_{\mathrm{SR}}}{y_{\mathrm{CR}}}, \tag{6.7}$$

where y_{SR} and y_{CR} represent the number of simulated $t\bar{t}$ events in the signal region and the control region, in which the $t\bar{t}$ contribution is normalised to data. Afterwards, the systematic uncertainty δ from radiation, hard scatter and hadronisation effects is determined from the ratio of the transfer factors of the alternative and nominal $t\bar{t}$ events:

$$\delta = \frac{TF' - TF}{TF} = \frac{TF'}{TF} - 1. \tag{6.8}$$

Here, TF' is the transfer factor of the alternative $t\bar{t}$ background sample and TF is the transfer factor of the nominal $t\bar{t}$ sample.

No uncertainty on the cross section is needed, because the the $t\bar{t}$ background is normalised to data in a dedicated control region.

Modelling of the residual SM background sources

In both analyses, the number of events from SM backgrounds other than top quark pair processes is minor, with a total contribution of less than 15 % in the signal regions. Theoretical uncertainties on these processes are found to have a negligible impact on the final result. Thus, only a conservative uncertainty on the theoretical cross section of the respective SM background is applied in both analyses. The uncertainty on the cross section ranges from 25 %-100 %, depending on the SM background.

SUSY signal modelling

Since the targeted SUSY signal events and the dileptonic $t\bar{t}$ events are very similar in the considered 3-body decay phase space, additional radiation can have a significant impact on the kinematic variables and on the acceptance of the expected signal events in the signal regions. As a consequence, systematic uncertainties on the signal acceptance due to the modelling of initial- and final-state radiation and the variation of the factorisation and renormalisation scales are considered. In addition, an uncertainty on the expected signal cross section is taken into account in both analyses. In the first analysis, `bWN-subSet`, the signal cross sections are normalised to NLO accuracy and an uncertainty on the expected signal cross section is taken from an envelope of cross section predictions using different PDF sets as well as factorisation and renormalisation scales, as described in [121]. The signal cross sections in the second analysis, `bWN-fullSet`, are calculated to approximate next-to-next-to-leading order and the considered uncertainties on the cross sections are derived using the PDF4LHC15_mc PDF set, following the recommendations in [143].

Chapter 7

Search for top squark pair production in the 3-body decay mode with a single lepton final state using 36.1 fb^{-1} of pp collision data

In this chapter, the results from a search for top squark pairs in a compressed decay scenario are presented and discussed. As already mentioned, this analysis targets a signal model as illustrated in Figure 2.4, denoted as $\tilde{t}_1 \rightarrow bW\tilde{\chi}_1^0$. The compressed decay scenario is characterised by the potential mass difference of the $\tilde{t}_1$ and the $\tilde{\chi}_1^0$, $m_W + m_b \leq \Delta m = m_{\tilde{t}_1} - m_{\tilde{\chi}_1^0} \leq m_t$. The results of this search are published in [58] together with various other analyses that attempt to probe similar supersymmetric models and share common properties of the final state.

7.0.1 Event preselection

In analysis strategy is described in Section 4.1.

In order to monitor the modelling of important kinematic variables in the simulated events, a loose event preselection is defined. The selection criteria of the preselection are mainly driven by basic properties of the targeted SUSY model, so that SM background events which are very different from the target signal are excluded at an early stage of the analysis. The preselection is necessary to sort out events that did not pass particular quality requirements, and guarantees that only data is selected in which all sub-detectors were fully functional. The event sample at preselection level also serves as a starting point for the definition of the signal, control and validation region.

The event preselection is summarised in Table 7.1. Selection criteria are based on the physics objects which have to fulfil the respective quality requirements as discussed in Chapter 4. Events are required to have exactly one electron or muon with a minimum p_{T} of at least 25 GeV after the preselection. Events with additional leptons are suppressed by requiring no further baseline leptons in the sample. The analysis is performed on the dataset which is recorded using $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers only. In order to guarantee that the trigger is fully efficient, an additional requirement on $E_{\mathrm{T}}^{\mathrm{miss}} > 230$ GeV is defined. Selected events must have a minimum of 4 hadronic jets, and at least one jet must be tagged as a b-jet. The transverse momentum of the leading jet is required to be larger than 50 GeV, whereas the p_{T} threshold for all other hadronic jets in the selected events is $p_{\mathrm{T}} > 25$ GeV. Furthermore, the azimuthal angle between the direction of $E_{\mathrm{T}}^{\mathrm{miss}}$ and the two leading jets is required to be $|\Delta\phi(j_{1,2}, E_{\mathrm{T}}^{\mathrm{miss}})| > 0.4$, and selected events are also required to have a transverse mass of $m_{\mathrm{T}} > 130$ GeV. As a consequence, a potential contribution from QCD multi-jet events is effectively suppressed. In fact, a proper estimation of the multi-jet background in a data driven-manner indicated that the contribution of multi-jet processes is negligible due to the high $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold, the requirement on $|\Delta\phi(j_{1,2}, E_{\mathrm{T}}^{\mathrm{miss}})|$ and the tight cut on m_{T}. Moreover, the stringent requirement on m_{T} effectively suppresses contributions from W+jets processes and semi-leptonic $t\bar{t}$ events, as already discussed in Section 4.3.9. Events including hadronic τ candidates are vetoed by the requirement of $m_{\mathrm{T2}}^{\tau} > 80$ GeV.

Trigger	$E_{\mathrm{T}}^{\mathrm{miss}}$ trigger only
Second-lepton veto	no additional baseline leptons
Number of leptons	exactly 1 e or μ
Lepton p_{T}	> 25 GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 230 GeV
Number of jets (min. p_{T})	$\geq 4(25$ GeV$)$
Number of b-tagged jets	≥ 1
Leading jet p_{T}	> 50 GeV
$\lvert\Delta\phi(j_{1,2}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})\rvert$	> 0.4
m_{T2}^{τ}	> 80 GeV
m_{T}	> 130 GeV

Table 7.1: Summary of the event preselection criteria.

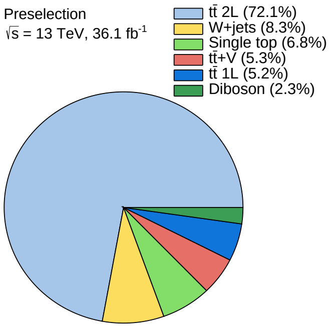
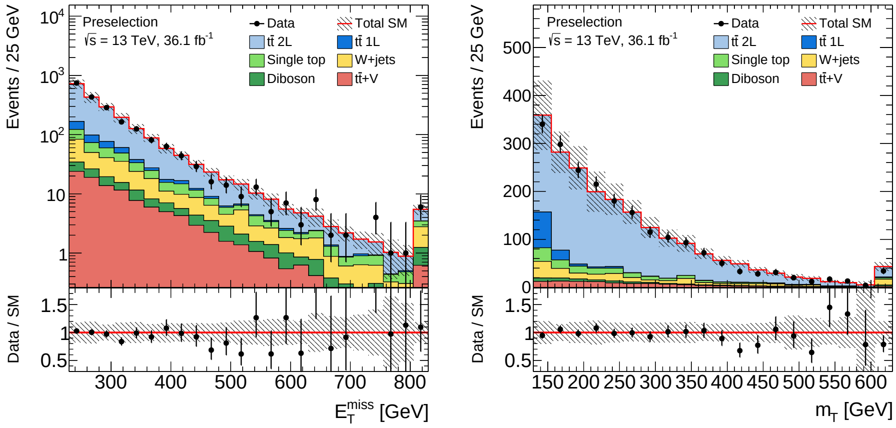


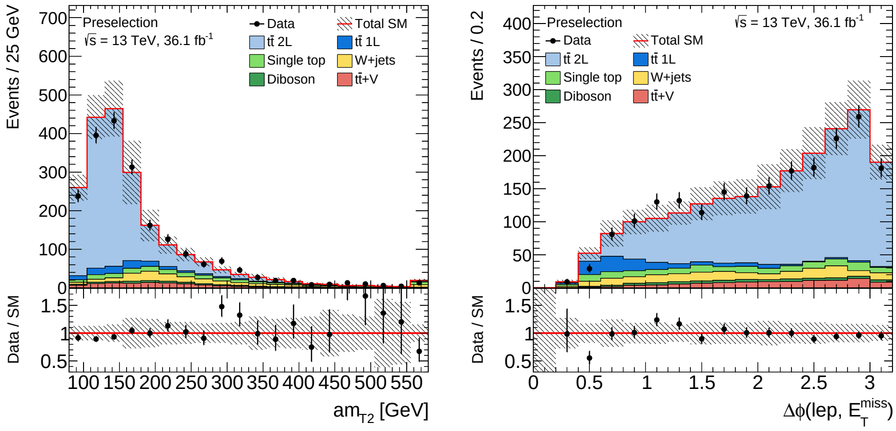
Figure 7.1: Relative contributions of the individual SM backgrounds after the preselection.

After the preselection, the dominant SM background stems from top quark pair production (see Figure 7.1). In fact, the dominant fraction of $t\bar{t}$ events are dileptonic top quark pair decays (~ 72 %), in which one lepton is either not reconstructed, not identified, removed by the overlap-removal procedure, or one W boson decays leptonically and the other via a hadronically decaying τ lepton. Distributions of the most important kinematic observables after the preselection are shown in Figure 7.2. In these distributions, the $t\bar{t}$ background prediction is scaled by the normalisation factor obtained from a likelihood fit of the CR, which is described in detail in Section 7.2. Other SM background processes are normalised with the respective theoretical cross sections. Good agreement between observed data and the SM prediction is obtained. The shaded band around the total SM prediction as well as the shaded band in the data/SM ratio include statistical and experimental uncertainties. $t\bar{t}$ modelling uncertainties which are defined only for the signal and validation region are not included in the uncertainty band.

In the following, simulated events are used to define the event selections for the signal, control and validation regions.



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) m_{T}



(c) am_{T2} (d) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$

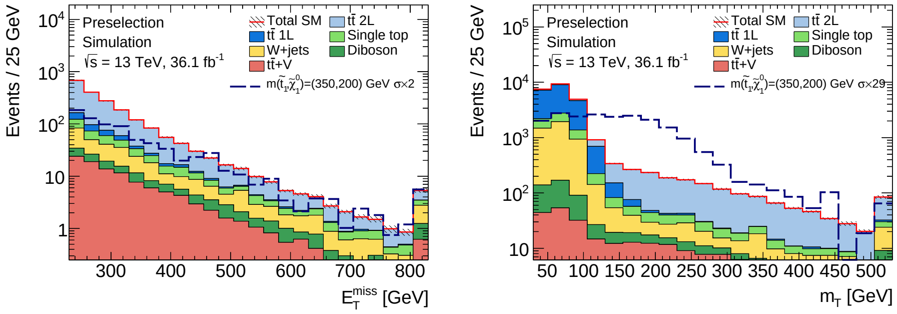
Figure 7.2: Kinematic distributions of key observables after the preselection is applied. The $t\bar{t}$ background prediction is scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross sections. The hatched band around the total SM prediction and the hatched band in the data/SM ratio include statistical and systematic uncertainties, except the $t\bar{t}$ modelling uncertainties, which are determined only for the VR and SR, respectively. Overflows are included in the last bin.

7.1 Signal region

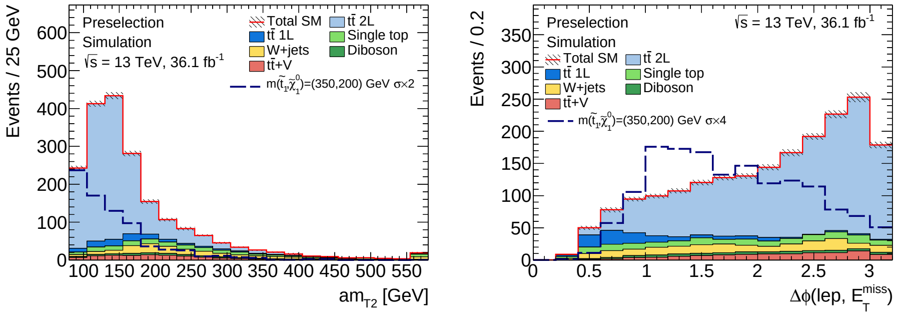
The metric for the optimisation is to maximise the discovery significance[1] for the $\tilde{t}_1 \rightarrow bW\tilde{\chi}_1^0$ decay scenario. For the optimisation procedure, a benchmark signal model with a top squark mass of 350 GeV and a neutralino mass of 200 GeV is used. The procedure for the optimisation of the signal region selection is based on an iterative algorithm. At first, the algorithm determines the maximum expected significance by selecting an upper or lower bound for all variables at preselection level, respectively. Afterwards all considered variables are ranked by their corresponding significance. The variable selection with the highest significance is then applied to the preselection and the procedure is repeated for all remaining variables afterwards. The algorithm stops if the expected significance does not change significantly by adding additional cut selections. Moreover, statistical uncertainties and a constant relative systematic uncertainty of 25 % on the SM background events are considered in the computation of the significance. The four kinematic variables which showed to most discriminative power between the signal and background are considered in the iterative optimisation approach and are illustrated in Figure 7.3. Since m_T indicated the best separation power after a very loose event selection it was decided to apply this variable in the initial event preselection. Furthermore, the distributions for signal processes are scaled by multiples of its nominal cross section, in order to stress the distinct kinematics compared to the SM background processes. Especially am_T2 and $\Delta\phi(\ell, E_\mathrm{T}^\mathrm{miss})$ tend to have good discriminating power. Due to the kinematic construction of the am_T2 variable (see Section 4.3.9) and the compressed mass splitting of the considered top squark decay scenario, the SUSY signal events peak below 100 GeV in am_T2, compared to the dileptonic $t\bar{t}$ background which has its maximum around the top quark mass. The separation power of the $\Delta\phi(\ell, E_\mathrm{T}^\mathrm{miss})$ variable is due to the neutralinos in the final state of the top squark pair decay, in which the $E_\mathrm{T}^\mathrm{miss}$ is no longer associated with the neutrino from the leptonic decay of the W boson. The signal peaks around the range of 1 to 1.5 radian while the dileptonic $t\bar{t}$ events tend to have a larger angular distance between the lepton and $E_\mathrm{T}^\mathrm{miss}$ above 2.5 radian.

---

[1] The expected significance for the benchmark signal is calculated using the `BinomialExpZ` function provided by the `RooStats` [196] framework.



(a) $E_\mathrm{T}^\mathrm{miss}$ (b) m_T



(c) am_T2 (d) $\Delta\phi(\ell, E_\mathrm{T}^\mathrm{miss})$

Figure 7.3: Kinematic distributions of the most important variables of the analysis after the preselection. Only simulated events are shown. For the distribution of m_T, the cut on the shown quantity is omitted in the event preselection in order to stress the discriminating power of this observable. Only statistical uncertainties are shown in the distributions. Overflows are included in the last bin. The distribution for signal events is scaled by multiples of its nominal cross section, to illustrate the distinct kinematics compared to the SM background.

The final signal region definition is listed in Table 7.2, and the remaining simulated SM background events and the expected number of events of the signal benchmark model are listed in Table 7.3, while Figure 7.4 shows the relative contribution from the individual

SM background sources. As already mentioned, dileptonic $t\bar{t}$ processes make up the largest contribution, while the fraction of semi-leptonic $t\bar{t}$ events is fairly small. As a consequence, the top quark pair processes are not divided into fractions in the following, except it is mentioned explicitly. Figure 7.5 shows the comparison of data and SM expectation before the fit. Plotted quantities are kinematic distributions of the variables used in the selection, the cut on the dedicated quantity is omitted, except for the distribution of the m_T variable, because a tight requirement on m_T is already applied at event preselection level. The vertical line in the plot shows where and in which direction the cut on this variable would be applied.

	Signal region definition
$E_\mathrm{T}^\mathrm{miss}$	$> 300\,\mathrm{GeV}$
m_T	$> 130\,\mathrm{GeV}$
am_T2	$< 110\,\mathrm{GeV}$
$\Delta\phi(\ell, E_\mathrm{T}^\mathrm{miss})$	< 2.5
Exclusion technique	shape-fit in am_T2
Bin boundaries in am_T2 [GeV]	$[80, 91, 97, 106, 118, 130]$

Table 7.2: Overview of the signal region selection. In addition, the event preselection as defined in Table 7.1 is applied. The last row shows the bin intervals for the shape-fit scenario.

Signal region	Events
$t\bar{t}$ 2L	57.8 ± 1.5
$t\bar{t}$ 1L	2.2 ± 0.5
Single top	1.9 ± 0.3
$t\bar{t} + V$	1.7 ± 0.2
W+jets	1.4 ± 0.3
Diboson	0.9 ± 0.1
Total SM	65.9 ± 1.9
$m(\tilde{t}, \tilde{\chi}_1^0) = (350, 200)$ GeV	60.9 ± 6.0

Table 7.3: Number of expected events in the signal region for 36.1 fb^{-1}. Only statistical uncertainties are shown.

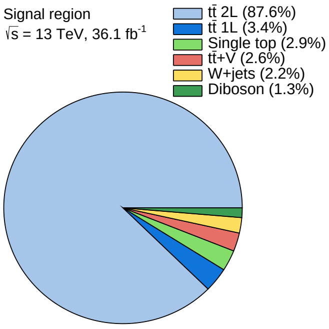
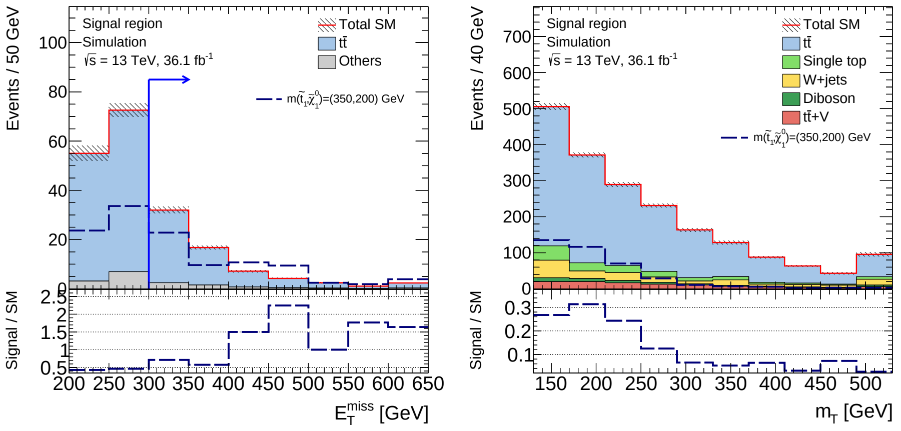
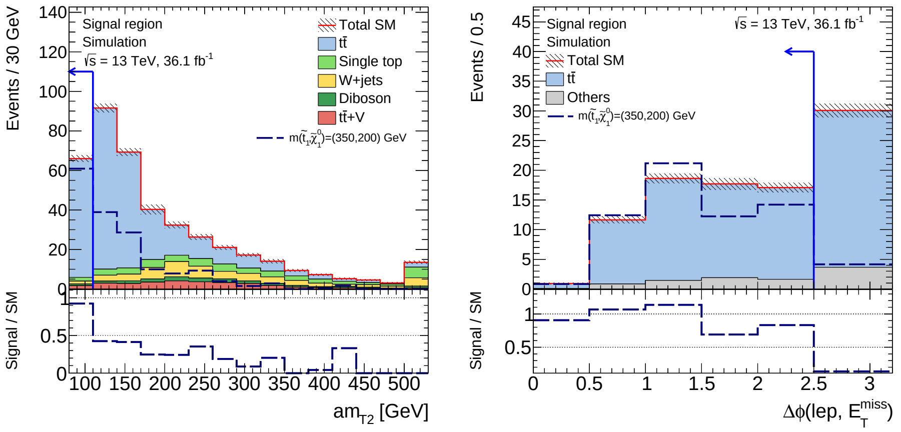


Figure 7.4: Relative contributions of the individual SM backgrounds in the signal region.



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) m_{T}



(c) am_{T2} (d) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$

Figure 7.5: Kinematic distributions of the variables used to define the signal region. The selection on the displayed quantity is omitted as indicated by the vertical line, except for the distribution of the m_{T} variable, because a tight requirement on m_{T} is already applied at event preselection level. The direction of the arrow on top of the vertical line indicates the region of accepted events in the signal region. Only statistical uncertainties are included. The category 'Others' in the legend represents the sum of minor SM backgrounds that contribute less than 5 % of the total SM background. The lower panel shows the ratio of signal to SM background in each bin respectively.

As shown in the last row of Table 7.2, a one-dimensional shape-fit is performed in am_{T2} including five bins in order to improve the exclusion sensitivity for the targeted SUSY signal, in addition to the single-bin strategy. The bins are designed such that the expected number of signal events is similar in all bins. In addition, all bins are required to contain no less than 15 expected background events. Expected significances for the single-bin and the shape-fit scenario are given in Table 7.4 for a few selected signal points with different $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$. It is found that, utilising different signal-to-background ratios in the bins of the shape-fit, the search sensitivity can be increased with respect to the single-bin configuration.

Signal point	Discovery scenario	Exclusion scenario
$m(\tilde{t},\tilde{\chi}_1^0) = (350, 185)$ GeV	1.91	2.04
$m(\tilde{t},\tilde{\chi}_1^0) = (350, 200)$ GeV	3.09	3.38
$m(\tilde{t},\tilde{\chi}_1^0) = (350, 230)$ GeV	2.70	3.18
$m(\tilde{t},\tilde{\chi}_1^0) = (350, 260)$ GeV	1.04	1.57

Table 7.4: Comparison of the expected significance for various signal points in the single-bin and shape-fit scenario. Statistical and systematics uncertainties are included in the computation.

7.2 Background estimation and validation

The dominant SM background process in the analysis originates from $t\bar{t}$ production. Therefore, $t\bar{t}$ processes are estimated in a dedicated control region in which the number of simulated $t\bar{t}$ events is normalised to the observed data with a likelihood fit. In order to guarantee a reliable estimation in the SR, the event selection for the CR has to be kinematically as close as possible to the SR. Other stringent requirements for the definition of the control region are that a potential signal contribution should be avoided, while the purity of top quark pair produced events should be enhanced. As a consequence, the event selection of the CR is based on the requirements of the SR, but requirements on key variables are either inverted or loosened.

Afterwards, the normalisation of $t\bar{t}$ events, obtained in the control region, is verified in a particular validation region. Again, the event selection of the VR has to be kinematically similar to the signal and the control region, with an enhanced purity of $t\bar{t}$ processes and a reduced contribution from potential SUSY signal events.

Since the signal and the control region are both used in the likelihood fit, the event selections for these regions are kept disjunct. The validation region is not part of the fit procedure, but its event selection is still defined to be statistically independent.

Other contributions to the total SM background are estimated from simulation.

7.2.1 Control region

The event selection of the control region is listed in Table 7.5 together with the corresponding event selections of the signal and validation region, in order to stress the kinematic differences. For the definition of the CR, the requirement on $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ is inverted to enhance the purity of $t\bar{t}$ processes. Potential signal contamination is reduced by shifting the am_{T2} selection to $130\,\mathrm{GeV} < \mathrm{am}_{\mathrm{T2}} < 170\,\mathrm{GeV}$. The upper cut on am_{T2} is necessary to avoid an overlap with other analyses in publication [58] which require high am_{T2}. In addition, the requirement on $E_{\mathrm{T}}^{\mathrm{miss}}$ is loosened to further enhance statistics of top quark pair processes. In general, the signal contamination has been tested for different signal points in the 2-, 3- and 4-body phase space of the analysis and was found to be less than 5%.

The predicted number of SM events in the control region is listed in the upper section of Table 7.6 and compared with the corresponding number of simulated SM events, which is listed in the lower section of the table. The number of observed events and predicted events are identical by construction, because the $t\bar{t}$ processes are normalised to data in this region. The event yields of SM backgrounds other than $t\bar{t}$ production are identical, because they are estimated from simulations. The uncertainties in the upper list include statistical and systematic uncertainties, while in the lower list only statistical uncertainties are given. The composition of the SM background in the control region is illustrated in

	SR	CR	VR
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 300 GeV	> 230 GeV	> 300 GeV
am_{T2}	< 110 GeV	$130 - 170$ GeV	$110 - 130$ GeV
$\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$	< 2.5	> 2.5	> 2.5

Table 7.5: Overview of the event selection of the signal region, the control region to estimate the normalisation of the $t\bar{t}$ background, and the validation region to verify the estimation method. The event preselection as defined in Table 7.1 is applied to all regions.

Figure 7.6. As intended, the purity of $t\bar{t}$ events is well above 90 % in this selection.

Control region	Events
Observed events	293
Predicted background events	292.9 ± 17.1
$t\bar{t}$	275.2 ± 20.0
Single top	5.5 ± 5.5
$t\bar{t} + V$	7.7 ± 7.6
W+jets	2.3 ± 2.3
Diboson	2.2 ± 0.7
Simulated background events	272.4 ± 6.9
$t\bar{t}$	254.7 ± 5.3
Single top	5.5 ± 0.5
W+jets	2.3 ± 4.3
$t\bar{t} + V$	7.7 ± 0.4
Diboson	2.2 ± 0.2

Table 7.6: Observed and predicted SM events in the control region. The individual uncertainties are correlated and do not add up quadratically to the total uncertainty. The lower section lists the number of SM events obtained from simulation.

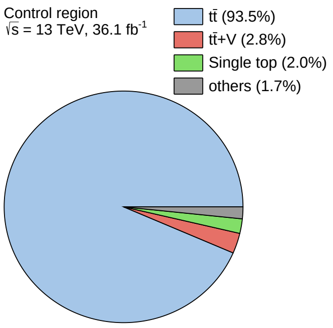
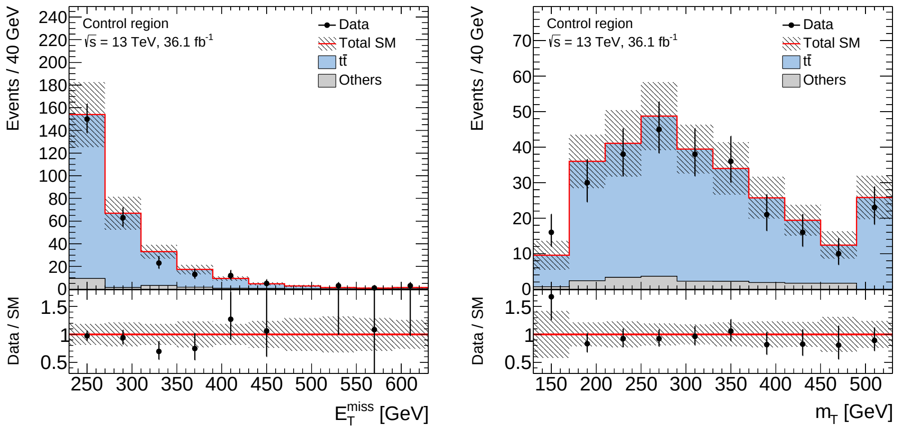
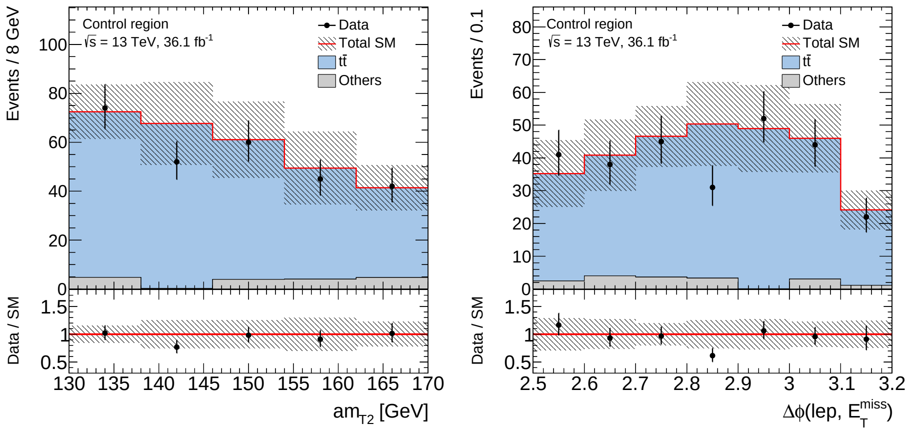


Figure 7.6: Relative contributions of the individual SM backgrounds in the control region.

The distribution of the four key variables in the control region is shown in Figure 7.7 for observed data events and the predicted SM events respectively. The observed data is in good agreement with the predicted SM events.



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) m_{T}



(c) am_{T2} (d) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$

Figure 7.7: Kinematic distributions of key variables for events that pass the control region selection. The $t\bar{t}$ events are scaled by the normalisation factor obtained in the CR. Residual SM background processes are normalised with the respective theoretical cross section. The category 'Others' in the legend represents the sum of minor SM backgrounds that contribute less than 5 % of the total SM background. The lower panel shows the ratio of data to the SM prediction. The uncertainty band includes statistical and experimental systematic uncertainties. $t\bar{t}$ modelling uncertainties are only determined for the VR and SR and hence not included in the total uncertainty. The last bins include the overflow.

7.2.2 Validation region

In order to test the background estimation method, the normalisation factor for the $t\bar{t}$ background, obtained in the CR, is applied in the validation region and the observed number of events is compared to the predicted number of events.

The event selection for the VR is listed in the third column of Table 7.5. To get kinematically closer to the signal region the requirement on $E_{\mathrm{T}}^{\mathrm{miss}}$ is kept at 300 GeV, while the selection on am_{T2} is shifted to 110 GeV $<$ am$_{\mathrm{T2}}$ $<$ 130 GeV to avoid potential contamination from signal events. For similar reasons, the $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ selection is still inverted. As in the other regions of interest, the dominant SM background source is top quark pair production with a purity above 90 %. The number of observed events as well as a comparison of predicted SM events to the number of events obtained from simulation are given in Table 7.7. $t\bar{t}$ processes are normalised according to the normalisation factor obtained in the control region. The event yields of all other SM backgrounds are identical in the upper and lower list, because they are estimated from simulations. All uncertainties are included in the upper part of the table and only statistical uncertainties are shown in the lower part of the table. Relative fractions of SM backgrounds contributing to the validation region is shown in Figure 7.8.

Validation region	Events
Observed events	41
Predicted SM events	41.2 ± 6.5
$t\bar{t}$	38.0 ± 6.6
Single top	0.8 ± 0.8
$t\bar{t} + V$	1.5 ± 1.5
W+jets	0.5 ± 0.3
Diboson	0.4 ± 0.2
Simulated background events	38.4 ± 1.4
$t\bar{t}$	35.1 ± 1.32
Single top	0.8 ± 0.2
$t\bar{t} + V$	1.5 ± 0.2
W+jets	0.5 ± 0.2
Diboson	0.4 ± 0.2

Table 7.7: Observed and predicted SM events in the validation region. The individual uncertainties are correlated and do not add up quadratically to the total uncertainty. Only statistical uncertainties are included in the simulated SM backgrounds listed in the lower part of the table.

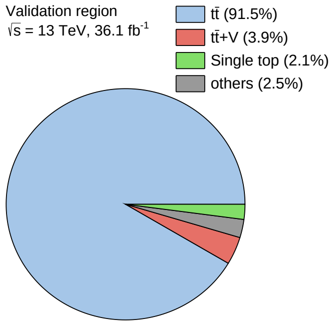
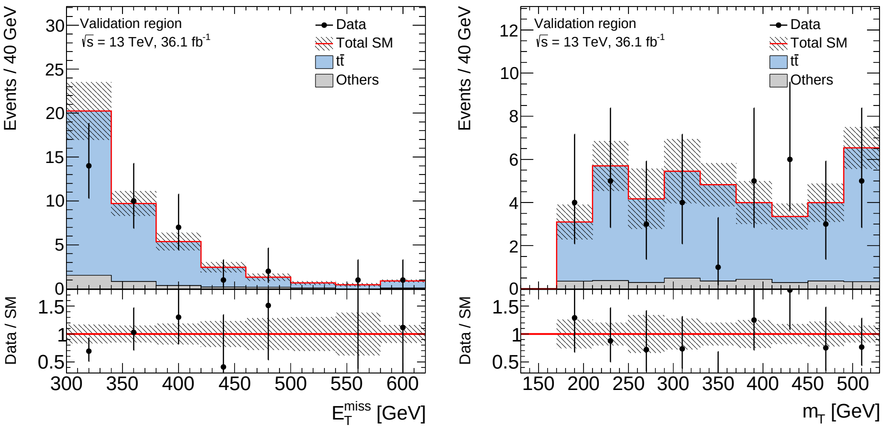
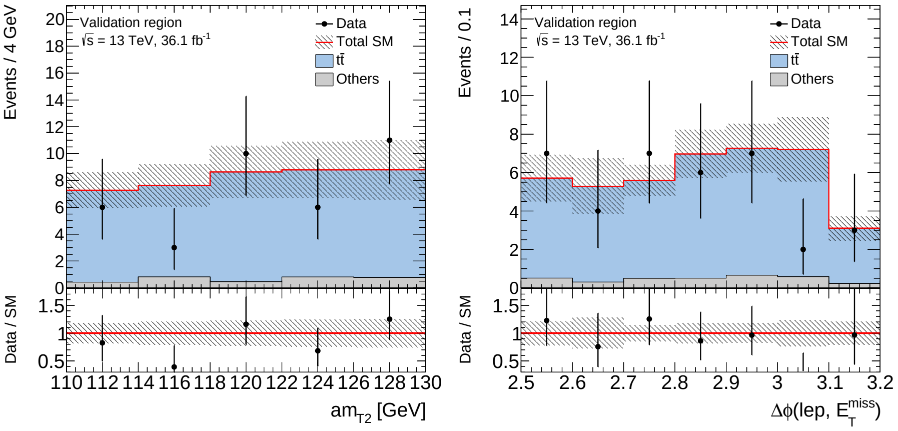


Figure 7.8: Relative contributions of the individual SM backgrounds in the validation region.

The kinematic distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$, m_{T}, am_{T2} and $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ in the validation region are shown in Figure 7.9. The observed events and predicted events match within statistical and systematic uncertainty.



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) m_{T}



(c) am_{T2} (d) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$

Figure 7.9: Kinematic distributions of key variables in the validation region. The $t\bar{t}$ events are scaled by the normalisation factor obtained in the CR. Residual SM background processes are normalised with the respective theoretical cross section. The category 'Others' in the legend represents the sum of minor SM backgrounds that contribute less than 5 % of the total SM background. The lower panel shows the ratio of data to the SM prediction. The uncertainty band includes statistical and systematic uncertainties. The last bins include the overflow.

7.3 Impact of systematic uncertainties

The various sources of systematic effects considered in the analysis are discussed in detail in Section 6.4. The basic methodology in order to evaluate the systematic effect is to compare the change in the number of predicted events in the regions of interest when varying the different sources of systematic uncertainties.

The dominant systematic uncertainties in the VR and SR are listed Table 7.8. The largest uncertainty is due to the modelling of the top quark pair production. This contribution is evaluated by comparing the results from alternative $t\bar{t}$ MC samples using the transfer-factor method (see Equation 6.7). By definition, the differences in the normalisation of the different MC samples cancel out in the transfer-factor method, and only differences in the shapes of kinematic distributions between the alternative $t\bar{t}$ samples influence the systematic uncertainty. However, the alternative $t\bar{t}$ samples are statistically limited in the tight event selection of the signal region, thus the original transfer-factor method is slightly modified. To reduce the statistical uncertainties, not the full signal region definition is used, but instead only the requirement of one key variable is applied on top of the preselection and the systematic uncertainty is computed. This procedure is repeated for all discriminating variables used for the signal region definition, and the corresponding uncertainties are summed up in quadrature. Although correlations between the variables are neglected, the systematic uncertainties still agree with the results from the original approach, only the statistical uncertainties are smaller.

Another large contribution to the total SM background uncertainty stems from the jet energy scale calibration. The reason is that the computation of $E_{\mathrm{T}}^{\mathrm{miss}}$ is affected by variations in the jet energy scales. Hence, also m_{T} and $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ are influenced by this systematic effect. Moreover, jet energy scales also impact the kinematic construction of am_{T2}.

Systematic uncertainty	VR	SR
Total predicted SM background	41.2	70.8
Total statistical ($\sqrt{N_{\mathrm{exp}}}$)	±6.4	±8.4
Total background uncertainty	±6.5 [15.8 %]	±15.7 [21.5 %]
$t\bar{t}$ radiation	±2.1 [5.1 %]	±9.0 [12.7 %]
$t\bar{t}$ hadronisation	±2.3 [5.6 %]	±8.6 [12.1 %]
$t\bar{t}$ generator	±4.0 [9.7 %]	±5.5 [7.8 %]
$t\bar{t}$ normalisation	±3.1 [7.5 %]	±5.3 [7.5 %]
Jet energy scale (1^{st} component)	±1.8 [4.4 %]	±4.7 [6.6 %]
Single top cross section	±0.8 [1.9 %]	±1.9 [2.7 %]
Statistics (simulated samples)	±1.5 [3.6 %]	±1.7 [2.4 %]
$t\bar{t} + V$ cross section	±1.5 [3.6 %]	±1.7 [2.4 %]
$E_{\mathrm{T}}^{\mathrm{miss}}$ resolution ($\perp$ component)	±0.8 [1.9 %]	±1.5 [2.1 %]
$E_{\mathrm{T}}^{\mathrm{miss}}$ resolution ($\parallel$ component)	±0.2 [0.5 %]	±1.2 [1.7 %]
Flavour tagging c mis-tag rate	±0.5 [1.2 %]	±1.2 [1.7 %]
Flavour tagging b-tagging efficiency	±0.8 [1.9 %]	±1.2 [1.7 %]
Jet energy resolution	±1.2 [2.9 %]	±1.1 [1.6 %]

Table 7.8: Summary of the dominant systematic uncertainties on the total predicted number of SM background events for the validation and signal region. Individual uncertainties may be correlated, and hence do not necessarily add up in quadrature to the total uncertainty. Systematic uncertainties which contribute less than 1 % are omitted in the list. The percentages show the size of the uncertainty relative to the total predicted SM background. The systematic uncertainties are ordered by the size of their relative contribution to the total predicted SM background in the SR.

The total statistical uncertainty of the total SM background ($\sqrt{N_{\mathrm{exp}}}$) is also given in the table in order to compare the size of the statistical uncertainty with the systematic uncertainty. The total relative uncertainty in the signal region is 21.5 %, which is larger than total statistical uncertainty. For completeness, the individual contributions of systematic effects to the total uncertainty for the five bins in the shape-fit are summarised in Table 7.9. As for the single-bin scenario, the largest systematic sources are the modelling of top quark pair production and the jet energy scale.

Systematic uncertainty	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Total predicted SM background	16.5	16	25.6	40.1	38.5
Total statistical ($\sqrt{N_{\mathrm{exp}}}$)	±4.1	±4.0	±5.1	±6.3	±6.2
Total background uncertainty	±4.5 [27.3 %]	±6.0 [37.5 %]	±5.3 [20.7 %]	±8.1 [20.2 %]	±8.3 [21.6 %]
$t\bar{t}$ hadronisation	±3.1 [18.8 %]	±2.4 [15.0 %]	±2.9 [11.3 %]	±3.5 [8.7 %]	±4.2 [10.9 %]
$t\bar{t}$ radiation	±2.2 [13.3 %]	±2.2 [13.8 %]	±3.1 [12.1 %]	±5.3 [13.2 %]	±4.7 [12.2 %]
$t\bar{t}$ normalisation	±1.2 [7.3 %]	±1.2 [7.5 %]	±1.9 [7.4 %]	±2.9 [7.2 %]	±2.9 [7.5 %]
Jet energy scale (1^{st} component)	±1.1 [6.7 %]	±0.9 [5.6 %]	±1.8 [7.0 %]	±2.0 [5.0 %]	±1.8 [4.7 %]
$t\bar{t}$ generator	±1.0 [6.1 %]	±4.7 [29.4 %]	±1.7 [6.6 %]	±2.4 [6.0 %]	±3.6 [9.4 %]
Statistics (simulated samples)	±0.8 [4.8 %]	±0.7 [4.4 %]	±1.1 [4.3 %]	±1.7 [4.2 %]	±1.3 [3.4 %]
Jet energy resolution	±0.8 [4.8 %]	±0.4 [2.5 %]	±0.01 [0.05 %]	±1.1 [2.7 %]	±1.0 [2.6 %]
Single top cross section	±0.5 [3.0 %]	±0.4 [2.5 %]	±0.7 [2.7 %]	±1.4 [3.5 %]	±0.9 [2.3 %]
$E_{\mathrm{T}}^{\mathrm{miss}}$ resolution ($\perp$ component)	±0.5 [3.0 %]	±0.5 [3.1 %]	±0.2 [0.8 %]	±0.5 [1.2 %]	±1.5 [3.9 %]
$t\bar{t}+V$ cross section	±0.5 [3.0 %]	±0.3 [1.9 %]	±0.6 [2.2 %]	±1.0 [2.5 %]	±1.2 [3.1 %]
Flavour tagging c mis-tag rate	±0.4 [2.4 %]	±0.2 [1.3 %]	±0.5 [2.0 %]	±0.5 [1.2 %]	±0.2 [0.5 %]
Pileup	±0.4 [2.4 %]	±0.02 [0.1 %]	±0.1 [0.4 %]	±0.2 [0.5 %]	±0.3 [0.8 %]
Jet energy scale (1^{st} component)	±0.4 [2.4 %]	±0.02 [0.1 %]	±0.02 [0.1 %]	±0.4 [1.0 %]	±0.2 [0.5 %]
Flavour tagging b-tagging efficiency	±0.3 [1.8 %]	±0.3 [1.9 %]	±0.3 [1.2 %]	±0.9 [2.2 %]	±0.9 [2.3 %]
$E_{\mathrm{T}}^{\mathrm{miss}}$ resolution ($\parallel$ component)	±0.2 [1.2 %]	±0.3 [1.9 %]	±0.3 [1.2 %]	±0.9 [2.2 %]	±0.3 [0.8 %]

Table 7.9: Summary of the dominant systematic uncertainties on the total predicted number of SM background events for the five bins in the shape-fit. Individual uncertainties may be correlated, and hence do not necessarily add up in quadrature to the total uncertainty. Systematic uncertainties which contribute less than 1 % are omitted in the list. The percentages show the size of the uncertainty relative to the total predicted SM background. The systematic uncertainties are ordered by their relative contribution to the total predicted SM background in the most sensitive bin (Bin 1) with respect to the SUSY signal benchmark.

The total relative uncertainty in the various bins ranges from 20.2 % to 37.5 %. Comparison to the total statistical uncertainty ($\sqrt{N_{\mathrm{exp}}}$) indicates that systematic effects are the dominant source of uncertainty in the shape-fit configuration, but compatible with the respective total statistical uncertainty.

7.4 Results and interpretation

The observed and predicted number of events in the signal region are given in Table 7.10. The observed data is well in agreement with SM expectation. In addition, the normalisation factor of the $t\bar{t}$ background $\mu_{t\bar{t}}$ obtained in the control region by the background-only fit, is shown in the table. Further on, a model-independent upper limit at 95 % confidence level on the expected ($S^{95}_{\mathrm{exp.}}$) and observed ($S^{95}_{\mathrm{obs.}}$) number of beyond-SM events is derived in the signal region. For this procedure, a generic signal model is assumed to contribute in the SR only. Neither experimental nor theoretical uncertainties, except for the luminosity, are considered for such a generic signal model.

In addition, the comparison between observed data and predicted SM background events in the validation as well as in the signal region is shown in Figure 7.10. Observed events are well in agreement with the SM prediction. The lower panel in the figure illustrates the difference between observed events (n_{obs}) and predicted SM background events (n_{exp}) divided by the total uncertainty (σ_{tot}) and is commonly referred to as pull. σ_{tot} is defined as $\sigma_{\mathrm{tot}} = \sqrt{\sigma^2 + \sigma^2_{n^{\mathrm{Poisson}}_{\mathrm{pred.}}}}$ and includes the total SM background uncertainty (σ^2) and a variation from the 68 % central confidence interval of a Poisson distribution ($\sigma^2_{n^{\mathrm{Poisson}}_{\mathrm{pred.}}}$). The pull represents an approximation of the significance of the p_0-value in the presence of a data excess.

	Signal region
Observed events	68
Total background	71 ± 16
$t\bar{t}$	65 ± 16
W+jets	1.41 ± 0.88
Single top	1.9 ± 1.9
$t\bar{t} + V$	1.7 ± 1.7
Multi boson	0.89 ± 0.28
$\mu_{t\bar{t}}$	1.08 ± 0.09
$S^{95}_{\text{exp.}}$	$35.2^{+10.5}_{-9.7}$
$S^{95}_{\text{obs.}}$	30.7

Table 7.10: Number of observed events in the signal region together with the predicted number of SM background events. All types of uncertainties are included. In addition, the normalisation factor ($\mu_{t\bar{t}}$) for the prediction of $t\bar{t}$ processes as obtained in the fit is listed, as well the expected ($S^{95}_{\text{exp.}}$) and observed ($S^{95}_{\text{obs.}}$) 95 % upper limit on the number of beyond-SM events.

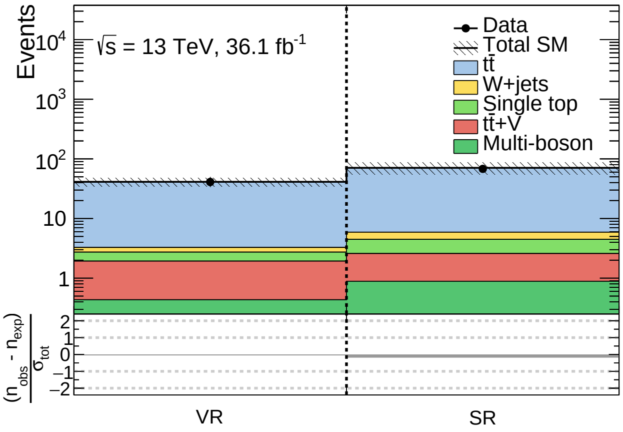
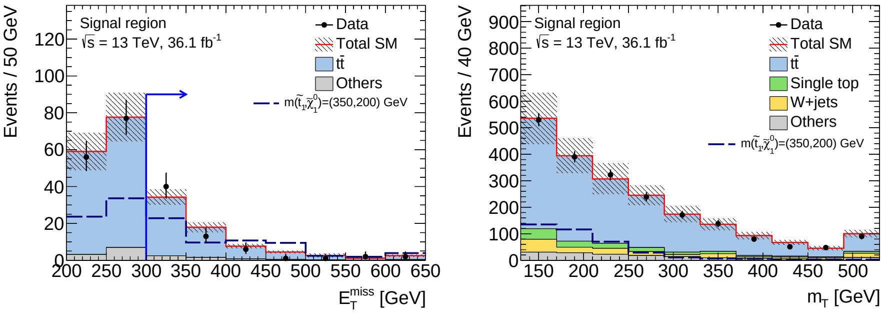


Figure 7.10: Comparison between the observed data (n_{obs}) and the predicted SM background (n_{exp}) events in the validation and the signal region. The shaded area around the SM expectation includes all uncertainties. The bottom panel shows the difference between observed and the predicted events divided by the total uncertainty (σ_{tot}).

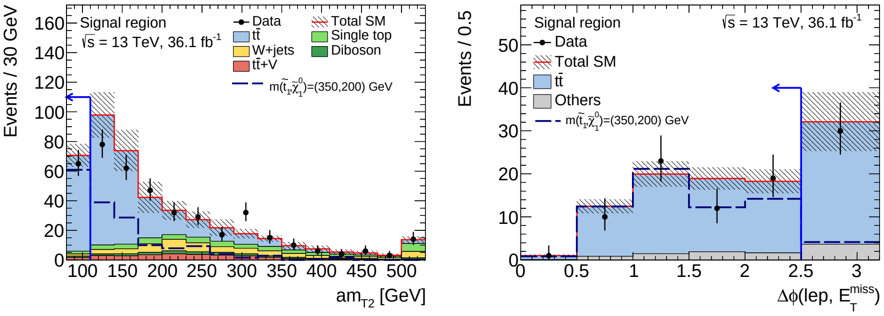
The kinematic distribution of $E_{\text{T}}^{\text{miss}}$, m_{T}, am_{T2} and $\Delta\phi(\ell, E_{\text{T}}^{\text{miss}})$ in the signal region are illustrated in Figure 7.11. Reasonable agreement between observed events and the SM prediction is found.

Since the result from the single-bin scenario does not show any hint for new physics beyond the standard model, the shape-fit procedure is performed in order to derive exclusion limits at 95 % confidence level. First, the numbers of observed events and the predicted numbers of SM background events for each bin contributing to the shape-fit are summarised in Table 7.11.

Figure 7.12 shows the am_{T2} distribution with a variable bin width matching the definition of the shape-fit. In all bins good agreement is found between the observed data and the SM prediction. Statistical as well as experimental and theoretical systematic uncertainties



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) m_{T}



(c) am_{T2} (d) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$

Figure 7.11: Kinematic distribution of relevant variables in the signal region. The requirement on the plotted quantity is omitted, but indicated by the vertical lines, except for the m_{T} distribution. The predicted $t\bar{t}$ background process is scaled with the normalisation factor obtained from the control region. Furthermore, the benchmark signal model is shown. The label 'Others' in the legend represents minor SM backgrounds that contribute less than 5% of the total background. All uncertainties are included and represented by the shaded area around the total SM expectation. The last bin contains overflows.

Exclusion scenario am_{T2} [GeV]	Bin 1 [80, 91]	Bin 2 [91, 97]	Bin 3 [97, 106]	Bin 4 [106, 118]	Bin 5 [118, 130]
Observed events	13	19	22	30	36
Total background	16.5 ± 4.5	16.0 ± 6.0	25.6 ± 5.3	40.1 ± 8.1	38.5 ± 8.3
$t\bar{t}$	15.1 ± 4.6	14.9 ± 6.0	23.4 ± 5.4	35.5 ± 8.1	35.3 ± 8.4
W+jets	0.1 ± 0.1	0.2 ± 0.2	0.7 ± 0.4	2.0 ± 1.1	0.6 ± 0.4
$t\bar{t} + V$	0.5 ± 0.5	0.3 ± 0.3	0.6 ± 0.6	1.1 ± 1.1	1.2 ± 1.1
Single top	0.5 ± 0.5	0.4 ± 0.4	0.7 ± 0.7	1.4 ± 1.4	0.9 ± 0.9
Diboson events	0.3 ± 0.1	0.2 ± 0.1	0.3 ± 0.1	0.2 ± 0.1	0.5 ± 0.2

Table 7.11: Number of observed events and predicted SM background events for 36.1 fb^{-1}, in each signal region contributing to the shape-fit. Individual uncertainties can be correlated and do not necessarily add up quadratically to the total background uncertainty.

are included. The bottom panel shows again the difference between data and the predicted SM background divided by the total uncertainty (σ_{tot}). In addition, the distribution of the benchmark signal model is shown.

Finally, exclusion limits are set for top squark pair production in the 3-body decay

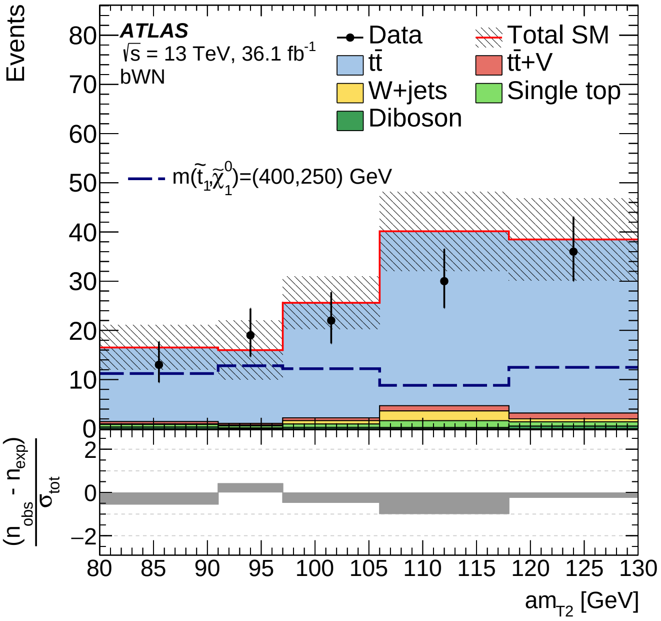


Figure 7.12: Kinematic distribution of am_{T2} for the shape-fit scenario. The event selections corresponding to the bins of the shape-fit are applied. $t\bar{t}$ events are scaled with the normalisation factor obtained from the control region. Statistical and systematic uncertainties are included. The last bin contains overflows. Additionally, the benchmark signal model is overlaid for comparison. The bottom panel shows the difference between observed and predicted events divided by the total uncertainty (σ_{tot}).

mode, as well as in the adjacent 2-body and 4-body phase space. Therefore, potential signal contributions to the control region and the five regions comprised by the shape-fit are taken into account. All uncertainties, except those for the theoretical signal cross section, are included in the fits.

The observed and expected exclusion contour as a function of the top squark mass and the neutralino mass in the targeted phase space are shown in Figure 7.13. The yellow uncertainty band ($\pm 1\,\sigma_{\mathrm{exp}}$) indicates the impact of the statistical and systematic uncertainties on the expected limit. The red dashed lines represent the uncertainty ($\pm 1\,\sigma_{\mathrm{th}}$) on the observed limit, which is obtained by varying the nominal signal cross section by one standard deviation of the theoretical cross section.

The grey shaded regions represent the exclusion contours from previous analyses that targeted the 2-, 3- and 4-body decay modes, respectively [173]. Those analyses were performed on the full Run 1 pp collision data recorded by the ATLAS collaboration at a centre-of-mass energy of 8 TeV and corresponds to an integrated luminosity of 20.3 fb^{-1}. The blue shaded region represents the exclusion contour from a search for direct top squark pair production performed on pp collision data at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 3.2 fb^{-1}. However, that analysis was optimised for signal models with a large top squark mass and a relatively small neutralino mass and is therefore not sensitive in the phase space of the 3-body decay mode.

Since there are no signal points below $m(\tilde{t}_1) = 220$ GeV considered in the 3- and 4-decay mode, the actual exclusion contour is bounded from the bottom, which is illustrated by the straight line of the actual exclusion contour at a top squark mass of approximately 200 GeV to 250 GeV.

The non-excluded area between the transition of the 3- and 4-body decay regions at a top squark mass of approximately 350 GeV is because of a reduction in the sensitivity. The reason of this decrease in sensitivity is because of the significant change of the kinematic properties of the signal models. For example, if the mass splitting $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$ decreases the momenta of the two b-tagged jets decrease. This results in a significant decrease of the acceptance of the p_{T} requirement on the b-tagged jets in the signal region.

In addition, Figure 7.14 shows the expected and observed exclusion contours in the plane spanned by the top squark mass and the mass splitting $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$. A greater level of detail is provided for the targeted 3-body decay region. The exclusion contour obtained from the results of the previous publication [173] is again illustrated by the grey hatched region. The acronyms *tN*, *bWN* and *bff'N* represent the different kinematics defined by $m(\tilde{t}_1) - m(\tilde{\chi}_1^0) > m(t)$, $m(t) > m(\tilde{t}_1) - m(\tilde{\chi}_1^0) > m(W) + m(b)$ and $m(W) + m(b) > m(\tilde{t}_1) - m(\tilde{\chi}_1^0) > 0$, respectively.

The performed analysis excludes top squark masses up to 460 GeV depending on the mass of the neutralino, under the assumption of $\mathcal{B}(\tilde{t}_1 \rightarrow bW\tilde{\chi}_1^0) = 100\%$.

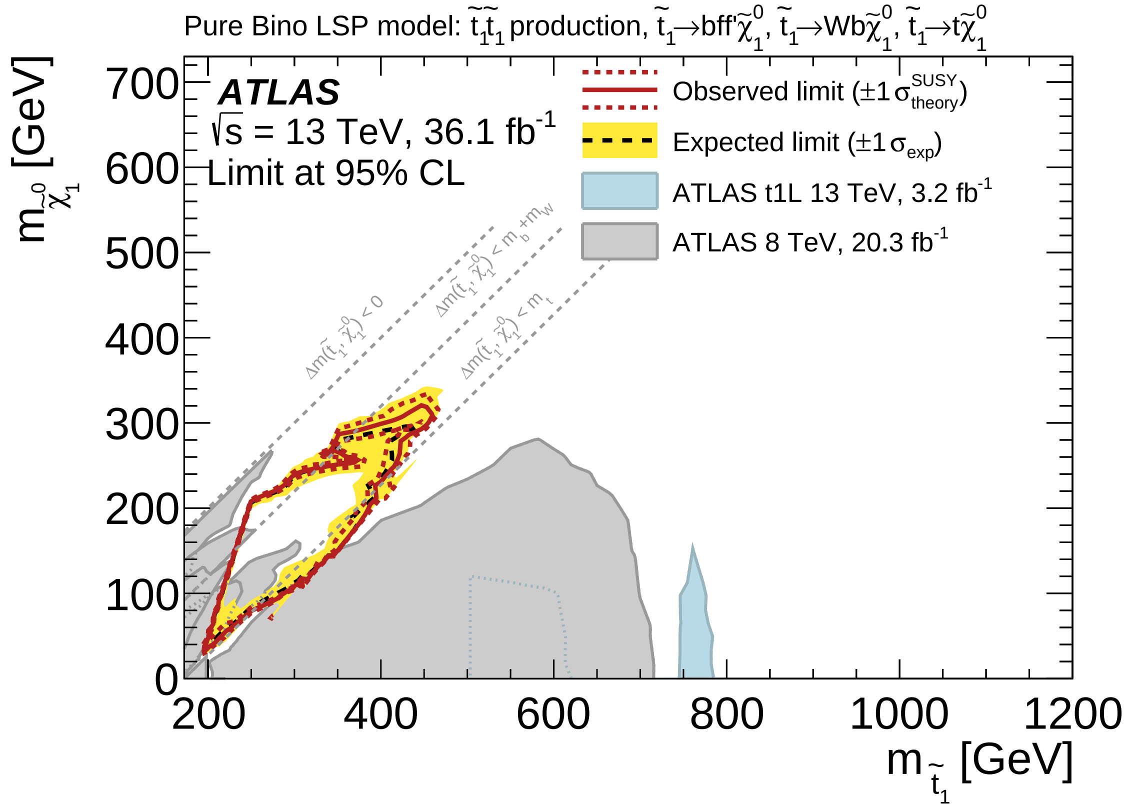


Figure 7.13: Expected (black dashed) and observed (red solid) exclusion limit at 95 % confidence level in the phase space spanned by $m(\tilde{\chi}_1^0)$ versus $m(\tilde{t}_1)$ for direct top squark pair production. Either $\tilde{t}_1 \rightarrow t\tilde{\chi}_1^0$, $\tilde{t}_1 \rightarrow bW\tilde{\chi}_1^0$, or $\tilde{t}_1 \rightarrow bff'\tilde{\chi}_1^0$ decay modes with a branching ratio of 100 % are assumed. The excluded regions from previous publications [55,173] are illustrated with the grey and blue shaded area, respectively.

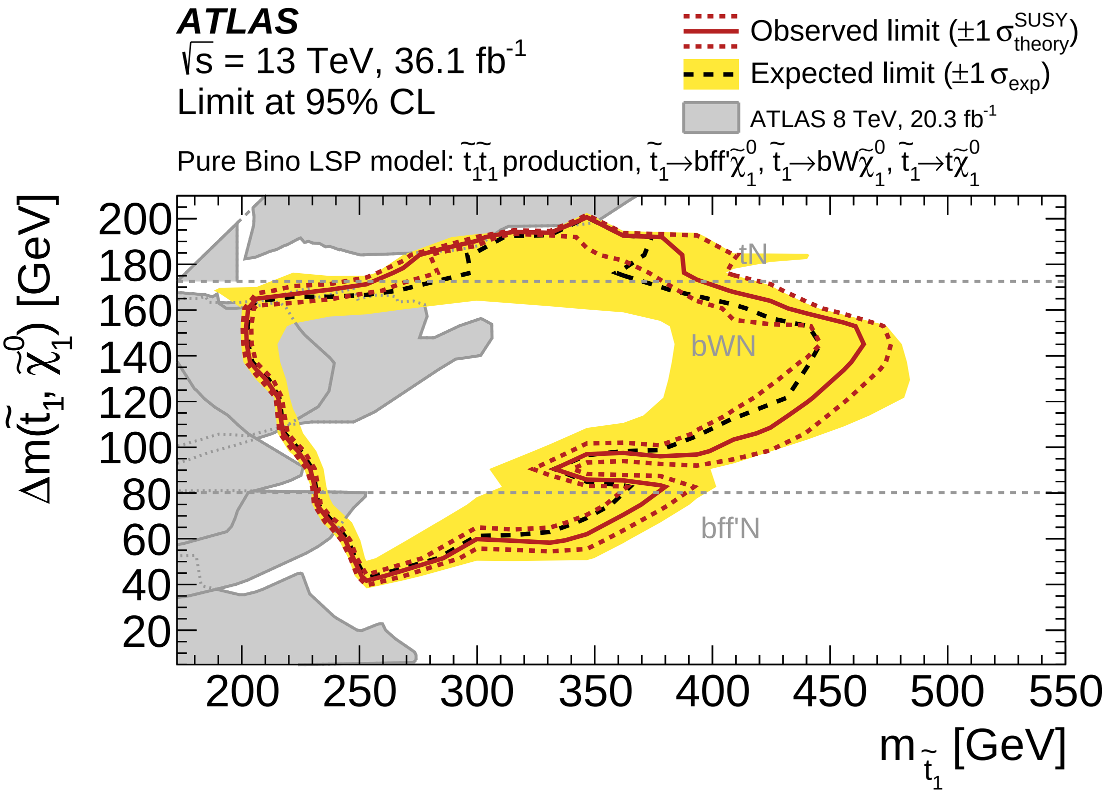


Figure 7.14: Expected (black dashed) and observed (red solid) 95% excluded regions in the plane of $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$ versus $m(\tilde{t}_1)$ for direct top squark pair production assuming either $\tilde{t}_1 \to t\tilde{\chi}_1^0$, $\tilde{t}_1 \to bW\tilde{\chi}_1^0$, or $\tilde{t}_1 \to bff'\tilde{\chi}_1^0$ decay with a branching ratio of 100 %. The excluded regions from previous publications [55, 173] are also illustrated with the grey shaded area.

Chapter 8

Search for top squark pair production in the 3-body decay mode with a single lepton final state using 139 fb^{-1} of pp collision data

In the following, a search for top squark pairs using the complete set of Run 2 pp collision data, recorded by the ATLAS experiment is presented. As in the previous chapter, the analysis targets the exact same signal model ($\tilde{t}_1 \to bW\tilde{\chi}_1^0$). This search forms the successor of analysis `bWN-subSet` and is further referred to as `bWN-fullSet`. Besides some similarities regarding analysis `bWN-subSet`, the major difference is, that a single discriminating variable, which is based on a machine learning classifier, is applied in order to separate the signal events from the dominant $t\bar{t}$ background. The presented results are also published by the ATLAS collaboration [197].

8.1 Discrimination between signal and background

The analysis strategy is described in Section 4.1. Again, observed results are statistically interpreted by a maximum likelihood fit. Therefore, a *discovery scenario* based on a single-bin signal region is defined in order to probe a potential data excess, and an *exclusion scenario*, comprising multiple bins, is defined to extract the signal exclusion in the absence of any evidence for new physics phenomena.

A significant change in this analysis compared to the previous analysis concerns the choice of the discriminating variables to define the signal region. Preliminary studies, performed on simulated events using a benchmark signal model with a larger $\tilde{t}_1$ mass, revealed that the discovery significance decreased by using the am_{T2} variable as the discriminating variable, even at an increased integrated luminosity. Possible reasons for this circumstance are that by definition am_{T2} has a lower boundary at 80 GeV, and since the signal selection was defined by an upper threshold in am_{T2} before, reducing the am_{T2} threshold reduces the signal efficiency as well. Furthermore, the top squark mass was raised for the benchmark signal model, which is used to define the signal region. The predicted cross section for $\tilde{t}_1$ pair production rapidly decreases with increasing $\tilde{t}_1$ mass. More precisely, the predicted cross section at $\sqrt{s} = 13$ TeV pp collisions at NNLO accuracy decreases from ~ 4.4 pb for $m_{\tilde{t}_1} = 350$ GeV down to ~ 1.1 pb for $m_{\tilde{t}_1} = 450$ GeV [118, 119, 141, 142], which obviously could not be compensated by the increase of integrated luminosity. As a consequence of the low signal sensitivity, the signal region definition in the following analysis is based on a machine learning approach.

Since the phase space of the target model has not changed, top quark pair processes still remain the dominant SM background. A dedicated control region is defined, to estimate the $t\bar{t}$ background in a data-driven manner and a validation region is defined to validate the extrapolation of the $t\bar{t}$ background normalisation from the CR to adjacent regions of phase space.

8.1.1 Event preselection

At first, a loose event preselection is defined, which depends mainly on the basic features of the decay products of the targeted SUSY signature ($\tilde{t}_1 \to bW\tilde{\chi}_1^0$). The modelling of important kinematic distributions is verified at preselection level. The subsequent definitions for the signal, control and validation region are also developed from simulated events that satisfy the preselection criteria.

The event preselection is based on the quality requirements for the physics objects as defined in Chapter 4 and is summarised in Table 8.1. Selected events are required to have exactly one electron or muon with a minimum p_{T} of at least 25 GeV. Moreover, no further baseline leptons are required in the sample. The dataset is recorded using $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers and an additional requirement of $E_{\mathrm{T}}^{\mathrm{miss}} > 230$ GeV ensures that the trigger is fully efficient. Furthermore, only events with a minimum of 4 jets, at which one of them is at least tagged as b-jet, are selected. The p_{T} threshold for those jets is $p_{\mathrm{T}} > 25$ GeV. In order to suppress a substantial contribution from QCD multi-jet events, the azimuthal angle between the direction of $E_{\mathrm{T}}^{\mathrm{miss}}$ and the two leading jets is required to be $|\Delta\phi(j_{1,2}, E_{\mathrm{T}}^{\mathrm{miss}})| > 0.4$. Events including hadronic τ candidates are vetoed by the requirement of $m_{\mathrm{T2}}^{\tau} > 80$ GeV. As already outlined and illustrated in Figure 4.6(b), the contributions from W+jets events and semi-leptonic $t\bar{t}$ processes are reduced by a requirement on the transverse mass, $m_{\mathrm{T}} > 110$ GeV.

Preselection	
Trigger	$E_{\mathrm{T}}^{\mathrm{miss}}$ trigger only
Second-lepton veto	no additional baseline leptons
Number of leptons	exactly 1 e or μ
Lepton p_{T}	> 25 GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 230 GeV
Number of jets (min. p_{T})	$\geq 4(25$ GeV)
Number of b-tagged jets	≥ 1
$\lvert\Delta\phi(j_{1,2}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})\rvert$	> 0.4
m_{T2}^{τ}	> 80 GeV
m_{T}	> 110 GeV

Table 8.1: Summary of the event preselection criteria.

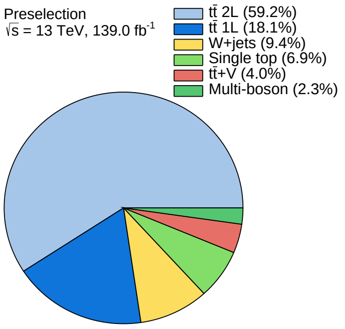
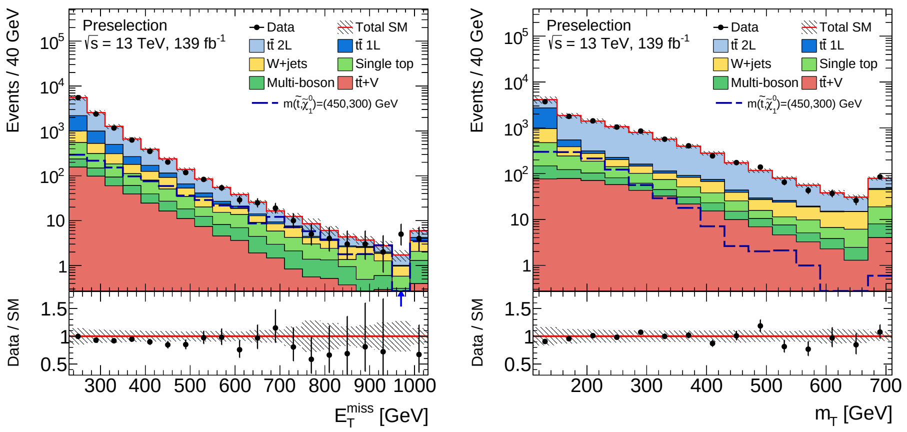


Figure 8.1: Relative contributions of the individual SM backgrounds after the preselection.

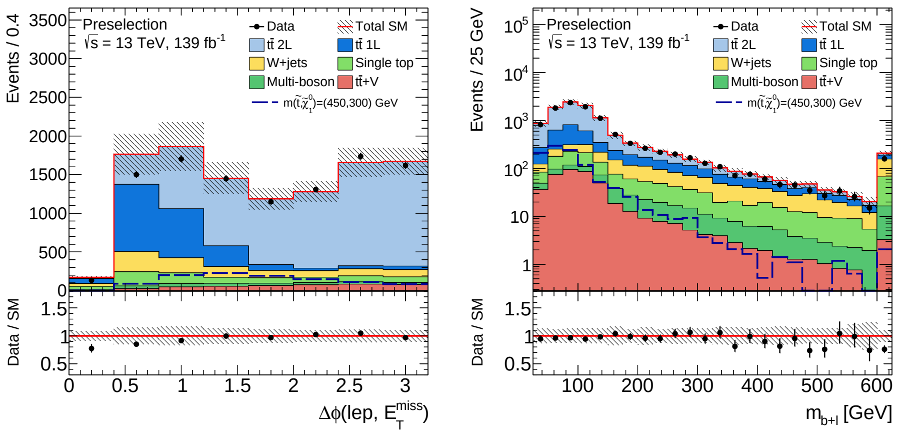
As indicated in Figure 8.1, $t\bar{t}$ production represents the dominant SM background after the preselection.

Distributions of kinematic observables at preselection level are illustrated in Figure 8.2. The $t\bar{t}$ background prediction in those distributions is scaled by the normalisation factor obtained from a likelihood fit of the CR, which is described in detail in Section 8.4, while all other SM background processes are normalised with the respective theoretical cross section. In general, good agreement between observed data and SM prediction is obtained. In addition, the expected distribution of the benchmark signal model is overlaid in order to stress the fact that the individual variables have low discriminating power. Initial studies of discriminating variables at preselection have been performed on simulated events only and the observed data was hidden until the signal region was defined in order to avoid a bias by the experimenter. Figure 8.2(b) indicates, that the contribution from semi-leptonic $t\bar{t}$ events saturates the region at lower values of m_{T}. Furthermore, it will be shown, that this essential feature is also learned by the ML classifier, so that the semi-leptonic $t\bar{t}$ background can be effectively suppressed in the analysis. Distributions of remaining variables, used to train the ML algorithm can be found in Section A.1 of the Appendix.



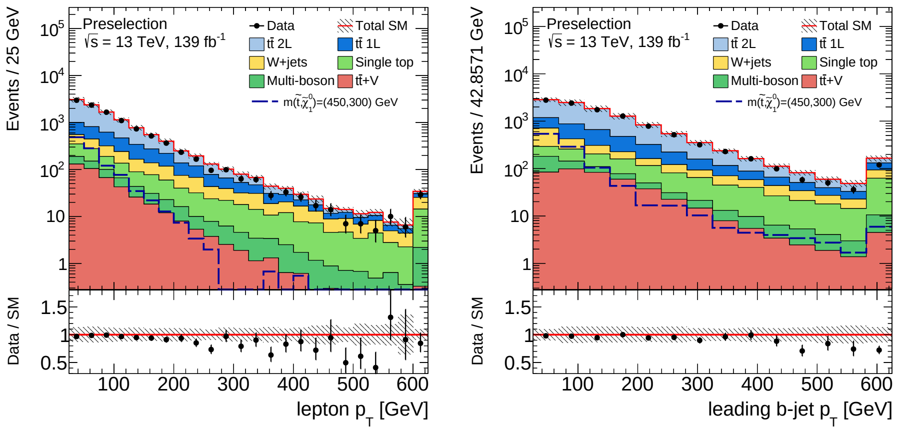
(a) $E_{\mathrm{T}}^{\mathrm{miss}}$

(b) m_{T}



(c) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$

(d) m_{bl}



(e) Lepton p_{T}

(f) Leading b-jet p_{T}

Figure 8.2: Kinematic distributions of key observables after the preselection is applied. The $t\bar{t}$ background prediction is scaled by the normalisation factor obtained from a likelihood fit of the CR, while residual SM background processes are normalised with the respective theoretical cross section. The hatched area around the total SM prediction and the hatched band in the ratio include statistical and systematic uncertainties, except for the $t\bar{t}$ modelling uncertainties, which are defined only for the signal and validation region. Additionally, the expected distribution of the SUSY benchmark signal model is overlaid (blue dashed line). Overflows are included in the last bin.

8.2 Signal extraction with machine learning techniques

In order to discriminate the SUSY signal events from the large amount of SM background processes, the SR selection is optimised by employing a machine learning approach. The goal of the ML classifier is to maximise the discovery sensitivity by learning the difference in the event topology between the signal and background. The ML classifier attempts to extract correlations amongst a set of kinematic distributions, which are considered as input variables for the training procedure. In order to perform the ML classification studies and the subsequent definition of the SR a signal benchmark model with $m_{\tilde{t}_1} = 450$ GeV and $m_{\tilde{\chi}_1^0} = 300$ GeV is chosen.

8.2.1 Truth-assisted ML training

As detailed in Chapter 5, to classify events as either signal-like or background-like from multi-variate data itself, the ML model needs to be trained on labelled events, and its output needs to be tested on an established mixture of signal and background events. The training and testing phase are performed on simulated events.

Any ML method requires sufficient statistics of the training sample to achieve a proper performance. However, generating the full simulation samples with adequate statistics is computationally expensive. As a consequence, signal events generated only at particle-level (Truth) were used for the training to enhance the signal statistics. Whereas, for the SM background processes, fully simulated and reconstructed events (Reco) were used with sufficient statistics.

In order to guarantee reliable results from the ML discriminator, the distributions of particle-level signal events should be well in agreement with those at fully reconstructed-level. To prevent possible deviations between particle-level events and detector-level events, the particle-level events were corrected to the reconstructed events using a dedicated smearing procedure.

Particle-level smearing procedure

Particle-level events are corrected by parametrisations for reconstruction and identification efficiencies, which are obtained based on dedicated ATLAS measurements. Furthermore, energies or transverse energies of physics objects at particle-level are smeared with a Gaussian term using a standard deviation corresponding to the energy (or E_{T}) resolution of the detector. Particle-level electrons, muons and jets are smeared according to their respective p_{T}, η as well as their applied identification working point. In addition, flavour tagging efficiencies are also applied on particle-level jets. Finally, the vector of the missing transverse energy is re-computed from all smeared particle-level objects, including also an approximation for the track soft term.

In the case of electrons, the identification efficiencies depend on p_{T}, η and on the applied reconstruction working point. The efficiency is determined in η bins of constant width, while a linear interpolation between adjacent values is used in p_{T} in the range from 7 GeV to 120 GeV. If the electron is beyond the given p_{T} range, the efficiency is associated to the respective upper or lower bound of the interval. The fraction of electrons which are mistakenly reconstructed from particle-level jets, depend on p_{T}, η and the applied electron isolation working point. The same interval for the interpolation in p_{T} is used for the isolation efficiency. The energy of the particle-level electron is smeared through a Gaussian term. The mean of the Gaussian distribution corresponds to the particle-level energy, while standard deviation corresponds to the energy resolution, which in turn depends on η and p_{T}.

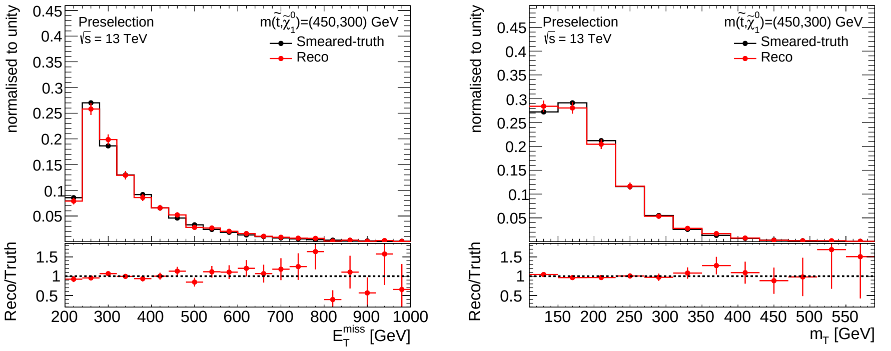
The identification efficiencies for particle-level muon candidates are binned in the η range. According to the applied identification working point, different values for the efficiencies are approximated. Likewise, the muon E_{T} is smeared by a Gaussian distribution. The two parameters of the Gaussian distribution are defined by the transverse energy and resolution of the particle-level muon candidate. Distinct approximations for the energy resolution are applied, depending whether the muon is located either in the barrel or the end-cap region.

A similar smearing based on a Gaussian term is also applied for the correction of the energy of particle-level jets. The approximation for the energy resolution relies on five equidistant η bins within the interval of $|\eta| \leq 4.5$. The p_{T} range for the smearing procedure of particle-level jets extends from 10 GeV to 1500 GeV.

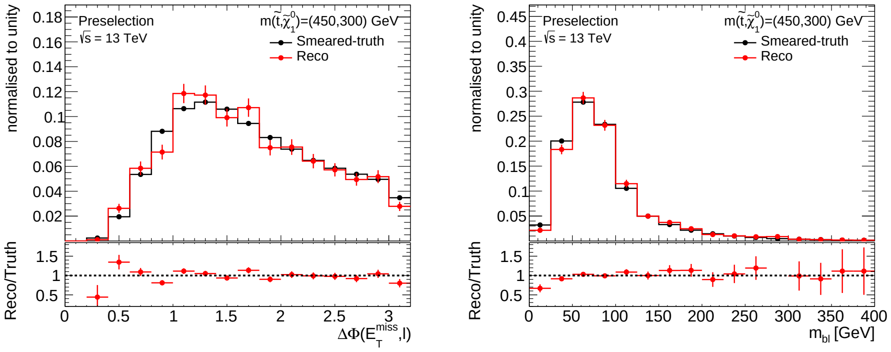
Approximations of flavour tagging efficiencies are also determined for particle-level jets with a transverse momentum above 20 GeV, and depend on the different flavour tagging working points. The corresponding approximations incorporate the various tagging efficiencies for light-, bottom-, charm- and tau-jets.

At last, the vector of the missing transverse momentum is derived from the transverse momenta of all smeared particle-level objects. An approximation for the track-based soft component, which constitutes energy remnants not associated to any physics object (see Section 4.3.7), is also included. Approximations for the soft term depend on the momentum balance between hard and soft components in an event, $p_{\mathrm{T}}^{\mathrm{hard}} = -p_{\mathrm{T}}^{\mathrm{soft}}$. Where $p_{\mathrm{T}}^{\mathrm{hard}}$ represents the sum of all smeared particle-level objects, and $p_{\mathrm{T}}^{\mathrm{soft}}$ is the track-based soft term. $p_{\mathrm{T}}^{\mathrm{hard}}$ is smeared by a Gaussian term, in which the standard deviation corresponds to the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. There are two components of the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution which are parallel and perpendicular aligned with respect to the $p_{\mathrm{T}}^{\mathrm{hard}}$. The approximated contribution for the track-based soft component is finally defined as the difference between the smeared and initial $p_{\mathrm{T}}^{\mathrm{hard}}$ component.

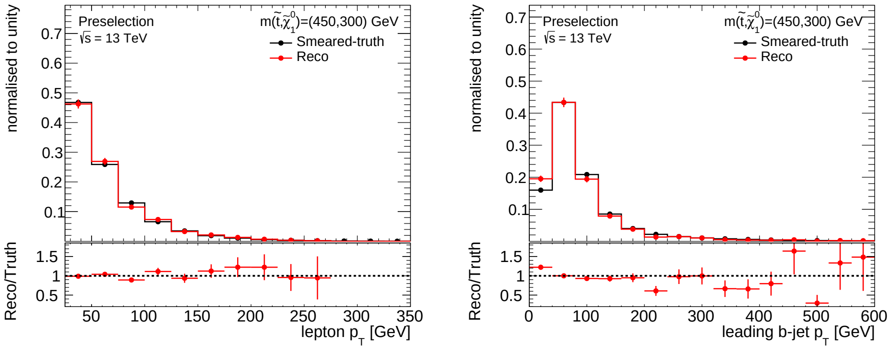
Figure 8.3 shows a comparison of important kinematic distributions between smeared particle-level signal events (black line) and fully reconstructed signal events (red line) after the preselection. These distributions nicely demonstrate the fact that no substantial differences between particle-level and detector-level events are observed. Kinematic distributions of all input variables at particle-level after smearing are found to be in agreement with the distributions at detector-level. Additionally, further distributions of input variables can be found in Section A.2 of the Appendix.



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) m_{T}



(c) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ (d) m_{bl}



(e) Lepton p_{T} (f) Leading b-jet p_{T}

Figure 8.3: Comparison of kinematic distributions between smeared particle-level signal events (“Smeared-truth”, shown as black dots on a black histogram) and fully reconstructed signal events (“Reco” shown as red dots on a red histogram) after preselection. Distributions are normalised to unity in order to investigate the shape of the distributions.

Network architecture and training phase

After applying the preselection on the smeared benchmark signal model, a total of about $\sim$ 200000 signal events were selected to perform the machine learning classification, accompanied by approximately $\sim$ 300000 fully reconstructed SM background events. In order to verify the result of the training procedure, the total statistics of signal and background events are splitted. 70 % of signal and background events are used to train the ML algorithm respectively, while 30 % of the total signal and background statistics are reserved as a test set.

According to the kinematical characteristics of the 3-body signal model (see Figure 2.4), two different types of jets play a crucial role in the analysis. On the one hand, a high energetic jet may recoil against the top squark pair system, and thus may induce a significant amount of $E_{\mathrm{T}}^{\mathrm{miss}}$. On the other hand, only low energetic jets may emerge from the decay of the top squark and the subsequent hadronically decaying W boson due to the compressed mass splitting of the SUSY parameters ($\Delta m = m_{\tilde{t}_1} - m_{\tilde{\chi}_1^0}$). However, the jet multiplicity in the signal events is not fixed. In order to process the variable length jet collection, the first stage of the ML architecture in this analysis is composed of a recurrent neural network. The reason is, because RNNs are known to extract information from sequences of arbitrary length, as already discussed in Section 5.2.2. Subsequently, the output vector of the jet-RNN is passed to a neural network in order to compute the signal and background probabilities at the output of the neural network. Further discriminating variables are applied as input neurons to the subsequent NN to further enhance the classification power. The conceptual architecture of the combination of the jet-RNN and the subsequent NN is illustrated in Figure 8.4.

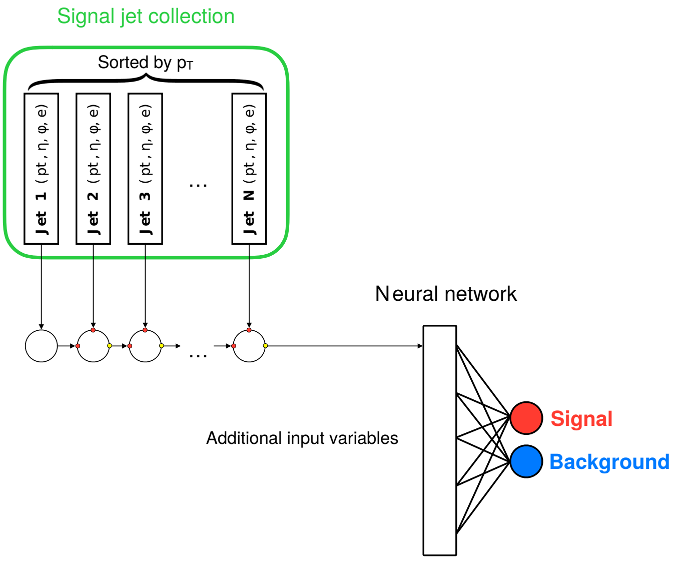


Figure 8.4: Conceptual layout of the machine learning architecture used to discriminate signal from SM background. In order to exploit the information of the whole jet collection, jet 4-vectors are processed in a recurrent neural network. The output vector of the jet-RNN is then passed as an input to the NN together with further discriminating variables.

Many different ML architectures have been tested and extensive parameter optimisation has been performed. The training parameters of the final ML classifier are defined as follows. The training is performed at preselection level. The 4-vectors (composed of p_{T}, η, ϕ and E) of all jets, which satisfy the dedicated jet reconstruction criteria as described in Section 4.3.4, are processed event-wise through a RNN using the aforementioned LSTM algorithm. The LSTM transforms the variable length sequence of jets into a 32 dimensional output vector, corresponding to a maximum of 8 jets output considered. The LSTM is then connected to the input layer of a NN together with 12 discriminating variables. A list of the 12 additional input variables used to train the neural network is shown in Table 8.2. The neural network is comprised of a single hidden layer with 128 neurons which are connected to two output neurons. The two output neurons represent the probabilities of the signal and background, respectively. The framework to train the ML classifier to discriminate between SM background processes and SUSY signal events is based on the `keras` [198] package using `tensorflow` [199] as a backend. During the training, the weight parameters are optimised by the Adam optimiser [185] with a learning rate of 10^{-3}. A leaky rectified linear unit with a slope coefficient of $\epsilon = 0.1$ was used as activation function in the hidden layer. A softmax activation function is used in the output layer to ensure that the output scores of the two output neurons sum up to unity. In addition, a few more parameters are applied to the NN, which may not necessarily improve the performance of the classifier, but rather accelerate the training procedure or prevent the network from overtraining. A small $L2$ regularisation term of $\lambda = 10^{-2}$ is applied in the NN. Regularisation is a common method to prevent overtraining by modifying the weights during the training phase. In order to enhance the training speed, the weight parameters corresponding to the hidden layer are initialised to a normal distribution [200] and the training was performed in mini-batches of 32 events each, with batch normalisation [201] applied. The parameters of the final neural network are summarised in Table 8.3.

Notable to mention, the jet-RNN is not trained separately before its output vector is forwarded to the neural network, the jet-RNN and NN are trained simultaneously in one step instead.

Input variables	Description
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse energy
$\phi(E_{\mathrm{T}}^{\mathrm{miss}})$	Azimuthal angle of the direction of $E_{\mathrm{T}}^{\mathrm{miss}}$
m_{T}	Transverse mass
$\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$	Azimuthal angle between the direction $E_{\mathrm{T}}^{\mathrm{miss}}$ and the lepton
m_{bl}	Invariant mass of the leading b-tagged jet and the lepton
$p_{\mathrm{T}}^{b-\mathrm{jet}}$	Transverse momentum of the leading b-tagged jet
n_{jet}	Jet multiplicity
$n_{b-\mathrm{jet}}$	b-jet multiplicity
$p_{\mathrm{T}}(\ell)$	Transverse momentum of the lepton
$\eta(\ell)$	Pseupdorapidity of the lepton
$\phi(\ell)$	Azimuthal angle of the lepton
$E(\ell)$	Energy of the lepton

Table 8.2: Discriminating variables applied as inputs to the NN. In addition, the output vector of the jet-RNN to process the arbitrary-length jet collection is also applied as an input.

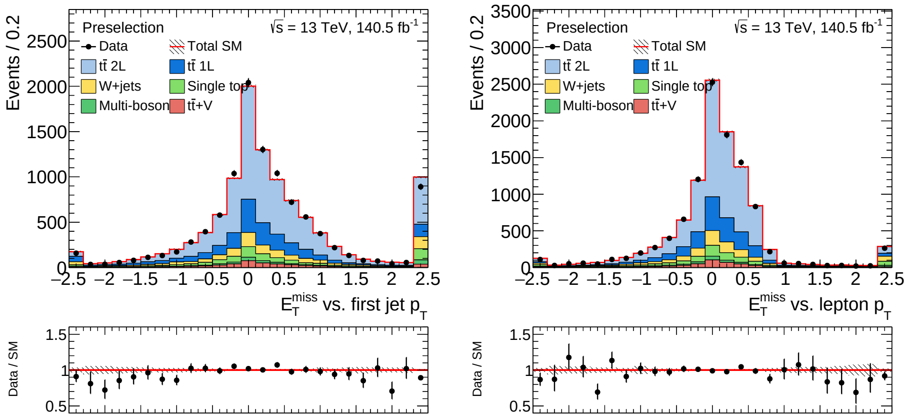
Architecture	Parameter set
Number of hidden layers	1
Neurons per hidden layer	128
Activation function	leaky ReLU ($\epsilon = 0.1$)
Learning rate	10^{-3}
Regularisation	$L2$ ($\lambda = 10^{-2}$)
Weight initialisation	Glorot normal
Batch size	32
Batch normalisation	Yes

Table 8.3: Architecture and parameters of the neural network.

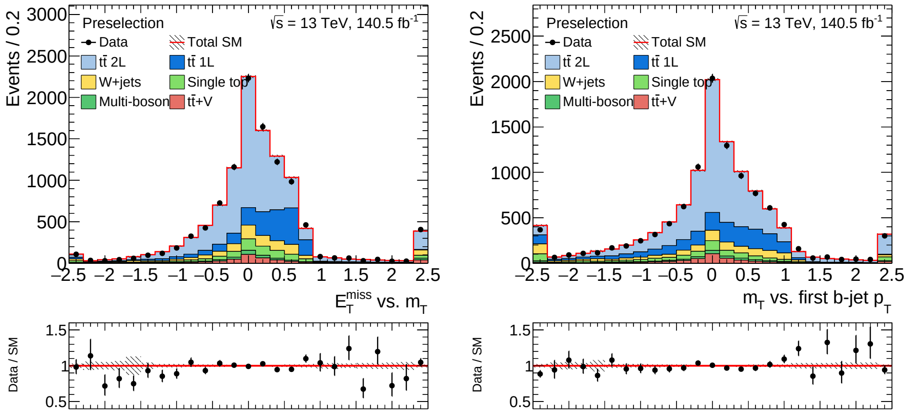
Performance metrics of the ML classifier

After the NN is successfully trained, several measures are performed to quantify the performance of the network.

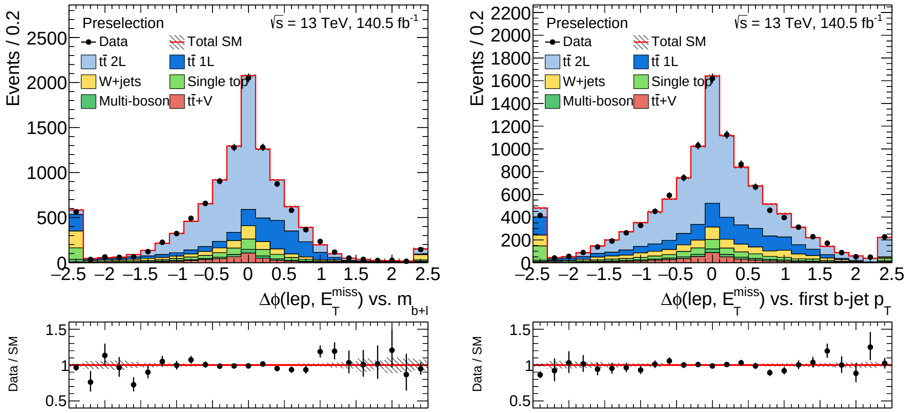
At first, the linear correlation between the input variables can be examined, even before the training procedure is accomplished. In order to compare the linear correlations of the inputs for observed data as well as simulated events, the quantity $(x - \langle x \rangle) \cdot (y - \langle y \rangle)/(\text{RMS}_x \cdot \text{RMS}_y)$ is computed. The corresponding distributions are given in Figure 8.5. Further correlation distributions of the various input variables are shown in Section A.3 of the Appendix. Despite some fluctuations in statistically limited bins, the observed events are well in agreement with the SM expectation.



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ vs. leading jet p_{T} (b) $E_{\mathrm{T}}^{\mathrm{miss}}$ vs. lepton p_{T}



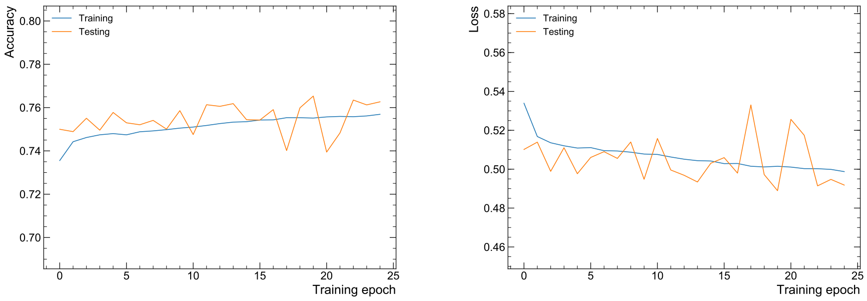
(c) $E_{\mathrm{T}}^{\mathrm{miss}}$ vs. m_{T} (d) m_{T} vs. leading b-jet p_{T}



(e) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ vs. m_{bl} (f) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ vs. leading b-jet p_{T}

Figure 8.5: Distributions of the linear correlation between two kinematic observables, respectively. Agreement between data and simulated background events is observed. Only statistical uncertainties are shown. Underflows and overflows are included in the first and last bins.

After the training phase is finished, the NN output score is evaluated for training events and independent test events respectively, in order to verify the classification power of the network. Figure 8.6(a) shows the accuracy metric, as a function of the number of training cycles, as discussed in Section 5.4. It can be seen that the accuracy, and thus the overall performance of the NN, increases until the training stopped after 25 training epochs. The larger fluctuations of the test accuracy is due to the smaller statistics in the test set. The decrease of the loss function of the ML classifier with each training epoch is illustrated in Figure 8.6(b). In both figures, the training and the test set show a similar behaviour, which indicates that the network does not tend to overtrain and does not learn from statistical fluctuations of single events included in the training sample. Given the progression of the loss function, it could be argued that even more training epochs may further increase the performance. Although, a stopping algorithm was implemented such that if the loss is not significantly improved after 5 consecutive training cycles, the training procedure is stopped. It turned out that with the aforementioned configuration of the ML classifier, a good performance is achieved in the whole phase space of 3-body decay mode, even if it was trained on a benchmark signal point ($\Delta m = 150$ GeV). In-depth optimisation on this signal point could have lead to lower sensitivities amongst signal points in the phase space with varying Δm.



(a) Accuracy (b) Loss

Figure 8.6: Distribution of the accuracy and loss of the NN as a function of the number of training epochs.

Figure 8.7 shows the distribution of the neuron of the output score of the ML classifier that represents the signal probabilities, referred to as NN_{bWN}. The distributions are separated into signal and background events, which either belong to the training or test sample. The respective distributions for SUSY signal events and SM background processes are well in agreement, which indicates that the learned attributes from the training set are valid on an independent sample as well. The distributions are normalised to unity in order to investigate any shape differences. In general, a good separation power between background and signal events is achieved.

The performance of the classification problem is further quantified by the receiver operating characteristic illustrated in Figure 8.8. The ROC curves are determined for the training and test set, respectively. The similarities of both curves indicate once again that the NN is not overtrained. The AUC values in the legend represent the area under ROC, which gives a measure of the overall performance of the classification problem.

The next metric discussed are the confusion matrices of the training and test set respectively, illustrated in Figure 8.9. Values close to unity in the diagonal term signify good performance of the classifier, while off-diagonal terms indicate the fraction of mis-classified events. The dataset is again split into events used in the training and test phases to analyse possible overtraining. The values in the diagonal and off-diagonal terms in both matrices

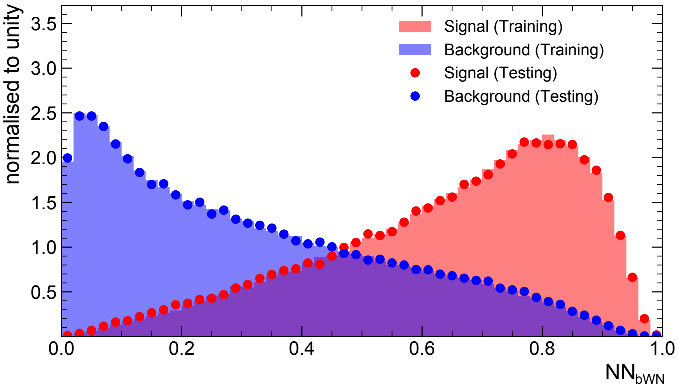


Figure 8.7: Distribution of the ML output score, referred to as $\mathrm{NN}_{\mathrm{bWN}}$. The events are separated into signal and background used for training and those used for testing, respectively. Distributions of training and test events are in good agreement, no overtraining is observed. Distributions are normalised to unity.

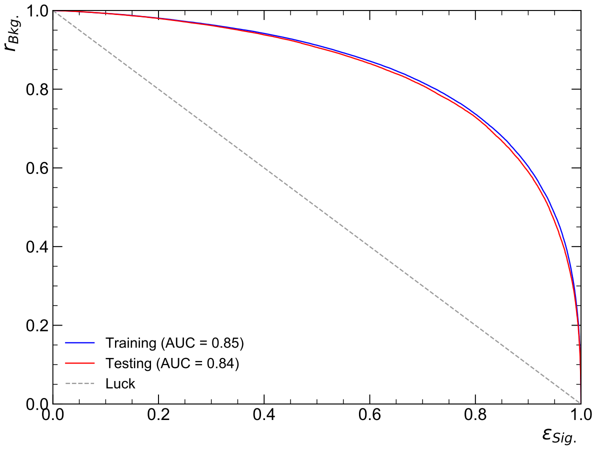
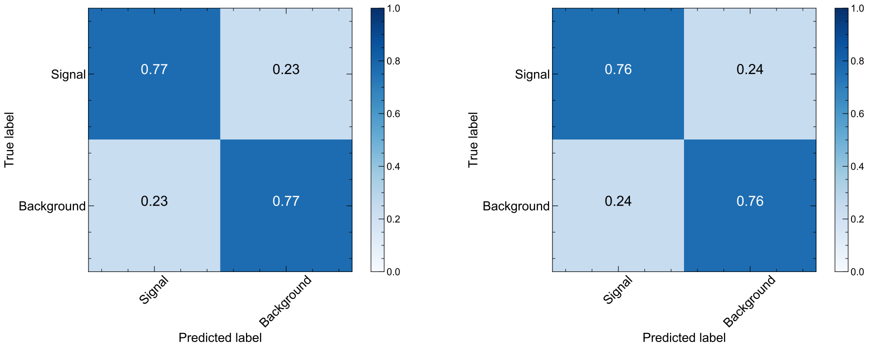


Figure 8.8: Receiver operating characteristic of the ML classifier. The events are separated into training and test set. Both curves are in good agreement, which indicates that the NN is not overtrained. The quantity in the legend denotes the area under the curve (AUC) and is a metric for the overall performance of the classification problem.

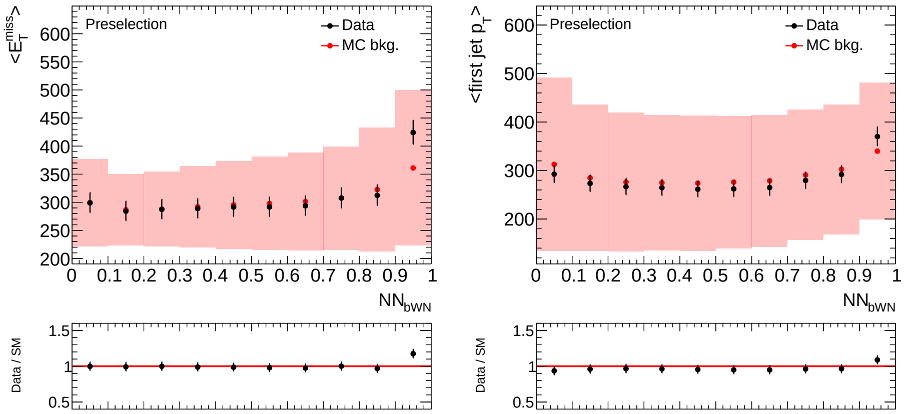
are in agreement.

Finally, the dependence of each input variable as a function of the ML output score is investigated. Figure 8.10 shows the mean value of the respective input variable as a function of $\mathrm{NN}_{\mathrm{bWN}}$ for SM background events and observed data, respectively. Overall, a good agreement between observed and simulated SM background processes is observed. In addition, the uncertainties represent the standard deviation (RMS) of the distribution in the corresponding bin. The dependence of the sensitive variable with respect to the ML output score can be nicely seen. Events with large $E_{\mathrm{T}}^{\mathrm{miss}}$ and a high leading jet p_{T} tend to accumulate the region with larger values of $\mathrm{NN}_{\mathrm{bWN}}$ (see Figures 8.10(a) and 8.10(b)), whereas events with a high $\mathrm{NN}_{\mathrm{bWN}}$ score also tend to have softer transverse momenta of further reconstructed objects such as the lepton, the leading b-jet or further light jets (see Figures 8.10(c), 8.10(d), 8.10(e) and 8.10(f)). Further profile distributions of the various input variables are shown in Section A.4 of the Appendix.

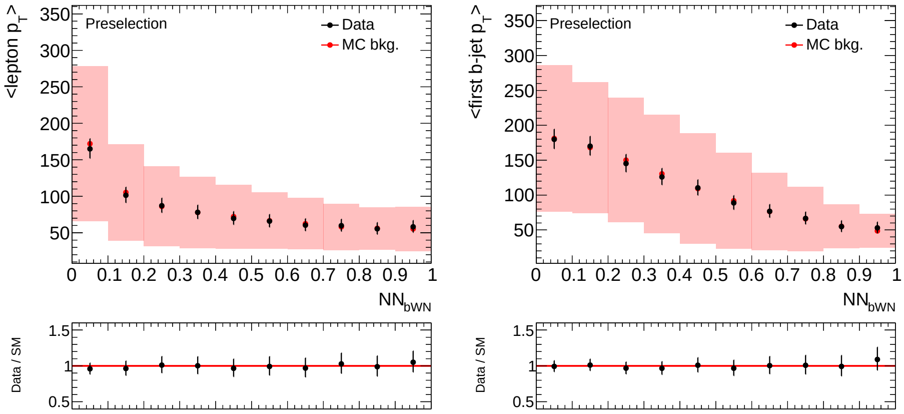


(a) Training set (b) Test set

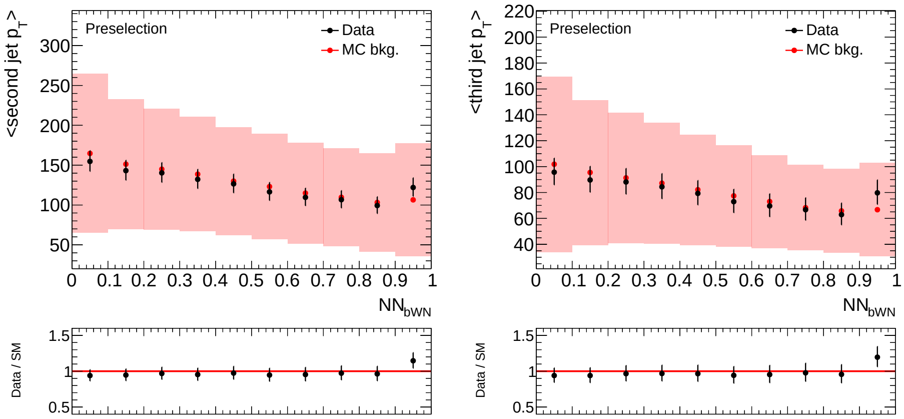
Figure 8.9: Confusion matrices of training set and test set. The values in the diagonal and off-diagonal terms represent the fraction of correctly and misidentified events of both classes.



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ profile (b) Leading jet p_{T} profile



(c) Lepton p_{T} profile (d) Leading b-jet p_{T} profile



(e) 2nd jet p_{T} profile (f) 3rd jet p_{T} profile

Figure 8.10: Distribution of the mean of the sensitive input variable as a function of $\mathrm{NN}_{\mathrm{bWN}}$ for data and simulated SM events. Good agreement between observed and simulated events is observed. The error bands represent the standard deviation (RMS) of the respective variable in the corresponding bin.

Since smeared particle-level signal events are applied in the training phase, a comparison of the shape of the $\mathrm{NN_{bWN}}$ distribution between smeared particle-level and fully reconstructed signal events is of particular interest. The distribution of $\mathrm{NN_{bWN}}$ is shown in Figure 8.11. Both distributions agree well within uncertainty. This figure indicates that no substantial differences between particle-level and detector-level events are induced during the learning procedure.

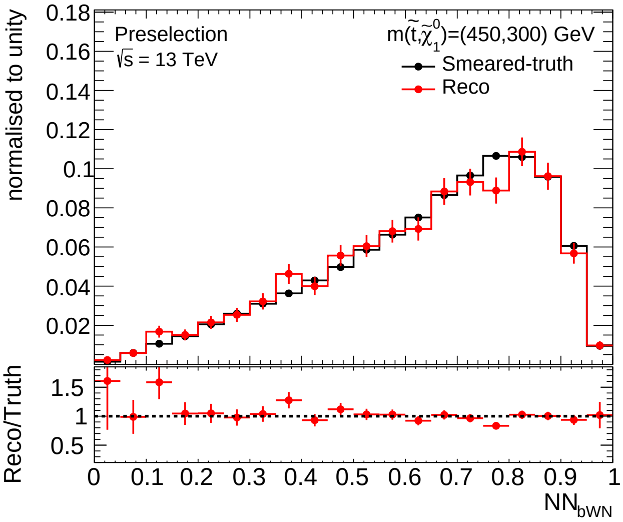


Figure 8.11: Comparison of the distribution of the ML output score between particle-level and reconstructed events of the benchmark signal model. Both distributions agree well within uncertainty. In order to investigate differences of the shapes of the distributions, the distributions are normalised to unity.

Comparison with alternative analysis techniques

In addition, the performance of the current ML-based analysis technique is also compared to alternative cut-based methods as well as concurrent multi-variate techniques. The methodology for this study is as follows. Three cut-based approaches are investigated. For those analyses, a signal region selection is defined, based on a set of discriminating variables. Moreover, four alternative multi-variate techniques are trained with a particular set of input variables and the dedicated signal region selection is defined, based on the output score of the corresponding multi-variate technique. The expected discovery significance Z[1] in the respective signal region is chosen as the figure of merit for the comparison of the concurrent analysis techniques. An additional 25 % relative uncertainty on the total SM background, supplementary to the statistical uncertainty is considered in the calculation of the significance. In the following, the various analysis techniques are briefly described:

- `bWN-subSet` refers to the exact same signal region selection as it is discussed in Chapter 7 with the only difference, that the simulated events are normalised to 139 fb^{-1}.

---

[1] The significance Z of observing n events for a prediction of $b \pm \sigma$ is computed following the profile likelihood method in [202]:
$Z = \sqrt{2(n \ln[\frac{n(b+\sigma^2)}{b^2+n\sigma^2}] - \frac{b^2}{\sigma^2}\ln[1 + \frac{\sigma^2(n-b)}{b(b+\sigma^2)}])}$ when $n \geq b$, or
$Z = -\sqrt{2(n \ln[\frac{n(b+\sigma^2)}{b^2+n\sigma^2}] - \frac{b^2}{\sigma^2}\ln[1 + \frac{\sigma^2(n-b)}{b(b+\sigma^2)}])}$ when $n < b$.

- `bWN-subSet (re-optimised)` uses the set of discriminating variables as `bWN-subSet`, but for the signal region definition the dedicated thresholds have been optimised for the larger integrated luminosity.
- `Cut-based` uses the same 12 input variables as described above for the actual NN approach (see Table 8.2), but the signal region is based on particular cuts on the different variables.
- `NN (alt. input)` refers to a neural network which has been optimised and trained with a different set of input variables.
- `NN incl. 4 jets` represents again a neural network. In addition to the 12 input variables listed in Table 8.2, the 4-vectors of the four leading jets are applied as input variables, and as a consequence the jet-RNN is omitted.
- `BDT` defines a signal region based the on the output score of a boosted decision tree, which is trained and optimised with the same 12 input variables as the final $\mathrm{NN}_{\mathrm{bWN}}$.
- `NN w/o jet-RNN` refers to a SR which is based on a neural network using again the 12 main input variables, but without the jet-RNN applied.
- `NN` represents the final ML method.

Further details on the alternative methods can be found the Appendix A.5. The expected number of SM background and SUSY signal events, as well as the expected discovery significance is listed in Table 8.4. For multi-variate analysis techniques, the same truth-smeared benchmark signal model as the final $\mathrm{NN}_{\mathrm{bWN}}$ is used to train the respective algorithm, whereas for the cut-based methods, fully reconstructed events of the same signal point are applied. However, the expected number of events from a signal model with slightly different kinematics ($m(\tilde{t}_1, \tilde{\chi}_1^0) = (500, 380)$ GeV) are listed in the Table, in order to stress the generalisation of the respective analysis technique in the 3-body phase space. The reason is, because, some of the analyses may isolate the considered signal model very well, but could do worse on adjacent signal points in the 3-body phase space due to slightly different kinematics. The final $\mathrm{NN}_{\mathrm{bWN}}$ approach gives the largest significance compared to other analysis techniques. In addition, the performances of all considered methods are illustrated in Figure 8.12 in terms of signal efficiency ($\epsilon_{\mathrm{Sig.}}$) and background rejection ($r_{\mathrm{Bkg.}}$). The solid lines correspond to the ROC curves of the dedicated multi-variate technique, while the points represent the corresponding signal region selection.

140 fb^{-1}	`bWN-subSet`	`bWN-subSet(re-optimised)`	`Cut-based`	`NN (alt. input)`
Total SM	257.04 ± 2.88	77.66 ± 1.41	22.45 ± 0.99	20.89 ± 0.68
$m(\tilde{t}_1,\ \tilde{\chi}_1^0) = (500, 380)$	54.83 ± 3.21	27.70 ± 2.27	31.31 ± 2.44	31.87 ± 3.97
expected Z	0.62	1.02	3.10	3.31
	`NN (incl. 4 jets)`	`BDT`	`NN (w/o jet-RNN)`	`NN`
Total SM	6.04 ± 0.14	5.22 ± 0.16	5.57 ± 0.17	4.56 ± 0.09
$m(\tilde{t}_1,\ \tilde{\chi}_1^0) = (500, 380)$	13.88 ± 1.76	13.69 ± 1.79	14.88 ± 2.10	16.21 ± 2.56
expected Z	3.33	3.49	3.65	4.27

Table 8.4: Comparison of the discovery significance of $\mathrm{NN}_{\mathrm{bWN}}$ with alternative cut-based and multi-variate approaches.

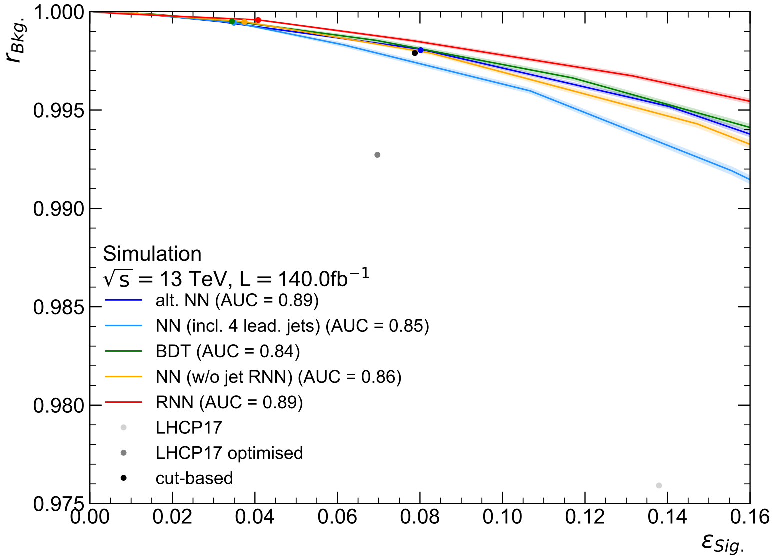


Figure 8.12: Illustration of the ROC curves and working points, associated with a signal region, of the primary ML classifier and various alternative methods.

8.3 Signal region definition

After training the ML classifier and validating its performance metrics, it is evaluated on all simulated SM background processes as well as on SUSY signal events. Figure 8.13 represents the distribution of the classifier output score. In addition to the distribution of the benchmark signal point, another signal model is shown, which has slightly different kinematics, in order to indicate the universal sensitivity of $\mathrm{NN_{bWN}}$ in the signal phase space. As expected the SM background processes dominate at low values, in contrast to the SUSY signal models which saturate at large $\mathrm{NN_{bWN}}$ scores. As a consequence, the signal region is defined by a stringent requirement on $\mathrm{NN_{bWN}}$.

The exact event selection for the single-bin SR as well as for the shape-fit configuration is listed in Table 8.5.

Scenario	Signal region definition and its binning
Discovery	$\mathrm{NN_{bWN}} > 0.9$
Exclusion	$\mathrm{NN_{bWN}} \in [0.65^*, 0.7^*, 0.75^*, 0.8, 0.82, 0.84, 0.86, 0.88, 0.9, 0.92, 1]$

Table 8.5: Overview of the signal region definition. A single-bin SR is defined for a potential data excess (Discovery), while a shape-fit is used to enhance the exclusion sensitivity (Exclusion). Square brackets indicate the interval in the $\mathrm{NN_{bWN}}$ score. An additional requirement of $m_{\mathrm{T}} > 150$ GeV is applied to bins marked by *, in order to suppress potential contamination of semi-leptonic $t\bar{t}$ events.

A single cut is placed above 0.9 of $\mathrm{NN_{bWN}}$, for a possible discovery scenario. Table 8.6 lists the expected events of the respective SM background process as well as a potential signal model. The expected number of the given signal model does not exactly match the number of expected signal events which is listed in Table 8.4, because the requirement on $\mathrm{NN_{bWN}}$ in the study above was tighter. However, it was decided to loosen the requirement on $\mathrm{NN_{bWN}}$ a little bit in order to gain sensitivity across the full phase space of the 3-

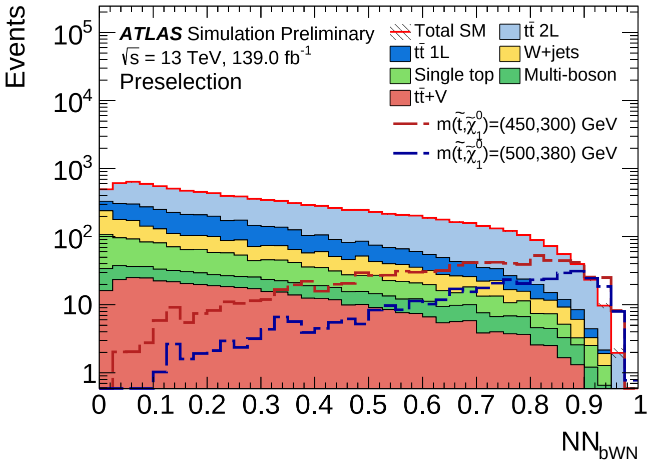


Figure 8.13: Distribution of the trained ML output score after the preselection. Values close to 1 indicate that the classified event is more signal-like while small values determine events as background-like. The $t\bar{t}$ background is not yet normalised in the CR and only statistical uncertainties are illustrated by the hatched band.

body decay. In addition, Figure 8.14 shows the relative contribution of the individual SM background processes.

Signal region	Events
$t\bar{t}$ 2L	28.2 ± 0.9
$t\bar{t}$ 1L	1.41± 0.25
Single top	1.9 ± 0.7
W+jets	1.50 ± 0.6
$t\bar{t}$+V	0.85 ± 0.18
Multi-boson	1.24 ± 0.11
Total SM	35.1 ± 1.30
$m(\tilde{t},\tilde{\chi}_1^0) = (500, 380)$	52.5 ± 3.1

Table 8.6: Expected simulated events in the signal region for 139 fb^{-1}. Only the statistical uncertainties are given.

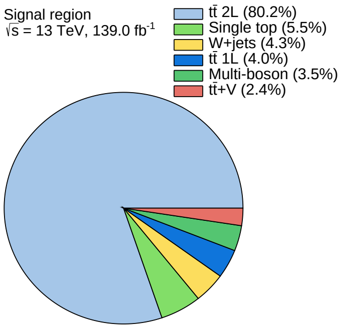


Figure 8.14: Breakdown of the individual SM contributions to the signal region.

To test the significance across the full phase space of the 3-body decay mode, the set of considered signal points is also extended by additional signal points from the adjacent 2-body and 4-body phase space. The expected significance in the phase space characterised by 20 GeV < Δm < 200 GeV is shown in Figure 8.15. The black contour line indicates the threshold where the expected significance reaches $Z = 3$. Only the statistical uncertainty of the background and a 25 % relative uncertainty on the absolute SM background is considered in the computation of Z. A universal sensitivity across the full signal phase space is achieved by the ML classifier.

In order to increase the potential exclusion sensitivity in the absence of any data excess compared to the SM prediction, the shape of the NN$_{\mathrm{bWN}}$ distribution is exploited in the shape-fit while expanding bins to lower NN$_{\mathrm{bWN}}$ than the single-bin signal region. A dedicated binning in NN$_{\mathrm{bWN}}$ is defined within the interval from 0.65 to 1.0 in the classifier output score. Figure 8.16 shows the m_{T} distributions for the fraction of events within NN$_{\mathrm{bWN}} \in [0.65, 0.8]$ and above NN$_{\mathrm{bWN}} > 0.8$, respectively. A significant amount of semi-

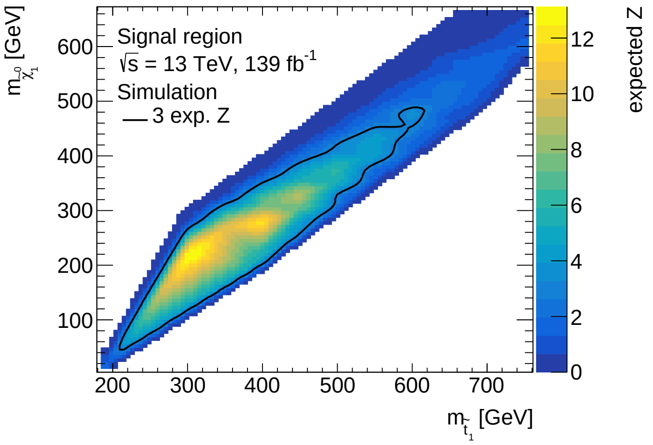
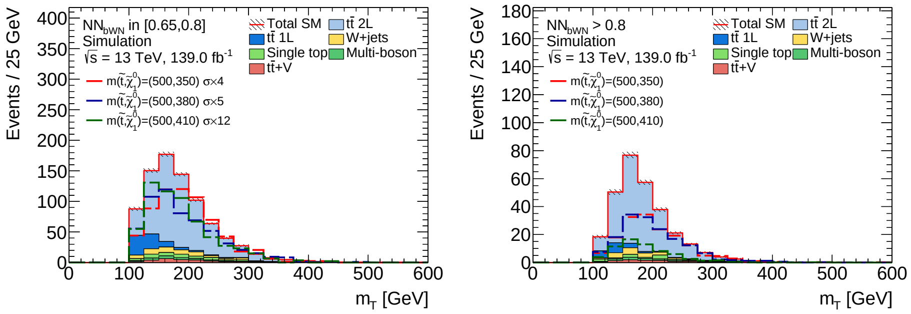


Figure 8.15: Expected significance across the signal phase space spanned by $m(\tilde{t}_1)$ versus $m(\tilde{\chi}^0_1)$. The considered phase space spans over signal points characterised by 20 GeV < Δm < 200 GeV. The black contour line illustrates the threshold for $Z > 3$.

leptonic $t\bar{t}$ events is located at m_{T} < 150 GeV in the phase space with moderate ML classifier output scores (Figure 8.16(a)), while this background contribution is mainly suppressed for higher values of $\mathrm{NN}_{\mathrm{bWN}}$. This indicates, that the classifier successfully learned, that SUSY signal events tend to have larger values of m_{T}. However, for the three lowest bins in the shape-fit configuration, an additional requirement of $m_{\mathrm{T}} > 150$ GeV is applied, in order to keep the composition of the different SM contributions similar across the various bins.



(a) m_{T} distribution in $\mathrm{NN}_{\mathrm{bWN}} \in [0.65, 0.8]$ (b) m_{T} distribution in $\mathrm{NN}_{\mathrm{bWN}} > 0.8$

Figure 8.16: Distribution of m_{T} for simulated SM backgrounds as well as for three different signal points in the interval of $\mathrm{NN}_{\mathrm{bWN}} \in [0.65, 0.8)$ (a) and $\mathrm{NN}_{\mathrm{bWN}} \geq 0.8$ (b). At moderate values of the output scores (a), a significant amount of semi-leptonic $t\bar{t}$ processes saturates at m_{T} < 150 GeV, whereas such events are suppressed in the region characterised by large values of $\mathrm{NN}_{\mathrm{bWN}}$ (b).

A signal region composed of multiple bins can further constrain the SM prediction in the fit, if the the ratios of the SUSY signal to the SM background are distinct in the various bins. In order to test the impact of the number of bins contributing to the shape-fit, several exclusion fits for two different signal points has been performed, in which the number of bins has been increased, respectively. Figure 8.17 shows the expected significance Z for rejecting the background+signal hypothesis determined from exclusion fits for two different signal points as a function of the number of bins included in the fit procedure and indicates that the Z significantly increases with the number of bins.

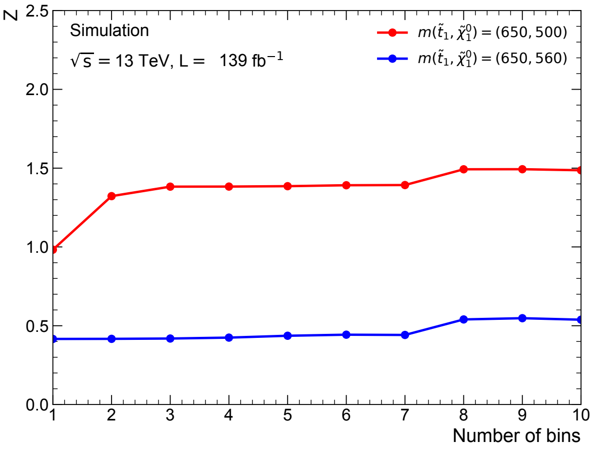


Figure 8.17: Expected significance Z for rejecting the background+signal hypothesis for two different signal points as a function of bins included to the maximum likelihood fit. Z significantly increases if multiple bins are added to the fit procedure. Proper systematic uncertainties accounting for $t\bar{t}$ radiation, fragmentation and hard scattering are applied respectively, as well as systematic uncertainties for flavor tagging, lepton identification, isolation and reconstruction, and pileup are included. $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainties as well as uncertainties on jet energy scales and resolutions are also applied. Signal acceptance uncertainties and uncertainties on the cross section of the minor SM backgrounds are included as well.

Figure 8.16 shows the distribution of $\mathrm{NN_{bWN}}$ for simulated SM backgrounds as well as for three different signal points. The chosen binning is identical as defined in the shape-fit configuration. Emphasis is on the distinct shapes of $\mathrm{NN_{bWN}}$ for SM processes and signal events. In addition, the total amount of simulated SM background processes and a hypothetical supersymmetric signal in the corresponding bins are listed in Table 8.7.

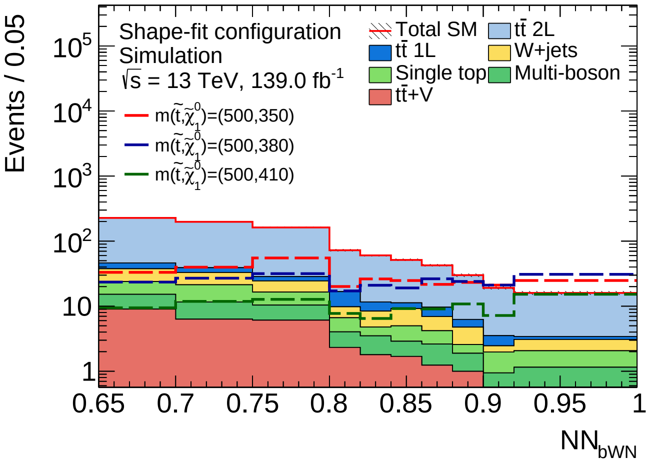


Figure 8.18: Distribution of $\mathrm{NN_{bWN}}$ in the range from 0.65 to 1.0 for SM background as well as for signal points with three different mass splittings. An additional requirement on $m_{\mathrm{T}} > 150$ GeV is applied on the first three bins ranging from 0.65 to 0.8. Only the statistical uncertainty of the simulated events is shown in the shaded area around the total SM expectation. The $t\bar{t}$ background is not yet normalised to the CR.

Shape-fit	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
$t\bar{t}$	189.93± 3.09	165.21± 2.79	137.88± 2.49	62.67± 1.70	52.02± 1.38
Single top	8.71± 1.49	9.74± 1.59	6.10± 1.25	2.61± 0.80	1.28± 0.52
W+jets	13.70± 1.82	11.64± 1.71	7.97± 1.12	3.15± 0.87	3.63± 0.84
$t\bar{t}$+V	9.02± 0.48	6.32± 0.44	6.32± 0.39	2.32± 0.24	1.80± 0.23
Multi-boson	6.20± 0.37	5.29± 0.63	4.30± 0.25	1.71± 0.13	1.71± 0.15
Total SM	227.56± 3.93	198.19± 3.72	162.34± 3.04	72.47± 2.09	60.43± 1.72
$m(\tilde{t}, \tilde{\chi}_1 0) = (500, 380)$	23.49± 2.21	26.94± 2.30	31.68± 2.51	17.20± 1.82	20.98± 2.03
	Bin 6	Bin 7	Bin 8	Bin 9	Bin 10
$t\bar{t}$	42.19± 1.28	35.52± 1.24	25.25± 0.90	16.67± 0.64	12.91± 0.65
Single top	2.10± 0.74	1.58± 0.63	0.69± 0.40	1.03± 0.52	0.91± 0.48
W+jets	4.22± 1.03	2.75± 0.87	2.19± 0.99	0.48± 0.20	1.02± 0.53
$t\bar{t}$+V	1.69± 0.22	1.24± 0.18	1.00± 0.16	0.37± 0.13	0.48± 0.12
Multi-boson	1.20± 0.13	1.38± 0.17	0.89± 0.11	0.57± 0.08	0.67± 0.08
Total SM	51.39± 1.82	42.47± 1.67	30.02± 1.41	19.14± 0.86	16.00± 0.98
$m(\tilde{t}, \tilde{\chi}_1 0) = (500, 380)$	19.08± 1.89	26.38± 2.26	24.01± 2.14	21.08± 1.98	30.90± 2.39

Table 8.7: Expected events in the corresponding bin of the shape-fit configuration for 139 fb^{-1}. Only the statistical uncertainty of the simulated events is given. The $t\bar{t}$ background is not yet normalised to the CR.

In the following, the $t\bar{t}$ background is not divided into fractions, except where it is explicitly mentioned, because the dominant SM background source in the signal region originates from dileptonic $t\bar{t}$ production and semi-leptonic top quark pair processes are effectively suppressed.

8.4 Background estimation and validation

As in the previous chapter, the SM background originating from $t\bar{t}$ production is estimated with a data-driven method, while the remaining SM backgrounds are estimated from simulation. The number of simulated $t\bar{t}$ events is normalised to match the observed data in the control region and afterwards, the obtained normalisation factor is applied on simulated $t\bar{t}$ processes in order to estimate the $t\bar{t}$ background in the signal region.

In order to test the background estimation method, a validation region is defined in which the background is extrapolated from the CR, using the dedicated $t\bar{t}$ normalisation factor, and compared with the observed data. Hence, a potential contribution from SUSY signal events should be sufficiently small.

Since the output score of the ML classifier is used as the main discriminating variable to separate the SUSY signal from SM background processes, the main objective is to use the same variable to define the CR and VR. Moreover, the event selections of the control, validation as well as for the signal region are statistically independent and do not overlap in phase space.

8.4.1 Control region

The control region is defined as summarised in Table 8.8. In addition, the event selection of the signal and validation regions are listed in order to emphasise the kinematical differences.

The CR is characterised by selecting events in the range from 0.40 to 0.60 in $\mathrm{NN}_{\mathrm{bWN}}$. By shifting the event selection to lower $\mathrm{NN}_{\mathrm{bWN}}$ values, a potential signal contamination

	SR	CR	VR
m_{T}	> 110 GeV	> 150 GeV	> 150 GeV
$\mathrm{NN}_{\mathrm{bWN}}$	> 0.9 (or > 0.65 for shape-fit)	$0.40 - 0.60$	$0.60 - 0.65$

Table 8.8: Summary of the event selection for the signal region and the associated control and validation region. The event preselection as defined in Table 8.1 is applied to all regions, but not explicitly shown.

could be reduced. To preserve the composition of the SM background events as similar as in the signal region an additional requirement on $m_{\mathrm{T}} > 150$ GeV is applied.

The upper part of Table 8.9 lists the number of observed events and the number of predicted SM events in the control region, in which $t\bar{t}$ processes are normalised accordingly. The corresponding number of simulated SM events is listed in the lower part of the table for comparison. The event yields of SM backgrounds other than $t\bar{t}$ production are identical, because they are estimated from simulations and normalised to their theoretical cross section. The uncertainties in the upper list include statistical and systematic uncertainties, while in the lower list only statistical uncertainties are given. As illustrated in Figure 7.6, the relative contributions from all SM backgrounds in the control region are very similar to the composition of the SM background in the signal region.

Control region	Events
Observed events	1273
Predicted background events	1272.9 ± 35.7
$t\bar{t}$	1037.0 ± 100.8
Single top	60.6 ± 59.6
W+jets	83.2 ± 39.9
$t\bar{t} + V$	60.2 ± 59.3
Multi-boson	31.9 ± 7.9
Simulated background events	1218.2 ± 10.2
$t\bar{t}$	982.3 ± 7.3
Single top	60.6 ± 3.9
W+jets	83.2 ± 5.8
$t\bar{t} + V$	60.2 ± 1.2
Multi-boson	31.9 ± 0.7

Table 8.9: Observed and predicted SM events in the control region. The individual uncertainties in the upper table are correlated and do not add up quadratically to the total uncertainty. The lower section lists the number of SM events obtained from simulation.

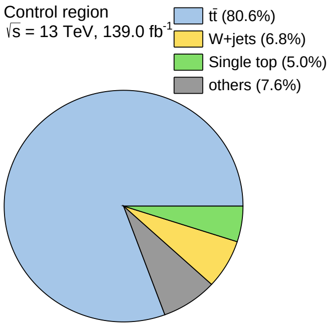


Figure 8.19: Relative contributions of the individual SM backgrounds in the control region.

Potential contributions from SUSY signal events in the control region are verified. The signal contamination in the phase space spanned by $m(\tilde{t}_1)$ versus $m(\tilde{\chi}_1^0)$ is shown in Figure 8.20. Either 2-,3- or 4-body decay modes are considered. The black contour line represents the limit of 5 % signal contamination. All signal points above this limit are already excluded by previous analyses.

Kinematic distributions of selected variables used to train the ML classifier are shown in Figure 8.21 in order to compare observed data events with the SM prediction. Distributions of other input variables in the control region are illustrated in Section A.6 of the Appendix.

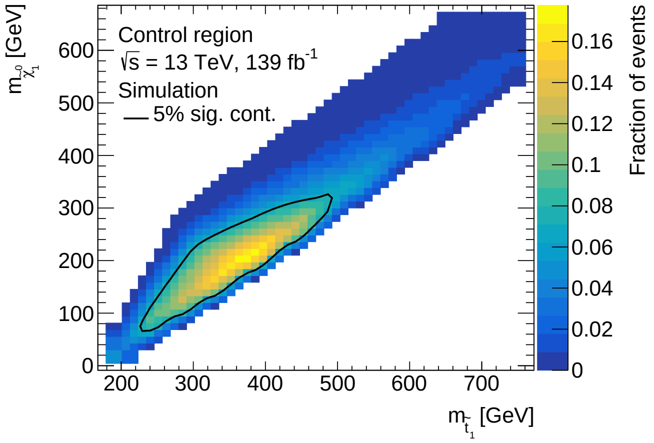
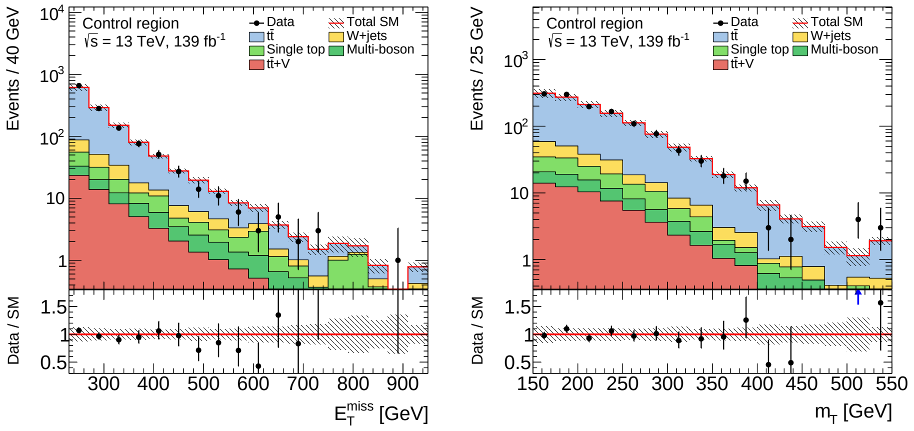


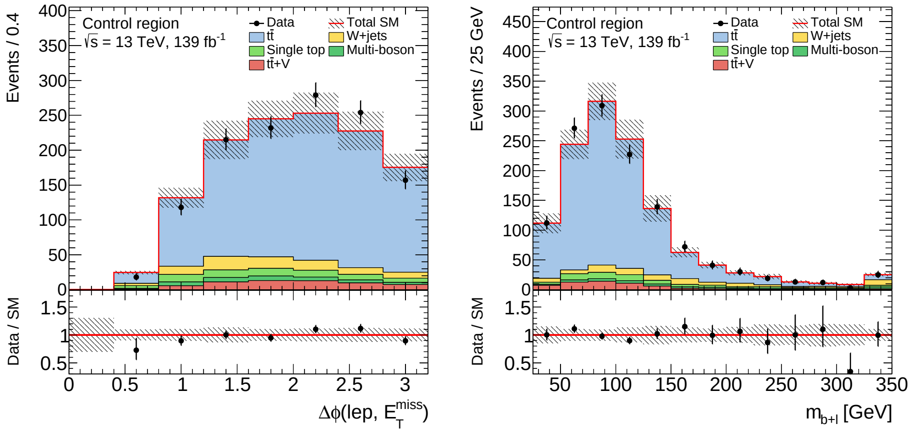
Figure 8.20: Signal contamination in the control region illustrated in the phase space spanned by $m(\tilde{t}_1)$ versus $m(\tilde{\chi}_1^0)$. The black contour line illustrates the threshold for a signal fraction of 5%. Signal points where the contamination is above 5% are already excluded by previous analyses.

In general, a reasonable agreement between data and SM expectation is found for all input variables.



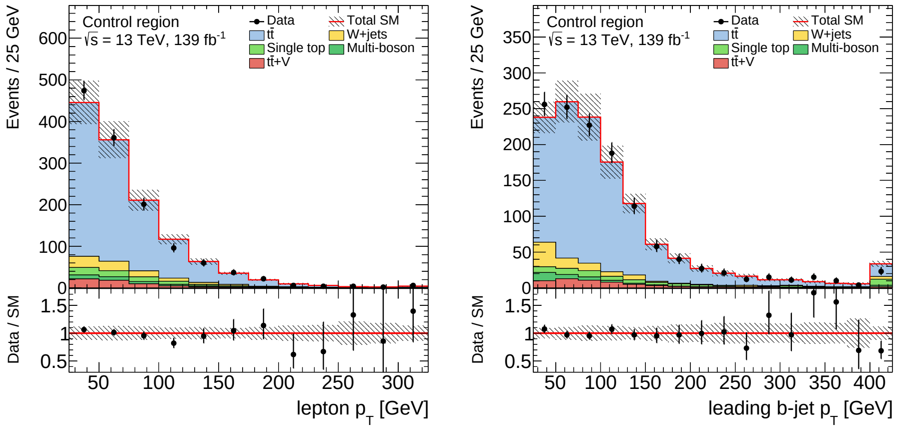
(a) $E_{\mathrm{T}}^{\mathrm{miss}}$

(b) m_{T}



(c) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$

(d) m_{bl}



(e) Lepton p_{T}

(f) Leading b-jet p_{T}

Figure 8.21: Kinematic distributions of selected observables in the control region. The $t\bar{t}$ background prediction is scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross section. The hatched band around the total SM prediction and the hatched band in the data/SM ratio include statistical and systematic uncertainties, except for the $t\bar{t}$ modelling uncertainties, which are only defined for the signal and validation region. Overflows are included in the last bin.

8.4.2 Validation region

In order to validate the extrapolation of the background prediction from the control to the signal region, a validation region is defined. The event selection for the VR is also given in the corresponding column of Table 8.8. The selection on $\mathrm{NN_{bWN}}$ is defined by the interval from 0.6 to 0.65. Alike the definition of the CR, the stringent requirement on m_T is kept at a threshold above 150 GeV in order to suppress semi-leptonic $t\bar{t}$ processes.

A comparison of observed events with predicted SM events (upper part) and the number of events obtained from simulation (lower part) are given in Table 8.10. The $t\bar{t}$ events in the upper list are scaled by the normalisation factor obtained in the CR. Statistical and systematic uncertainties are given in the upper section, while only statistical uncertainties are shown for the number of simulated SM background events in the lower table. The number of predicted SM events agree with the observed data within uncertainty.

Additionally, relative fractions of SM backgrounds contributing to the validation region are shown in Figure 8.22. The composition of the various SM backgrounds in the validation region is similar to that of the control and signal region.

Validation region	Events
Observed events	276
Predicted SM events	257.2 ± 18.3
$t\bar{t}$	217.4 ± 24.0
Single top	7.9 ± 7.8
W+jets	15.6 ± 7.4
$t\bar{t}+V$	9.9 ± 9.7
Multi-boson	6.4 ± 1.6
Simulated background events	245.6 ± 4.3
$t\bar{t}$	205.9 ± 3.1
Single top	7.9 ± 1.4
W+jets	15.6 ± 2.5
$t\bar{t}+V$	9.9 ± 0.5
Multi-boson	6.4 ± 0.4

Table 8.10: Observed and predicted SM events in the validation region. The individual uncertainties in the upper list are correlated and do not add up quadratically to the total uncertainty. The lower section lists the number of SM events obtained from simulation.

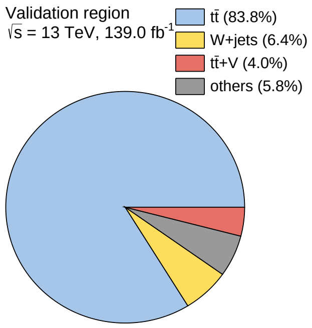


Figure 8.22: Relative contributions of the individual SM backgrounds in the validation region.

The contamination of SUSY signal events in the validation region is also verified. Figure 8.23 shows the fraction of the signal events contributing to the total number of simulated events in the validation region. Either 2-,3- or 4-body decay modes are considered. The black contour in the plot indicates a signal fraction of 10 %. Signal points above this limit are already excluded by previous analyses.

The kinematic distributions of selected variables, used to train the ML algorithm, in the validation region are shown in Figure 8.24. Distributions of the additional input variables in the validation region are given in Section A.7 of the Appendix. In general, a reasonable agreement between observed and predicted events is found for all input variables.

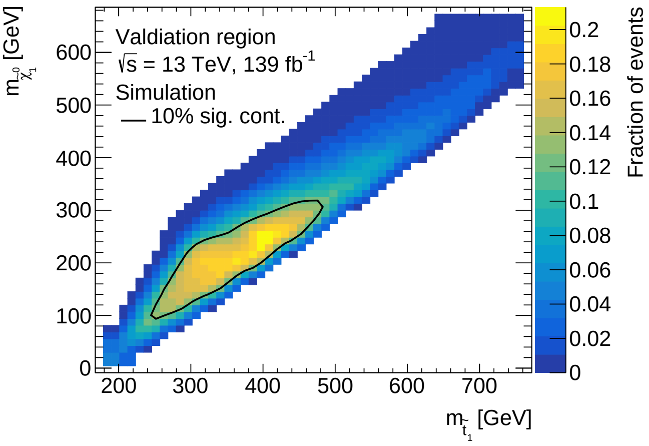
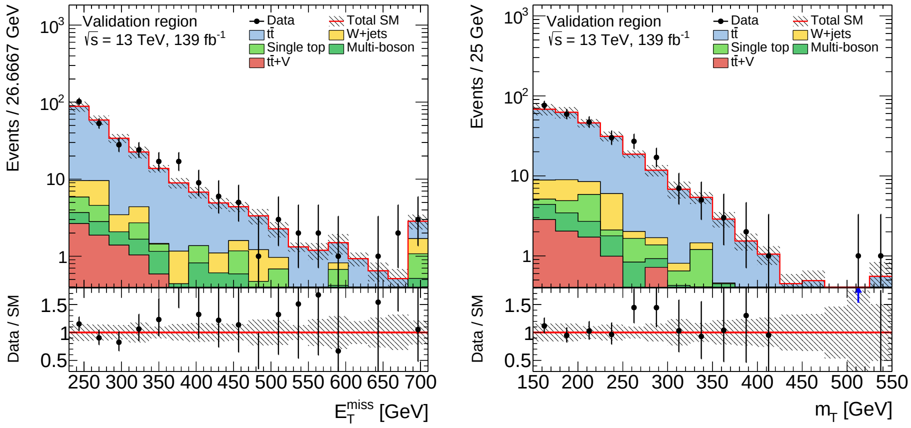
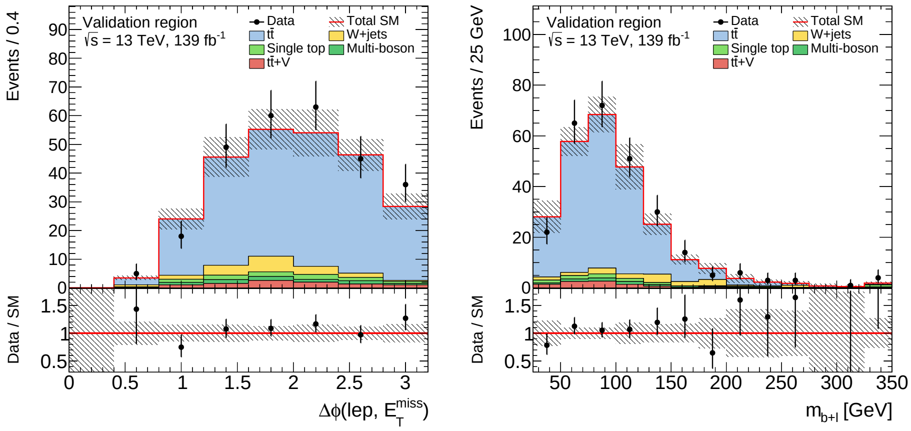


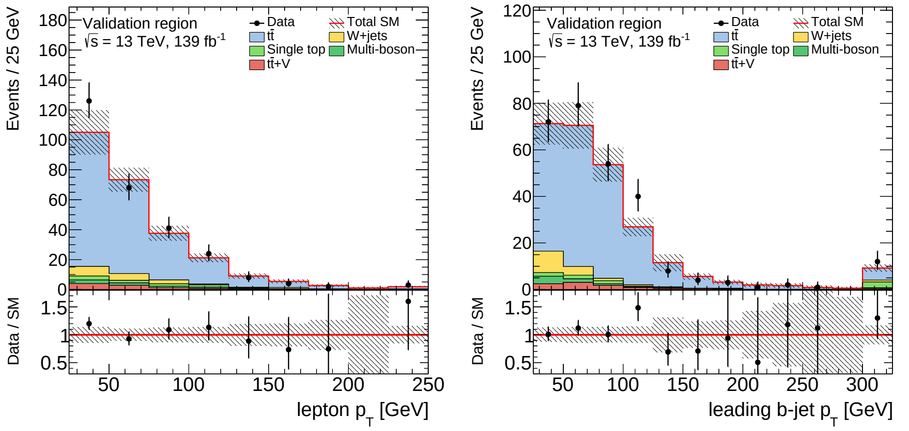
Figure 8.23: Signal contamination in the validation region illustrated in the phase space spanned by $m(\tilde{t}_1)$ versus $m(\tilde{\chi}_1^0)$. The black contour line illustrates the threshold for a signal fraction of 10%. Signal points where the contamination is above 10% are already excluded by previous analyses.



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) m_{T}



(c) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ (d) m_{bl}



(e) Lepton p_{T} (f) Leading b-jet p_{T}

Figure 8.24: Kinematic distributions of selected observables in the validation region. The $t\bar{t}$ background processes are scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross section. The shaded band around the total SM prediction as well as the shaded band in the data/SM ratio include statistical and systematic uncertainties. Overflows are included in the last bin.

8.5 Impact of systematic uncertainties

A comprehensive summary of the various sources of systematic effects, considered in the analysis can be found in Section 6.4. In general, the systematic uncertainties in the corresponding region are evaluated by comparing the nominal number of predicted events with the prediction when the different systematic effects are varied, respectively.

Table 8.11 summarises the dominant systematic uncertainties in the VR and SR. The total statistical uncertainty $\sqrt{N_{\text{exp}}}$ is just listed in the table in order to compare the impact of the overall statistics with the total systematic uncertainty. As in the previous analysis, the dominant systematic effect originates from the modelling and normalisation of the top quark pair production. The methodology to determine $t\bar{t}$ modelling uncertainties is based on the transfer-factor method (see Equation 6.7), in which the differences in the normalisation of the different simulated $t\bar{t}$ samples cancel out by construction. Hence, only differences in the shape of NN_{bWN} between the alternative $t\bar{t}$ samples affect the systematic uncertainty.

Systematic uncertainty	VR	SR
Total predicted SM background	257.2	36.8
Total statistical ($\sqrt{N_{\text{exp}}}$)	±16.0	±6.1
Total background uncertainty	15.0[5.8 %]	7.2[19.6 %]
$t\bar{t}$ hadronisation	±7.8 [3.0 %]	±6.1 [16.6 %]
$t\bar{t}$ normalisation	±21.2 [8.3 %]	±3.1 [8.4 %]
Single top cross section	±7.8 [3.0 %]	±1.9 [5.2 %]
Jet energy resolution (3^{rd} component)	±1.9 [0.7 %]	±1.6 [4.3 %]
$t\bar{t}$ radiation	±1.8 [0.7 %]	±1.4 [3.8 %]
Statistics (simulated samples)	±4.5 [1.7 %]	±1.4 [3.8 %]
Jet energy scale (2^{nd} component)	±2.6 [1.0 %]	±1.2 [3.3 %]
Jet energy resolution (2^{nd} component)	±2.4 [0.9 %]	±1.1 [3.0 %]
Jet energy scale flavour response	±1.5 [0.6 %]	±0.9 [2.4 %]
$t\bar{t} + V$ cross section	±9.7 [3.8 %]	±0.8 [2.2 %]
Jet energy scale (1^{st} component)	±2.1 [0.8 %]	±0.7 [1.9 %]
W+jets cross section	±7.3 [2.8 %]	±0.7 [1.9 %]
Jet energy resolution (1^{st} component)	±2.9 [1.1 %]	±0.7 [1.9 %]
$t\bar{t}$ generator	±2.2 [0.9 %]	±0.6 [1.6 %]
Flavour tagging b-tagging efficiency	±0.1 [0.04 %]	±0.6 [1.6 %]
Jet energy resolution (6^{th} component)	±1.6 [0.6 %]	±0.5 [1.4 %]
Jet energy resolution (7^{th} component)	±0.8 [0.3 %]	±0.5 [1.4 %]
Jet energy resolution (4^{th} component)	±0.5 [0.2 %]	±0.4 [1.1 %]

Table 8.11: Summary of the dominant experimental and theoretical systematic uncertainties on the total predicted SM background in the validation region and the signal region, respectively. Individual uncertainties may be correlated, and hence do not necessarily add up in quadrature to the total uncertainty. Systematic uncertainties which contribute less than 1 % to the total SM prediction are omitted in the list. The percentages show the size of the uncertainty relative to the total predicted SM background. The systematic effects are ordered by the size of their relative contribution to the total predicted SM background in the SR.

The total relative uncertainty in the signal region is 19.6 %, which is compatible with the total statistical uncertainty of the total SM background ($\sqrt{N_{\text{exp}}}$). This is an improvement compared to the previous analysis, in which the total systematic uncertainty was significantly larger than the statistical uncertainty.

The individual systematic uncertainties in each bin of the shape-fit configuration are listed in Table 8.12. Similar as in the single-bin scenario, the largest systematic effects are the modelling of the $t\bar{t}$ background and the calibration of the jet energy resolution. To avoid complexity, the various components contributing to jet energy scale and resolution, $E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution, as well as systematic effects associated to the flavour tagging efficiency are summarised into a single component. Systematic effects which contribute less than 1 % to the total SM prediction are omitted in the list.

The total uncertainty in the various bins ranges from 5.7 % to 28.7 %, relative to the total predicted SM background. Systematic effects are still the dominant source of uncertainty in the shape-fit configuration, but compatible with the respective total statistical uncertainty ($\sqrt{N_{\mathrm{exp}}}$).

Systematic uncertainty	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Total predicted SM background	238.1	207.4	170.0	76.0	63.3
Total statistical ($\sqrt{N_{\mathrm{exp}}}$)	±15.43	±14.40	±13.04	±8.71	±7.96
Total background uncertainty	±13.6 [5.7 %]	±17.3 [8.3 %]	±21.8 [12.8 %]	±7.4 [9.7 %]	±7.6 [12.0 %]
$t\bar{t}$ hadronisation	±6.8 [2.9 %]	±12.7 [6.1 %]	±17.9 [10.5 %]	±2.9 [3.8 %]	±5.4 [8.5 %]
$t\bar{t}$ generator	±2.6 [1.1 %]	±0.7 [0.3 %]	±4.9 [2.9 %]	±3.4 [4.5 %]	±1.5 [2.4 %]
Jet energy resolution	±3.9 [1.6 %]	±3.8 [1.8 %]	±7.6 [4.5 %]	±2.4 [3.2 %]	±3.1 [4.9 %]
$t\bar{t}$ normalisation	±19.6 [8.2 %]	±17.0 [8.2 %]	±14.2 [8.4 %]	±6.5 [8.6 %]	±5.4 [8.5 %]
Statistics (simulated sample)	±4.1 [1.7 %]	±3.9 [1.9 %]	±3.2 [1.9 %]	±2.2 [2.9 %]	±1.8 [2.8 %]
Jet energy scale	±3.3 [1.4 %]	±6.6 [3.2 %]	±3.5 [2.1 %]	±2.4 [3.2 %]	±0.3 [0.5 %]
Single top cross section	±8.6 [3.6 %]	±9.6 [4.6 %]	±6.0 [3.5 %]	±2.6 [3.4 %]	±1.3 [2.1 %]
$t\bar{t}$ radiation	±2.2 [0.9 %]	±2.2 [1.1 %]	±0.9 [0.5 %]	±1.2 [1.6 %]	±1.1 [1.7 %]
W+jets cross section	±6.4 [2.7 %]	±5.4 [2.6 %]	±3.7 [2.2 %]	±1.5 [2.0 %]	±1.7 [2.7 %]
$t\bar{t}+V$ cross section	±8.9 [3.7 %]	±6.2 [3.0 %]	±6.0 [3.5 %]	±2.3 [3.0 %]	±1.8 [2.8 %]
Flavour tagging efficiency	±1.7 [0.7 %]	±1.0 [0.5 %]	±0.8 [0.5 %]	±0.9 [1.2 %]	±0.9 [1.4 %]
Multi-boson cross section	±1.4 [0.6 %]	±1.2 [0.6 %]	±1.0 [0.6 %]	±0.4 [0.5 %]	±0.4 [0.6 %]
$E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution	±3.7 [1.6 %]	±1.3 [0.6 %]	±2.2 [1.3 %]	±2.7 [3.6 %]	±1.3 [2.1 %]

Systematic uncertainty	Bin 6	Bin 7	Bin 8	Bin 9	Bin 10
Total predicted SM background	53.7	44.4	31.4	20.0	16.7
Total statistical ($\sqrt{N_{\mathrm{exp}}}$)	±7.33	±6.67	±5.61	±4.48	±4.09
Total background uncertainty	±7.8 [14.5 %]	±7.1 [16.0 %]	±5.7 [18.2 %]	±3.2 [16.0 %]	±4.8 [28.7 %]
$t\bar{t}$ hadronisation	±6.7 [12.5 %]	±4.0 [9.0 %]	±3.9 [12.4 %]	±2.0 [10.0 %]	±4.0 [24.0 %]
$t\bar{t}$ generator	±2.1 [3.9 %]	±2.2 [5.0 %]	±2.1 [6.7 %]	±0.9 [4.5 %]	±1.5 [9.0 %]
Jet energy resolution	±0.6 [1.1 %]	±3.8 [8.6 %]	±2.3 [7.3 %]	±1.3 [6.5 %]	±1.4 [8.4 %]
$t\bar{t}$ normalisation	±4.4 [8.2 %]	±3.7 [8.3 %]	±2.6 [8.3 %]	±1.7 [8.5 %]	±1.3 [7.8 %]
Statistics (simulated samples)	±1.9 [3.5 %]	±1.7 [3.8 %]	±1.5 [4.8 %]	±0.9 [4.5 %]	±1.0 [6.0 %]
Jet energy scale	±1.6 [3.0 %]	±2.6 [5.9 %]	±1.4 [4.5 %]	±0.8 [4.0 %]	±1.0 [6.0 %]
Single top cross section	±2.1 [3.9 %]	±1.6 [3.6 %]	±0.7 [2.2 %]	±1.0 [5.0 %]	±0.9 [5.4 %]
$t\bar{t}$ radiation	±0.9 [1.7 %]	±1.0 [2.3 %]	±0.6 [1.9 %]	±0.9 [4.5 %]	±0.5 [3.0 %]
W+jets cross section	±2.0 [3.7 %]	±1.3 [2.9 %]	±1.0 [3.2 %]	±0.2 [1.0 %]	±0.5 [3.0 %]
$t\bar{t}+V$ cross section	±1.7 [3.2 %]	±1.2 [2.7 %]	±1.0 [3.2 %]	±0.4 [2.0 %]	±0.5 [3.0 %]
Flavour tagging efficiency	±0.7 [1.3 %]	±0.5 [1.1 %]	±0.5 [1.6 %]	±0.3 [1.5 %]	±0.4 [2.4 %]
Multi-boson cross section	±0.3 [0.5 %]	±0.3 [0.7 %]	±0.2 [0.6 %]	±0.1 [0.5 %]	±0.2 [1.2 %]
$E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution	±0.4 [0.7 %]	±1.3 [2.9 %]	±1.2 [3.8 %]	±0.4 [2.0 %]	±0.1 [0.6 %]

Table 8.12: Summary of the dominant systematic uncertainties on the total predicted number of SM background events for the bins included in the shape-fit. Individual uncertainties may be correlated, and hence do not necessarily add up in quadrature to the total uncertainty. In order to avoid complexity, the various components for jet energy scale and resolution, $E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution, as well as systematic effects affecting the flavour tagging efficiency are summarised accordingly. Systematic uncertainties which contribute less than 1 % to the total SM prediction are omitted in the list. The percentages show the size of the uncertainty relative to the total predicted SM background. The sources of systematic uncertainties are ordered by their relative contribution to the total predicted SM background in the most sensitive bin (Bin 10) with respect to the SUSY signal benchmark.

8.6 Results and interpretation

The number of observed events and the number of predicted SM background events from the background-only fit are illustrated in Figure 8.25 for the validation and signal region, respectively. In addition, the significance of the observed events given the predicted SM background events is shown in the bottom panel. Data matches the SM prediction within uncertainties. A detailed summary of the number of observed events and the predicted number of SM background events in the SR is listed in Table 8.13. Furthermore, the fit parameter that controls the normalisation of the $t\bar{t}$ background ($\mu_{t\bar{t}}$) is given. The $t\bar{t}$ normalisation factor of 1.06 ± 0.1 is compatible with unity. Uncertainties, listed in the table, include statistical as well as systematic uncertainties. The last two rows in Table 8.13 represent the expected (S^{95}_{exp}) and observed (S^{95}_{obs}) upper limits on the number of beyond-SM events at 95 % confidence level. These numbers are derived from a profile-likelihood ratio test in which the level of agreement of the SM background-only hypothesis with the observed data in the SR is quantified. In this procedure, a generic signal model is assumed that contributes only to the SR and for which neither experimental nor theoretical systematic uncertainties except for the luminosity uncertainty are considered.

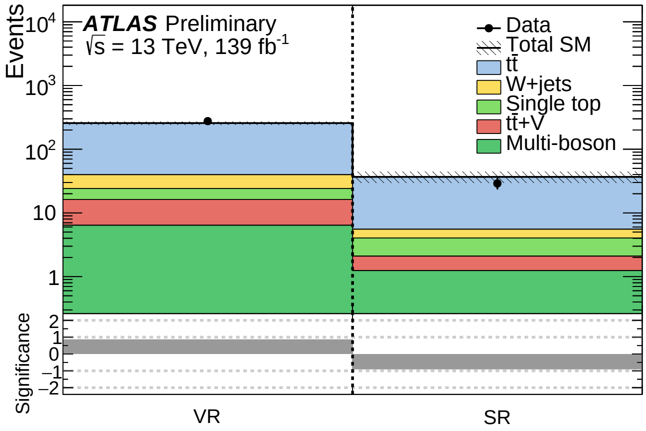
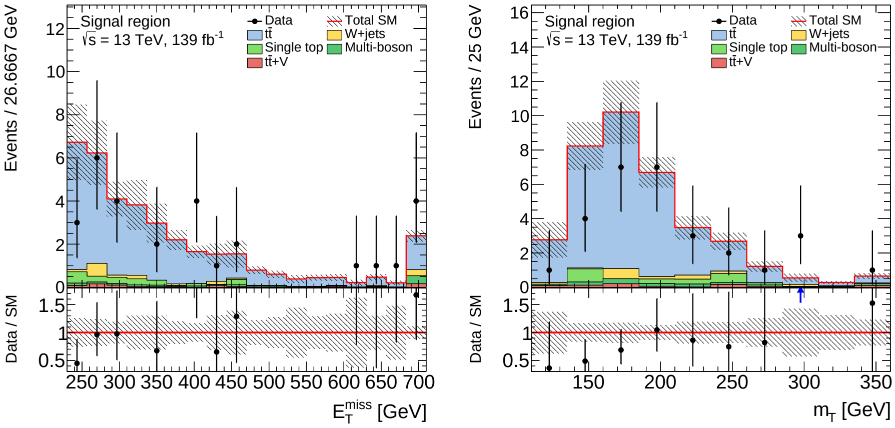


Figure 8.25: Comparison of observed events with the SM prediction in the VR and SR. The shaded area around the total SM background prediction includes all uncertainties. The bottom panel shows the significance of the observed data events given the number of predicted SM background events.

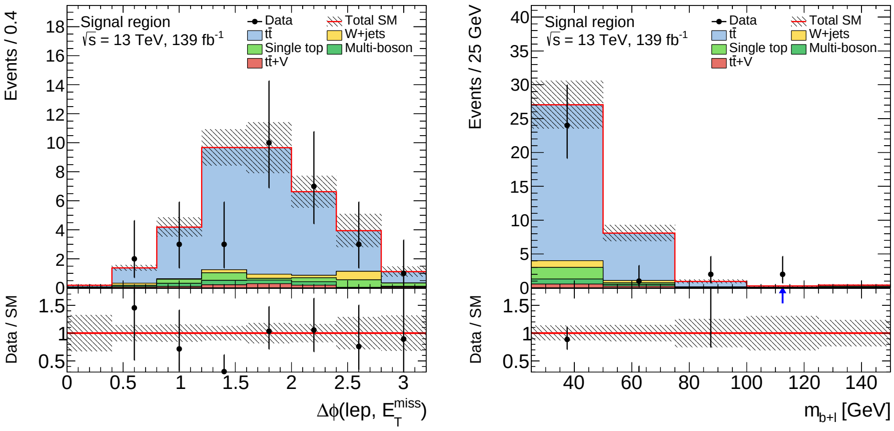
	Signal region
Observed events	29
Total background	36.8 ± 7.2
$t\bar{t}$	31.2 ± 7.6
Single top	2.0 ± 1.9
W+jets	1.50 ± 0.73
$t\bar{t} + V$	0.85 ± 0.84
Multi-boson	1.24 ± 0.31
$\mu_{t\bar{t}}$	1.06 ± 0.10
S^{95}_{exp}	$18.3^{+7.3}_{-4.9}$
S^{95}_{obs}	13.8

Table 8.13: Number of observed events in the signal region together with the number of predicted SM background events and their total uncertainties. In addition, the normalisation factor ($\mu_{t\bar{t}}$) for the prediction of the $t\bar{t}$ background is listed, and the expected (S^{95}_{exp}) and observed (S^{95}_{obs}) 95 % confidence level upper limits on the number of beyond-SM events.

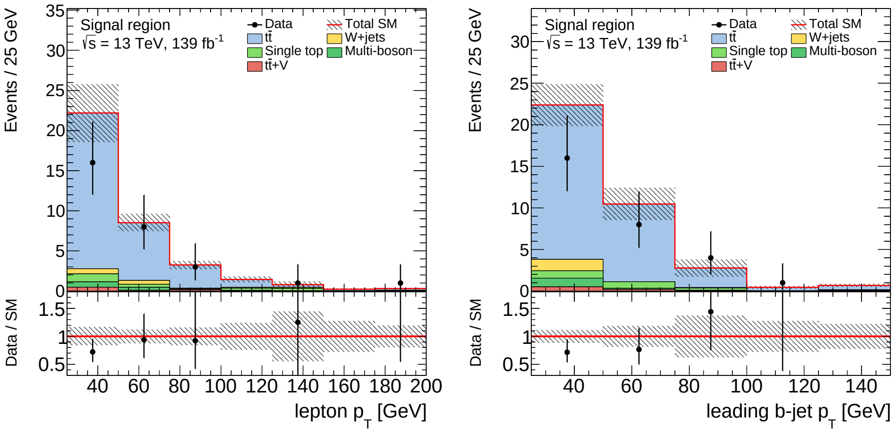
Figure 8.26 shows the kinematic distributions of some input variables for observed data events and the predicted SM background, respectively. A reasonable agreement between data and SM prediction is observed. Further kinematic distributions of variables used to train the ML algorithm are shown in Section A.8 of the Appendix.



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) m_{T}



(c) $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ (d) m_{bl}



(e) Lepton p_{T} (f) Leading b-jet p_{T}

Figure 8.26: Kinematic distributions of selected observables in the signal region. The $t\bar{t}$ background processes are scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross section. The shaded band around the total SM prediction as well as the shaded band in the data/SM ratio include statistical and systematic uncertainties. Overflows are included in the last bin.

The number of observed events and the number of predicted SM background events for each signal region included in the shape-fit configuration is summarised in Table 8.14. In addition, the breakdown of individual SM backgrounds that contribute to each SR in the shape-fit is also listed. Figure 8.27 shows the comparison between the observed and predicted events in the ML classifier distribution. The variable bin width is equivalent to the selection of the control, validation and the signal regions of the shape-fit. Good agreement between observed data and the SM background prediction is found. All uncertainties are included in the hatched area around the total SM prediction. The bottom panel shows the significance of the observed data events given the number of predicted SM background events. Additionally, the expected distribution from a representative SUSY signal model is overlaid.

Shape-fit	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
NN$_{\text{bWN}}$	$[0.65, 0.70]$	$[0.70, 0.75]$	$[0.75, 0.80]$	$[0.80, 0.82]$	$[0.82, 0.84]$
m_{T} [GeV]	$m_{\text{T}} > 150$	$m_{\text{T}} > 150$	$m_{\text{T}} > 150$	-	-
Observed events	196	215	173	61	63
Total SM background	238 ± 14	207 ± 17	170 ± 22	76 ± 7	63 ± 8
$t\bar{t}$	200.50 ± 22.22	174.41 ± 22.95	145.56 ± 25.15	66.16 ± 9.25	54.92 ± 8.68
Single top	8.71 ± 8.57	9.74 ± 9.58	6.10 ± 6.00	2.61 ± 2.57	1.28 ± 1.28
W+jets	13.69 ± 6.51	11.62 ± 5.47	7.96 ± 3.79	3.15 ± 1.50	3.63 ± 1.74
$t\bar{t}+V$	9.01 ± 8.87	6.31 ± 6.22	6.09 ± 6.00	2.32 ± 2.29	1.79 ± 1.77
Multi-boson	6.19 ± 1.55	5.28 ± 1.30	4.30 ± 1.07	1.71 ± 0.42	1.70 ± 0.45

Shape-fit	Bin 6	Bin 7	Bin 8	Bin 9	Bin 10
NN$_{\text{bWN}}$	$[0.84, 0.86]$	$[0.86, 0.88]$	$[0.88, 0.90]$	$[0.90, 0.92]$	$[0.92, 1.0]$
m_{T} [GeV]	-	-	-	-	-
Observed events	51	42	32	16	13
Total SM background	54 ± 8	44 ± 7	31 ± 6	20.1 ± 3.2	17 ± 5
$t\bar{t}$	44.54 ± 8.67	37.50 ± 7.77	26.66 ± 6.08	17.60 ± 3.43	13.63 ± 4.89
Single top	2.10 ± 2.07	1.58 ± 1.55	0.69 ± 0.68	1.04 ± 1.02	0.91 ± 0.90
W+jets	4.21 ± 2.01	2.74 ± 1.30	2.19 ± 1.05	0.48 ± 0.24	1.01 ± 0.50
$t\bar{t}+V$	1.69 ± 1.66	1.24 ± 1.22	1.00 ± 0.98	0.37 ± 0.37	0.48 ± 0.48
Multi-boson	1.20 ± 0.31	1.38 ± 0.34	0.89 ± 0.22	0.57 ± 0.14	0.67 ± 0.17

Table 8.14: Observed data events in each bin of the shape-fit configuration together with the predicted number of total SM background events and their uncertainties. The bin i (i = 1–10) corresponds to the i-th bin (from left to right) of the ML classifier output score used in the shape-fit. Contributions of the individual SM backgrounds are also given. Individual uncertainties may be correlated, and hence do not necessarily add up in quadrature to the total uncertainty.

No data excesses in any of the signal regions (single-bin or shape-fit) are observed. Hence, exclusion limits are derived for the particular top squark pair production models based on profile-likelihood ratio tests.

For the hypothesis tests, the signal uncertainties and potential signal contributions to all regions are taken into account. All uncertainties are included in the fit procedure, except for the uncertainty on the theoretical signal cross section, because the variation based on the theoretical signal cross section is evaluated separately. Exclusion limits at 95 % confidence level are obtained for each SUSY signal model, respectively. Finally, the exclusion contours are derived by interpolating in the CL$_{\text{s}}$ values of all considered signal points.

The expected and observed exclusion contour as a function of the top squark mass and the neutralino mass in the targeted 3-body decay mode is shown in Figure 8.28. Signal

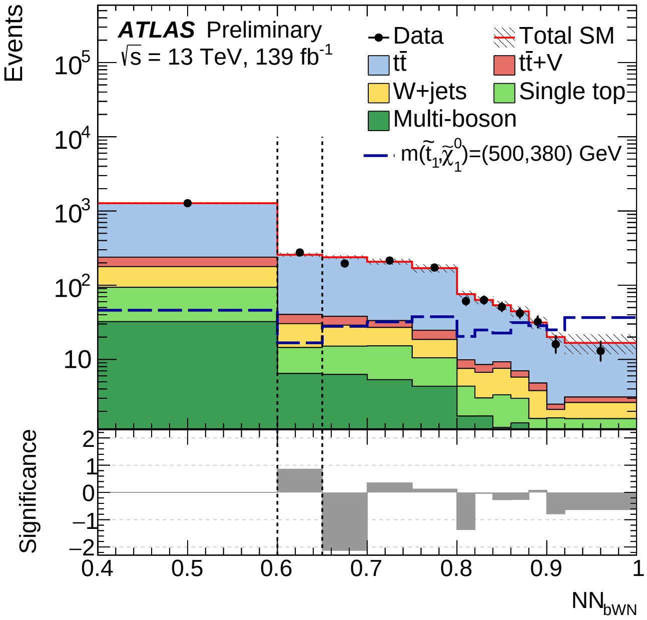


Figure 8.27: Distribution of NN$_{\mathrm{bWN}}$. The variable bin widths correspond to the definitions of the control, validation and signal regions. The shaded area around the total SM prediction includes all uncertainties. The expected distribution from a representative signal benchmark model is overlaid. The bottom panel shows the significance of the observed data events given the number of predicted SM background events.

points of the adjacent 2-body and 4-body decay region are also included in the interpolation to determine the exclusion contour. The yellow band represents the $\pm 1\,\sigma_{\mathrm{exp}}$ uncertainty in the expected limit and indicates how much the expected limit is affected by the systematic and statistical uncertainties. The up and down variation of the nominal theoretical signal cross section by its corresponding uncertainty is represented as the $\pm 1\,\sigma_{\mathrm{th}}$ uncertainty lines around the observed limit. The exclusion contour from a previous search for direct top squark pair production in the same phase space [58] is overlaid (grey shaded region).

To provide a greater level of detail, Figure 8.29 shows the expected and observed exclusion contours in the plane spanned by the top squark mass and the mass splitting $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$. The grey shaded region shows again the exclusion contour from the previous analysis [58].

As in the previous analysis, a non-excluded area between the transition of the 3- and 4-body decay mode is noticeable at top squark masses of approximately 600 GeV and 650 GeV. Due to the change of the kinematic properties of the signal models, the sensitivity decreases. For signal models which are characterised by a small mass splitting $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$, the momenta of the two b-tagged jets emerging from the top squark decay can decrease to 0, which affects the acceptance of the p_{T} requirement on the b-tagged jets in the signal region.

The region of a top squark mass up to 720 GeV with a neutralino mass of approximately 580 GeV (assuming $\mathcal{B}(\tilde{t}_1 \to bW\tilde{\chi}_1^0) = 100\%$) is excluded and hence these results improve upon previous exclusion limits.

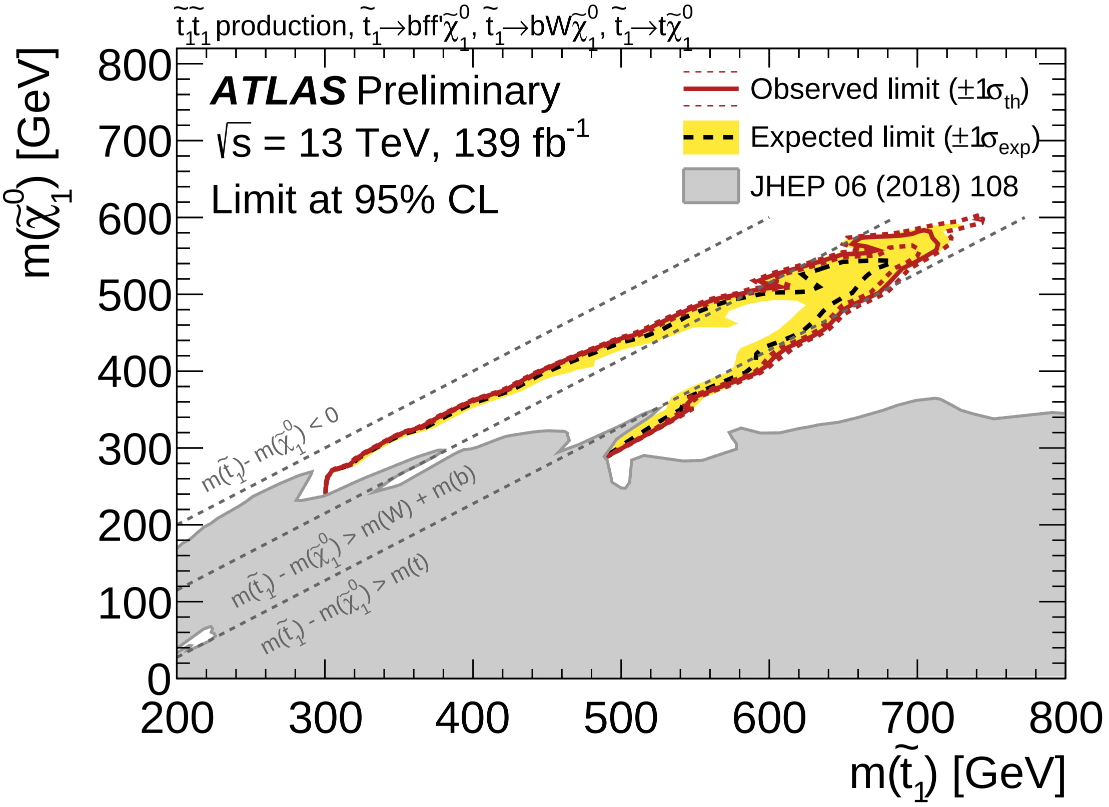


Figure 8.28: Expected (black dashed) and observed (red solid) excluded regions in the plane of $m(\tilde{\chi}_1^0)$ versus $m(\tilde{t}_1)$ at 95 % confidence level for direct top squark pair production. Either $\tilde{t}_1 \rightarrow t\tilde{\chi}_1^0$, $\tilde{t}_1 \rightarrow bW\tilde{\chi}_1^0$, or $\tilde{t}_1 \rightarrow bff'\tilde{\chi}_1^0$ decay modes are assumed with a branching ratio of 100 %. The grey shaded area illustrates the excluded region from a previous publication [58], including also the results obtained from the analysis described in Chapter 7.

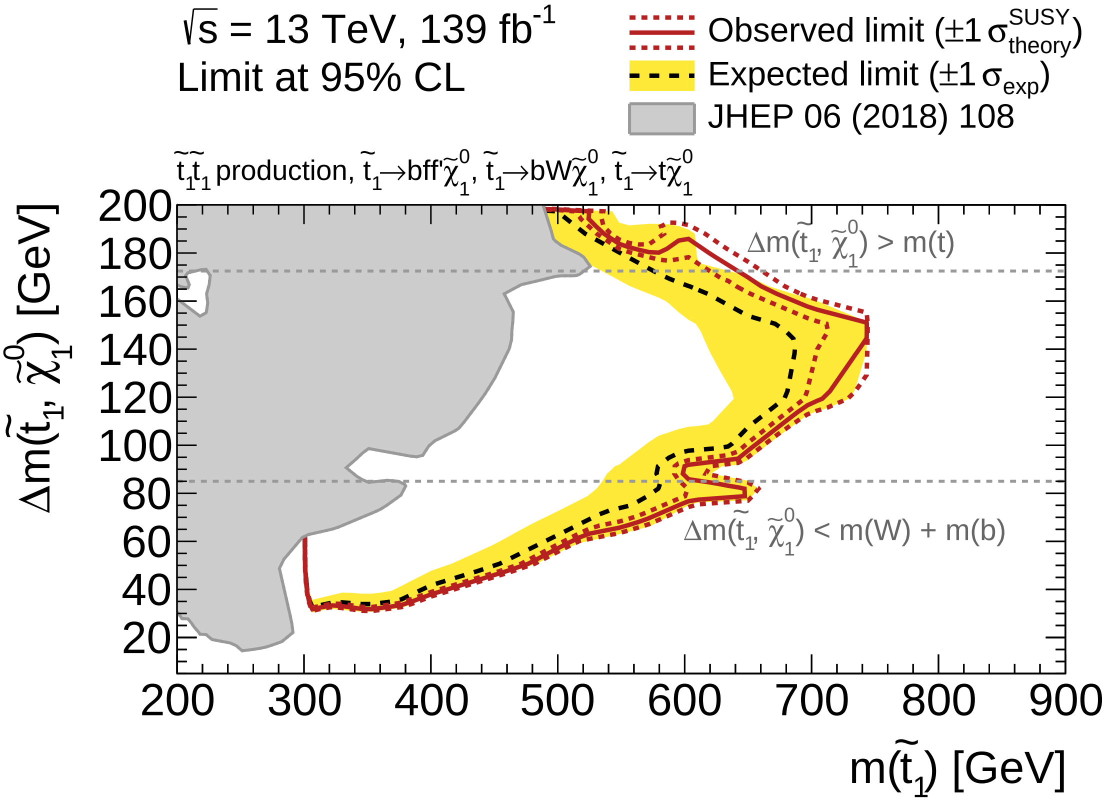


Figure 8.29: Expected (black dashed) and observed (red solid) excluded regions in the plane of $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$ versus $m(\tilde{t}_1)$ at 95 % confidence level for direct top squark pair production. Either $\tilde{t}_1 \to t\tilde{\chi}_1^0$, $\tilde{t}_1 \to bW\tilde{\chi}_1^0$, or $\tilde{t}_1 \to bff'\tilde{\chi}_1^0$ decay modes are assumed with a branching ratio of 100 %. The grey shaded area illustrates excluded regions from a previous publication [58], including also the results obtained from the analysis described in Chapter 7.

Chapter 9

Comparison of current results from searches for top squark pair production

The results obtained from the two analyses described in the previous chapters are compared in terms of exclusion limits in the following. The actual result is also compared with the most recent result from a similar search from the CMS collaboration.

In order to quantitatively compare the results from the two searches, the exact same shape-fit configuration in the am_{T2} variable as defined in Table 7.2, together with associated CR to normalise the $t\bar{t}$ background events (see Table 7.5) is performed on the SM background events and SUSY signal events using the simulated event samples from analysis `bWN-fullSet` (see Table 4.2). The same experimental uncertainties as they are derived in analysis `bWN-fullSet` are applied. The $t\bar{t}$ modelling uncertainties based on the transfer-factor method, as well as the theoretical uncertainties on the signal models are re-computed according to the event selections of the am_{T2} shape-fit and the dedicated CR.

The expected limit using the shape-fit configuration in am_{T2} is compared with the expected and observed limit which is derived from the shape-fit in $\mathrm{NN}_{\mathrm{bWN}}$ in Figure 9.1 in the plane spanned by $m(\tilde{t}_1)$ and $m(\tilde{\chi}_1^0)$. Both results correspond to an integrated luminosity of 139 fb^{-1}. The observed limit from the am_{T2}-shape-fit approach is omitted, because the analysis is not approved by the ATLAS collaboration. The grey shaded region represents the excluded region from the previous publication corresponding to 36.1 fb^{-1} [55]. For a greater level of detail, the respective limits are also illustrated as a function of $m(\tilde{t}_1)$ versus $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$ in Figure 9.2. The actual limit obtained from a shape-fit in $\mathrm{NN}_{\mathrm{bWN}}$ improves upon the results from a shape-fit using the am_{T2} variable by approximately 200 GeV in $m(\tilde{t}_1)$ depending on the mass of the neutralino.

Finally, the presented results are compared with the results from a search performed by the CMS collaboration. Figure 9.3 shows the exclusion limit at 95 % confidence level in the $m(\tilde{t}_1)$-$m(\tilde{\chi}_1^0)$ plane from an inclusive search for gluinos and squarks in fully hadronic events derived by the CMS collaboration [203]. The analysis is conducted on data from pp collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV recorded with the CMS detector corresponding to an integrated luminosity of 137 fb^{-1}. Emphasis of the comparison is on the region where 3-body decay modes are kinematically favoured, indicated by thin grey dashed lines and by the text "$m_{\tilde{t}_1} = m_t + m_{\tilde{\chi}_1^0}$" in the figure. The phase space in which the top squark decays via a 4-body decay is omitted in the figure. The exclusion limits are based on maximum likelihood fits to data in multiple disjunct signal regions. The signal regions are categorised by the hadronic energy in the events, the number of jets, the number of b-tagged jets and the value of M_{T2}. In the 3-body decay region, top squark masses of up to 590 GeV with a neutralino masses of approximately 500 GeV are excluded.

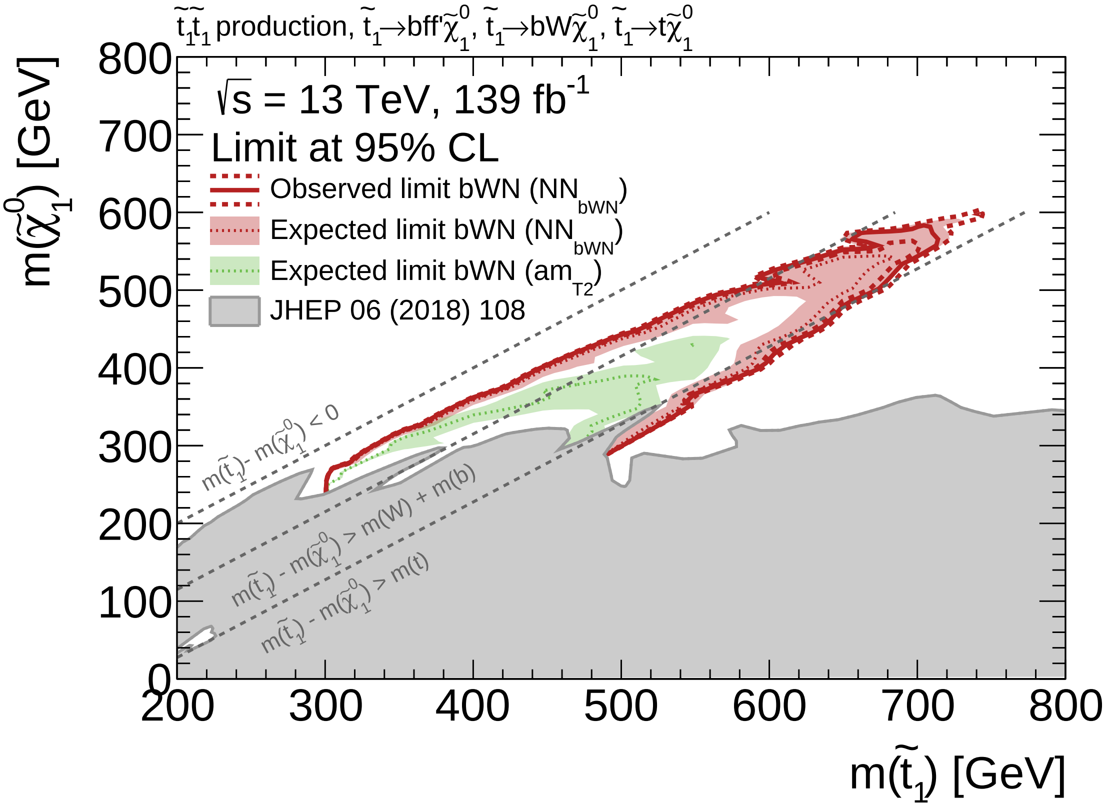


Figure 9.1: Comparison of the expected (green dashed) excluded region obtained from a shape-fit in am_{T2} with the expected (red dashed) and observed (red solid) excluded regions derived from a shape-fit in $\mathrm{NN_{bWN}}$ in the plane of $m(\tilde{t}_1)$ versus $m(\tilde{\chi}_1^0)$ for direct top squark pair production. The limits correspond to 95 % confidence level. Either $\tilde{t}_1 \to t\tilde{\chi}_1^0$, $\tilde{t}_1 \to bW\tilde{\chi}_1^0$, or $\tilde{t}_1 \to bff'\tilde{\chi}_1^0$ decay modes with a branching ratio of 100 % are assumed. The excluded region from the previous publication [55] is illustrated as the grey shaded region.

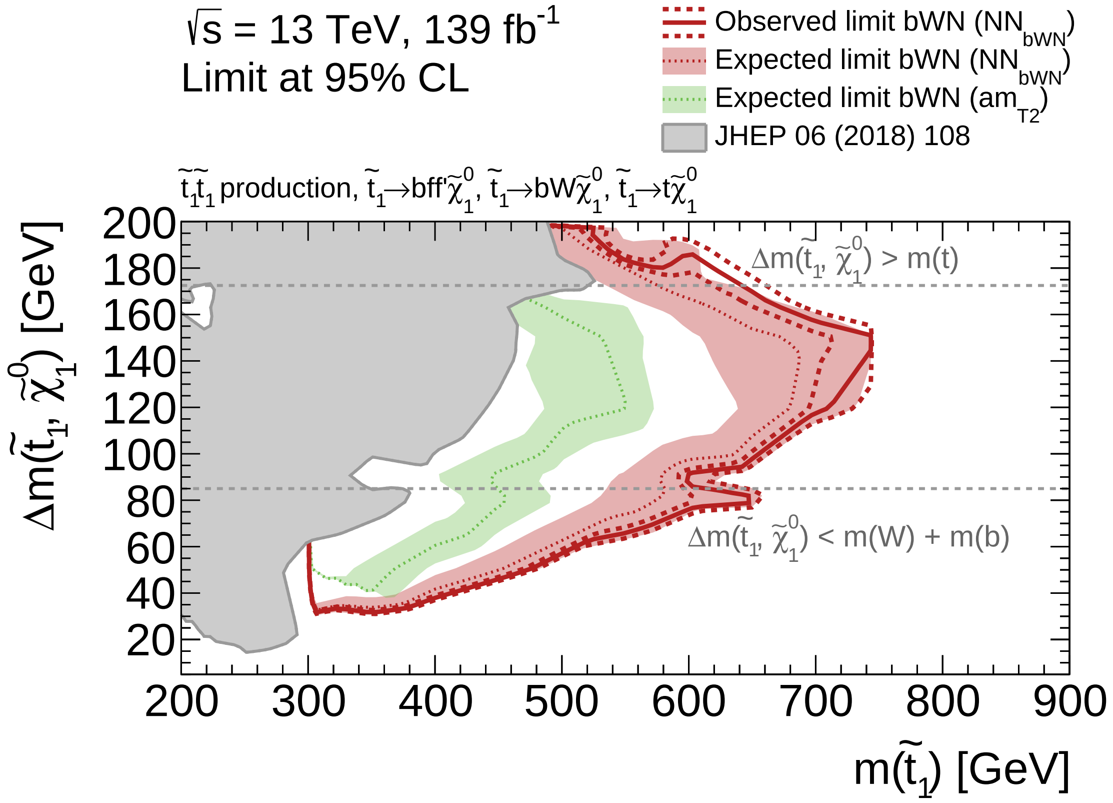


Figure 9.2: Comparison of the expected (green dashed) excluded region obtained from a shape-fit in am_{T2} with the expected (red dashed) and observed (red solid) excluded regions derived from a shape-fit in NN$_{\mathrm{bWN}}$ in the plane of $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$ versus $m(\tilde{t}_1)$ for direct top squark pair production. The limits correspond to 95 % confidence level. Either $\tilde{t}_1 \to t\tilde{\chi}_1^0$, $\tilde{t}_1 \to bW\tilde{\chi}_1^0$, or $\tilde{t}_1 \to bff'\tilde{\chi}_1^0$ decay modes with a branching ratio of 100 % are assumed. The excluded region from the previous publication [55] is illustrated as the grey shaded region.

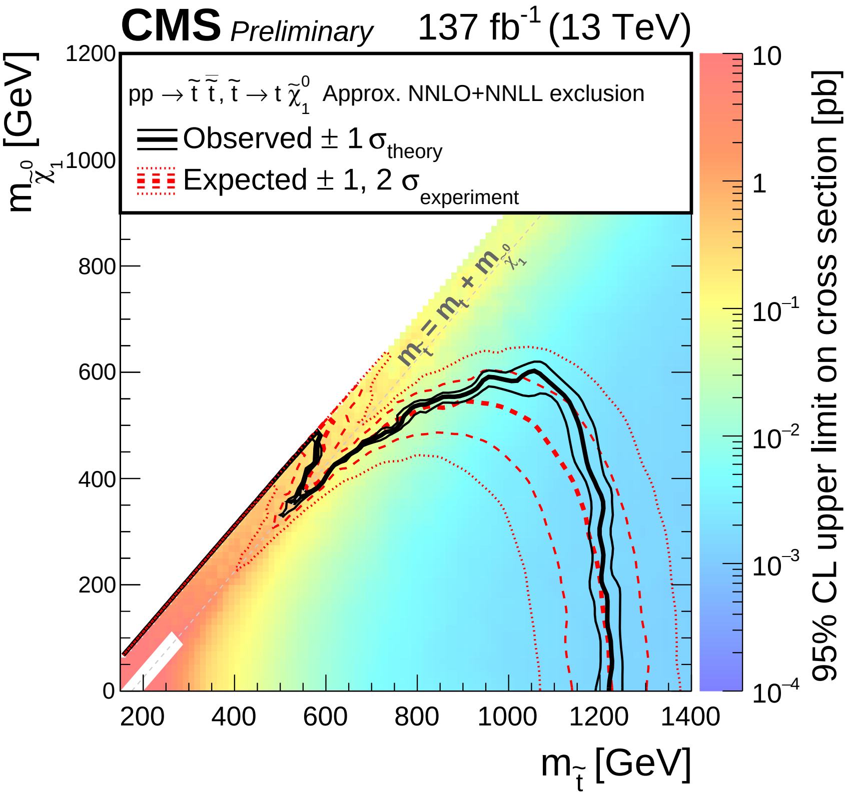


Figure 9.3: Expected (red dashed) and observed (black solid) 95% exclusion limit for top squark pair production [203] in the plane spanned by $m(\tilde{t}_1)$ and $m(\tilde{\chi}^0_1)$. Thin red dashed lines indicate the $\pm 1, 2\,\sigma_{\text{experiment}}$ uncertainty in the expected limit. The thin black lines represent the $\pm 1\,\sigma_{\text{theory}}$ uncertainty based on the theoretical uncertainties on the signal cross section.

Chapter 10

Conclusion

In this thesis, the results from two individual searches for direct top squark pair production are presented. The searches target the same supersymmetric signal model in which the top squark decays via a 3-body process into a b quark, a W boson, and a neutralino. Such a decay scenario is kinematically favoured, if the mass difference between the top squark and the lightest supersymmetric particle is smaller than the top quark mass. In both analyses, events with one isolated electron or muon, multiple hadronic jets, and large missing transverse momentum in the final state are considered. The lepton in the final state is assumed to originate from the decay of a W boson.

The first analysis is performed on data from proton-proton collisions delivered by the Large Hadron Collider at a centre-of-mass energy of $\sqrt{s} = 13$ TeV within the years 2015 and 2016 and recorded by the ATLAS detector, corresponding to an integrated luminosity of 36.1 fb^{-1}. In this search, am_{T2} is used as the main discriminating variable, which aims to reconstruct the top quark mass in dileptonic $t\bar{t}$ events. The discriminating character of the am_{T2} variable between supersymmetric signal events and dileptonic $t\bar{t}$ events is due to the small mass difference between the top squark and the neutralino, leading to smaller am_{T2} values for signal events. Since no significant excess in data upon the SM prediction is found, exclusion limits at 95 % confidence level for the targeted simplified model are determined by a shape-fit approach in the am_{T2} variable. For the shape-fit configuration, 5 disjoint signal regions in the interval from 80 GeV to 130 GeV in am_{T2} are included. By exploiting the different shapes of the signal and background distribution in the shape-fit, the background could be further constrained which results in a better exclusion sensitivity. In the region where the 3-body decay scenario is kinematically favoured, top squark masses up to 460 GeV are excluded, under the assumption of $\mathcal{B}(\tilde{t}_1 \rightarrow bW\tilde{\chi}_1^0) = 100\%$.

The second analysis discussed in this thesis is performed on the full Run 2 dataset of pp collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV within the period from 2015 to 2018, corresponding to an integrated luminosity of 139 fb^{-1}. The analysis strategy of this search is based on an innovative machine learning classifier in order to separate SUSY signal events from SM background events, because a conventional signal region based on am_{T2} indicated less discriminating power for signal models with higher top squark masses. The ML classifier is based on a neural network which is trained with several kinematic observables, such as the lepton 4-vector, jet 4-vectors and $E_{\mathrm{T}}^{\mathrm{miss}}$. The result from this search do not hint at any evidence for new physics phenomena. The observed events are in agreement with the events from the SM prediction. As a consequence, the results are interpreted as exclusion limits at 95% confidence level for the considered simplified model. In the considered phase space, which is characterised by the mass splitting $m(W)+m(b) \leq \Delta m(\tilde{t}_1, \tilde{\chi}_1^0) \leq m(t)$, the observed exclusion limits improve upon previous exclusion limits. Top squark masses up to 720 GeV for neutralino masses up to 580 GeV are excluded in the context of simplified models assuming a 100% decay branching ratio of $\tilde{t}_1 \rightarrow bW\tilde{\chi}_1^0$.

At this time, maintenance work is carried out at the LHC and the collider experiments in preparation for the next data taking period. The recorded dataset from Run 2 is currently intensively investigated by many scientists and many exciting high precision measurements and results from various searches beyond the standard model are yet to come.

Only a small fraction of data has been analysed in this work compared to the total expected data which will be delivered by the Large Hadron Collider in the next two decades. Also the centre-of-mass energy of proton-proton collisions will be increased to 14 TeV. Due to these reasons, searches of new physics phenomena beyond the standard model are more exciting than ever. It also opens up the possibility for more exotic and unconventional searches and advanced analysis strategies, such as shape-fit configurations and machine learning techniques. These innovative methods are essential in order to benefit from the incredible amount of data that lies ahead of us.

Bibliography

[1] S. L. Glashow. “Partial-symmetries of weak interactions”. *Nuclear Physics*, 22(4): 579 – 588, 1961. doi:10.1016/0029-5582(61)90469-2.

[2] S. Weinberg. “A Model of Leptons”. *Phys. Rev. Lett.*, 19: 1264–1266, 1967. doi:10.1103/PhysRevLett.19.1264.

[3] A. Salam. “Gauge Unification of Fundamental Forces”. *Rev. Mod. Phys.*, 52: 525–538, 1980. doi:10.1103/RevModPhys.52.525.

[4] A. Salam and J. C. Ward. “Electromagnetic and weak interactions”. *Phys. Lett.*, 13: 168–171, 1964. doi:10.1016/0031-9163(64)90711-5.

[5] The Particle Data Group Collaboration. “Review of particle physics”. *Phys. Rev. D*, 98: 030001, 2018. doi:10.1103/PhysRevD.98.030001.

[6] The Super-Kamiokande Collaboration. “Evidence for oscillation of atmospheric neutrinos”. *Phys. Rev. Lett.*, 81: 1562–1567, 1998. doi:10.1103/PhysRevLett.81.1562.

[7] The SNO Collaboration. “Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory”. *Phys. Rev. Lett.*, 87: 071301, 2001. doi:10.1103/PhysRevLett.87.071301.

[8] The LHCb Collaboration. “Observation of J/ψp Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^-$p Decays”. *Phys. Rev. Lett.*, 115: 072001, 2015. doi:10.1103/PhysRevLett.115.072001.

[9] M. Srednicki. *Quantum Field Theory.* Cambridge Univ. Press, 2007.

[10] D. Griffiths. *Introduction to elementary particles.* Wiley-VCH, 2008.

[11] I. J. R. Aitchison and A. J. G. Hey. *Gauge theories in particle physics: a practical introduction: Volume 1 From relativistic quantum mechanics to QED; 4th ed.* CRC Press, 2013.

[12] I. J. R. Aitchison and A. J. G. Hey. *Gauge theories in particle physics: a practical introduction: Volume 2 Non-Abelian gauge theories QCD and the electroweak theory; 4th ed.* CRC Press, 2013.

[13] M. Thomson. *Modern particle physics.* Cambridge University Press, 2013.

[14] G. ’t Hooft. “Renormalizable Lagrangians for Massive Yang-Mills Fields”. *Nucl. Phys.*, B35: 167–188, 1971. doi:10.1016/0550-3213(71)90139-8.

[15] M. Creutz. *Quarks, gluons and lattices.* Cambridge Monographs on Mathematical Physics. Cambridge Univ. Press, 1985.

[16] S. M. Bilenky and J. Hosek. “Glashow-Weinberg-Salam theory of electroweak interactions and the neutral currents”. *Phys. Rept.*, 90: 73–157, 1982. doi:10.1016/0370-1573(82)90016-3.

[17] M. Goldhaber, L. Grodzins, and A. W. Sunyar. “Helicity of neutrinos”. *Phys. Rev.*, 109: 1015–1017, 1958. doi:10.1103/PhysRev.109.1015.

[18] T. Nakano and K. Nishijima. “Charge independence for v-particles*”. *Progress of Theoretical Physics*, 10(5): 581–582, 1953. doi:10.1143/PTP.10.581.

[19] K. Nishijima. “Charge independence theory of v particles*”. *Progress of Theoretical Physics*, 13(3): 285–304, 1955. doi:10.1143/PTP.13.285.

[20] M. Gell-Mann. “The interpretation of the new particles as displaced charge multiplets”. *Nuovo Cim.*, 4(S2): 848–866, 1956. doi:10.1007/BF02748000.

[21] F. Englert and R. Brout. “Broken symmetry and the mass of gauge vector mesons”. *Phys. Rev. Lett.*, 13: 321–323, 1964. doi:10.1103/PhysRevLett.13.321.

[22] P. W. Higgs. “Broken symmetries, massless particles and gauge fields”. *Phys. Lett.*, 12: 132–133, 1964. doi:10.1016/0031-9163(64)91136-9.

[23] P. W. Higgs. “Broken symmetries and the masses of gauge bosons”. *Phys. Rev. Lett.*, 13: 508–509, 1964. doi:10.1103/PhysRevLett.13.508.

[24] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global conservation laws and massless particles”. *Phys. Rev. Lett.*, 13: 585–587, 1964. doi:10.1103/PhysRevLett.13.585.

[25] ATLAS and CMS Collaboration. “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments”. *Phys. Rev. Lett.*, 114: 191803, 2015. doi:10.1103/PhysRevLett.114.191803.

[26] 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: 1–29, 2012.

[27] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. *Physics Letters B*, 716: 30–61, 2012.

[28] S. P. Martin. “A Supersymmetry primer”. *Adv. Ser. Direct. High Energy Phys.*, 18: 1, 1998. doi:10.1142/9789812839657_0001, 10.1142/9789814307505_0001.

[29] E. Witten. “Dynamical Breaking of Supersymmetry”. *Nucl. Phys.*, B188: 513, 1981. doi:10.1016/0550-3213(81)90006-7.

[30] M. J. G. Veltman. “The Infrared - Ultraviolet Connection”. *Acta Phys. Polon.*, B12: 437, 1981.

[31] R. K. Kaul. “Gauge Hierarchy in a Supersymmetric Model”. *Phys. Lett.*, 109B: 19–24, 1982. doi:10.1016/0370-2693(82)90453-1.

[32] F. Zwicky. “On the Masses of Nebulae and of Clusters of Nebulae”. *Astrophys. J.*, 86: 217–246, 1937. doi:10.1086/143864.

[33] The Planck Collaboration. “Planck 2018 results. VI. Cosmological parameters”. arXiv.org e-print archive, `arXiv:astro-ph.CO/1807.06209`, 2018.

[34] P. J. E. Peebles. “Tests of Cosmological Models Constrained by Inflation”. *Astrophys. J.*, 284: 439–444, 1984. doi:10.1086/162425.

[35] H. Goldberg. “Constraint on the Photino Mass from Cosmology”. *Phys. Rev. Lett.*, 50: 1419, 1983. doi:10.1103/PhysRevLett.50.1419.

[36] J. R. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. A. Olive, and M. Srednicki. “Supersymmetric Relics from the Big Bang”. *Nucl. Phys. B*, 238: 453, 1984. doi:10.1016/0550-3213(84)90461-9.

[37] P. Langacker. “Grand Unified Theories and Proton Decay”. *Phys. Rept.*, 72: 185, 1981. doi:10.1016/0370-1573(81)90059-4.

[38] N. Sakai. “Naturalness in Supersymmetric Guts”. *Z. Phys. C*, 11: 153, 1981. doi:10.1007/BF01573998.

[39] S. Dimopoulos, S. Raby, and F. Wilczek. “Supersymmetry and the Scale of Unification”. *Phys. Rev. D*, 24: 1681, 1981. doi:10.1103/PhysRevD.24.1681.

[40] L. E. Ibanez and G. G. Ross. “Low-Energy Predictions in Supersymmetric Grand Unified Theories”. *Phys. Lett. B*, 105: 439, 1981. doi:10.1016/0370-2693(81)91200-4.

[41] S. Dimopoulos and H. Georgi. “Softly Broken Supersymmetry and SU(5)”. *Nucl. Phys. B*, 193: 150, 1981. doi:10.1016/0550-3213(81)90522-8.

[42] Yu. A. Golfand and E. P. Likhtman. “Extension of the Algebra of Poincare Group Generators and Violation of p Invariance”. *JETP Lett.*, 13: 323, 1971.

[43] D. V. Volkov and V. P. Akulov. “Is the Neutrino a Goldstone Particle?”. *Phys. Lett. B*, 46: 109, 1973. doi:10.1016/0370-2693(73)90490-5.

[44] J. Wess and B. Zumino. “Supergauge Transformations in Four-Dimensions”. *Nucl. Phys. B*, 70: 39, 1974. doi:10.1016/0550-3213(74)90355-1.

[45] J. Wess and B. Zumino. “Supergauge Invariant Extension of Quantum Electrodynamics”. *Nucl. Phys. B*, 78: 1, 1974. doi:10.1016/0550-3213(74)90112-6.

[46] P. Fayet. “Supersymmetry and Weak, Electromagnetic and Strong Interactions”. *Phys. Lett. B*, 64: 159, 1976. doi:10.1016/0370-2693(76)90319-1.

[47] P. Fayet. “Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions”. *Phys. Lett. B*, 69: 489, 1977. doi:10.1016/0370-2693(77)90852-8.

[48] G. R. Farrar and P. Fayet. “Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry”. *Phys. Lett. B*, 76: 575, 1978. doi:10.1016/0370-2693(78)90858-4.

[49] L. Girardello and M. T. Grisaru. “Soft Breaking of Supersymmetry”. *Nucl. Phys. B*, 194: 65, 1982. doi:10.1016/0550-3213(82)90512-0.

[50] K. Inoue, A. Kakuto, H. Komatsu, and S. Takeshita. “Aspects of Grand Unified Models with Softly Broken Supersymmetry”. *Prog. Theor. Phys.*, 68: 927, 1982. doi:10.1143/PTP.68.927.

[51] J. R. Ellis and S. Rudaz. “Search for Supersymmetry in Toponium Decays”. *Phys. Lett. B*, 128: 248, 1983. doi:10.1016/0370-2693(83)90402-1.

[52] J. Alwall, M.-P. Le, M. Lisanti, and J. G. Wacker. “Searching for Directly Decaying Gluinos at the Tevatron”. *Phys. Lett. B*, 666: 34, 2008. doi:10.1016/j.physletb.2008.06.065.

[53] J. Alwall, P. Schuster, and N. Toro. “Simplified Models for a First Characterization of New Physics at the LHC”. *Phys. Rev. D*, 79: 075020, 2009. doi:10.1103/PhysRevD.79.075020.

[54] D. Alves et al. “Simplified Models for LHC New Physics Searches”. *J. Phys. G*, 39: 105005, 2012. doi:10.1088/0954-3899/39/10/105005.

[55] ATLAS Collaboration. “Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV *pp* collisions with the ATLAS detector”. *Phys. Rev.*, D94(5): 052009, 2016. doi:10.1103/PhysRevD.94.052009.

[56] ATLAS Collaboration. “Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector”. *JHEP*, 01: 126, 2018. doi:10.1007/JHEP01(2018)126.

[57] ATLAS Collaboration. “Search for a scalar partner of the top quark in the jets plus missing transverse momentum final state at $\sqrt{s}$=13 TeV with the ATLAS detector”. *JHEP*, 12: 085, 2017. doi:10.1007/JHEP12(2017)085.

[58] ATLAS Collaboration. “Search for top-squark pair production in final states with one lepton, jets, and missing transverse momentum using 36 fb^{-1} of $\sqrt{s} =$ 13 TeV pp collision data with the ATLAS detector”. *JHEP*, 06: 108, 2018. doi:10.1007/JHEP06(2018)108.

[59] ATLAS Collaboration. “Search for direct top squark pair production in final states with two leptons in $\sqrt{s} = 13$ TeV *pp* collisions with the ATLAS detector”. *Eur. Phys. J.*, C77(12): 898, 2017. doi:10.1140/epjc/s10052-017-5445-x.

[60] ATLAS Collaboration. “Search for supersymmetry in final states with charm jets and missing transverse momentum in 13 TeV *pp* collisions with the ATLAS detector”. *JHEP*, 09: 050, 2018. doi:10.1007/JHEP09(2018)050.

[61] CMS Collaboration. “Search for top squark pair production in pp collisions at $\sqrt{s} = 13$ TeV using single lepton events”. *JHEP*, 10: 019, 2017. doi:10.1007/JHEP10(2017)019.

[62] CMS Collaboration. “Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JHEP*, 10: 005, 2017. doi:10.1007/JHEP10(2017)005.

[63] CMS Collaboration. “Search for top squarks decaying via four-body or chargino-mediated modes in single-lepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JHEP*, 09: 065, 2018. doi:10.1007/JHEP09(2018)065.

[64] ATLAS Collaboration. "Summary of the dedicated ATLAS searches for top squark pair production based on 36.1 fb^{-1} of *pp* collision data taken at a centre-of-mass energy 13 TeV". `http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SUSY/ATLAS_SUSY_Stop_tLSP/ATLAS_SUSY_Stop_tLSP.pdf`, 2018, Accessed: May 2019.

[65] L. R. Evans and P. Bryant. "LHC Machine". *JINST*, 3: S08001. 164 p, 2008.

[66] E. A. Mobs. "The CERN accelerator complex. Complexe des accélérateurs du CERN". 2016. General Photo.

[67] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl. *LHC Design Report: Volume 3, The LHC injector chain.* CERN Yellow Reports: Monographs. CERN, 2004.

[68] ALICE Collaboration. "The ALICE experiment at the CERN LHC. A Large Ion Collider Experiment". *JINST*, 3: S08002. 259 p, 2008.

[69] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". *JINST*, 3: S08003. 437 p, 2008.

[70] CMS Collaboration. "The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment". *JINST*, 3: S08004. 361 p, 2008.

[71] LHCb Collaboration. "The LHCb Detector at the LHC". *JINST*, 3(LHCb-DP-2008-001. CERN-LHCb-DP-2008-001): S08005, 2008.

[72] T. Schoerner-Sadenius. *The Large Hadron Collider : Harvest of Run 1.* Springer International Publishing, 2015.

[73] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, et al. *LHC Design Report: Volume 1, The LHC main ring.* CERN Yellow Reports: Monographs. CERN, 2004.

[74] ATLAS Collaboration. "ATLAS Public Luminosity Results for Run 2". `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2`, Accessed: May 2019.

[75] ATLAS Collaboration. "ATLAS Public Luminosity Results for Run 1". `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults`, Accessed: May 2019.

[76] ATLAS Collaboration. *ATLAS detector and physics performance: Technical Design Report, 1.* Technical Design Report ATLAS. CERN, 1999.

[77] ATLAS Collaboration. *ATLAS detector and physics performance: Technical Design Report, 2.* Technical Design Report ATLAS. CERN, 1999.

[78] ATLAS Collaboration. "Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System". Technical Report CERN-LHCC-2013-018. ATLAS-TDR-023, 2013. `http://cds.cern.ch/record/1602235`.

[79] J. Pequenao. "Computer generated image of the whole ATLAS detector". 2008.

[80] J. Pequenao and P. Schaffner. "An computer generated image representing how ATLAS detects particles". 2013.

[81] ATLAS Collaboration. “ATLAS Insertable B-Layer Technical Design Report”. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, 2010. `https://cds.cern.ch/record/1291633`.

[82] ATLAS Collaboration. “Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2”. *Eur. Phys. J.*, C77(10): 673, 2017. doi:10.1140/epjc/s10052-017-5225-7.

[83] J. Pequenao. “Computer Generated image of the ATLAS calorimeter”. 2008.

[84] J. Pequenao. “Computer generated image of the ATLAS Muons subsystem”. 2008.

[85] ATLAS Collaboration. “Performance of the ATLAS Trigger System in 2015”. *Eur. Phys. J.*, C77(5): 317, 2017. doi:10.1140/epjc/s10052-017-4852-3.

[86] ATLAS Collaboration. “Improved luminosity determination in pp collisions at $\sqrt{s} =$ 7 TeV using the ATLAS detector at the LHC”. *Eur. Phys. J.*, C73(8): 2518, 2013. doi:10.1140/epjc/s10052-013-2518-3.

[87] ATLAS Collaboration. “Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC”. *Eur. Phys. J.*, C76(12): 653, 2016. doi:10.1140/epjc/s10052-016-4466-1.

[88] G. Avoni et al. “The new lucid-2 detector for luminosity measurement and monitoring in atlas”. *JINST*, 13(07): P07017, 2018. doi:10.1088/1748-0221/13/07/P07017.

[89] A. Buckley et al. “General-purpose event generators for LHC physics”. *Phys. Rept.*, 504: 145–233, 2011. doi:10.1016/j.physrep.2011.03.005.

[90] V. N. Gribov and L. N. Lipatov. “Deep inelastic e p scattering in perturbation theory”. *Sov. J. Nucl. Phys.*, 15: 438–450, 1972.

[91] Y. L. Dokshitzer. “Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.”. *Sov. Phys. JETP*, 46: 641–653, 1977.

[92] G. Altarelli and G. Parisi. “Asymptotic Freedom in Parton Language”. *Nucl. Phys.*, B126: 298–318, 1977. doi:10.1016/0550-3213(77)90384-4.

[93] J. Rojo et al. “The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II”. *J. Phys.*, G42: 103103, 2015. doi:10.1088/0954-3899/42/10/103103.

[94] J. C. Collins, D. E. Soper, and G. F. Sterman. “Factorization of Hard Processes in QCD”. *Adv. Ser. Direct. High Energy Phys.*, 5: 1–91, 1989. doi:10.1142/9789814503266_0001.

[95] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. “Parton Fragmentation and String Dynamics”. *Phys. Rept.*, 97: 31–145, 1983. doi:10.1016/0370-1573(83)90080-7.

[96] B. Andersson. “The Lund model”. *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, 7: 1–471, 1997.

[97] D. Amati and G. Veneziano. “Preconfinement as a Property of Perturbative QCD”. *Phys. Lett.*, 83B: 87–92, 1979. doi:10.1016/0370-2693(79)90896-7.

[98] The GEANT4 Collaboration. “GEANT4: A Simulation toolkit”. *Nucl. Instrum. Meth.*, A506: 250–303, 2003. doi:10.1016/S0168-9002(03)01368-8.

[99] ATLAS Collaboration. “The ATLAS Simulation Infrastructure.”. *Eur. Phys. J. C*, 70(arXiv:1005.4568. CERN-PH-EP-2010-044): 823–874. 53 p, 2010.

[100] S. Alioli, P. Nason, C. Oleari, and E. Re. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. *JHEP*, 1006: 043, 2010. doi:10.1007/JHEP06(2010)043.

[101] E. Re. “Single-top Wt-channel production matched with parton showers using the POWHEG method”. *Eur. Phys. J. C*, 71: 1547, 2011. doi:10.1140/epjc/s10052-011-1547-z.

[102] S. Frixione, P. Nason, and G. Ridolfi. “A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”. *JHEP*, 09: 126, 2007. doi:10.1088/1126-6708/2007/09/126.

[103] R. Frederix, E. Re, and P. Torrielli. “Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO”. *JHEP*, 09: 130, 2012. doi:10.1007/JHEP09(2012)130.

[104] S. Alioli, P. Nason, C. Oleari, and E. Re. “NLO single-top production matched with shower in POWHEG: s- and t-channel contributions”. *JHEP*, 09: 111, 2009. doi:10.1007/JHEP02(2010)011, 10.1088/1126-6708/2009/09/111.

[105] Gleisberg, T. and Höche, Stefan. and Krauss, F. and Schönherr, M. and Schumann, S. and others. “Event generation with SHERPA 1.1”. *JHEP*, 0902: 007, 2009. doi:10.1088/1126-6708/2009/02/007.

[106] T. Gleisberg and S. Höche. “Comix, a new matrix element generator”. *JHEP*, 12: 039, 2008. doi:10.1088/1126-6708/2008/12/039.

[107] F. Cascioli, P. Maierhofer, and S. Pozzorini. “Scattering Amplitudes with Open Loops”. *Phys. Rev. Lett.*, 108: 111601, 2012. doi:10.1103/PhysRevLett.108.111601.

[108] S. Schumann and F. Krauss. “A Parton shower algorithm based on Catani-Seymour dipole factorisation”. *JHEP*, 03: 038, 2008. doi:10.1088/1126-6708/2008/03/038.

[109] S. Höche, F. Krauss, M. Schönherr, and F. Siegert. “QCD matrix elements + parton showers: The NLO case”. *JHEP*, 04: 027, 2013. doi:10.1007/JHEP04(2013)027.

[110] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. *JHEP*, 1407: 079, 2014. doi:10.1007/JHEP07(2014)079.

[111] ATLAS Collaboration. “Simulation of top-quark production for the ATLAS experiment at $\sqrt{s} = 13$ TeV”. ATL-PHYS-PUB-2016-004, 2016.

[112] ATLAS Collaboration. “Monte Carlo Generators for the Production of a W or Z/γ^* Boson in Association with Jets at ATLAS in Run 2”. ATL-PHYS-PUB-2016-003, 2016.

[113] ATLAS Collaboration. “Multi-boson simulation for 13 TeV ATLAS analyses”. ATL-PHYS-PUB-2016-002, 2016.

[114] ATLAS Collaboration. “Modelling of the $t\bar{t}H$ and $t\bar{t}V(V = W, Z)$ processes for $\sqrt{s} = 13$ TeV ATLAS analyses”. ATL-PHYS-PUB-2016-005, 2016.

[115] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk. “Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations”. *JHEP*, 03: 015, 2013. doi:10.1007/JHEP03(2013)015.

[116] Ian Low. “Polarized Charginos (and Tops) in Stop Decays”. *Phys. Rev. D*, 88, 2013. doi:10.1103/PhysRevD.88.095018.

[117] Maxim Perelstein, Andreas Weiler. “Polarized Tops from Stop Decays at the LHC”. *JHEP*, 2009, 2009. doi:10.1088/1126-6708/2009/03/141.

[118] W. Beenakker, M. Kramer, T. Plehn, M. Spira, and P. M. Zerwas. “Stop production at hadron colliders”. *Nucl. Phys. B*, 515: 3, 1998.

[119] W. Beenakker, S. Brensing, M. Kramer, A. Kulesza, E. Laenen, et al. “Supersymmetric top and bottom squark production at hadron colliders”. *JHEP*, 08: 098, 2010. doi:10.1007/JHEP08(2010)098.

[120] W. Beenakker, S. Brensing, M. Kramer, A. Kulesza, E. Laenen, et al. “Squark and gluino hadroproduction”. *Int. J. Mod. Phys. A*, 26: 2637, 2011. doi:10.1142/S0217751X11053560.

[121] C. Borschensky, M. Kr??mer, A. Kulesza, M. Mangano, S. Padhi, et al. “Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13$, 14, 33 and 100 TeV”. *Eur. Phys. J. C*, 74(12): 3174, 2014. doi:10.1140/epjc/s10052-014-3174-y.

[122] ATLAS Collaboration. “The ATLAS Simulation Infrastructure”. *Eur. Phys. J. C*, 70: 823, 2010. doi:10.1140/epjc/s10052-010-1429-9.

[123] D. J. Lange. “The EvtGen particle decay simulation package”. *Nucl. Instrum. Meth. A*, 462: 152, 2001. doi:10.1016/S0168-9002(01)00089-4.

[124] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, et al. “New parton distributions for collider physics”. *Phys. Rev. D*, 82: 074024, 2010. doi:10.1103/PhysRevD.82.074024.

[125] T. Sjöstrand, S. Mrenna, and P. Z. Skands. “PYTHIA 6.4 Physics and Manual”. *JHEP*, 0605: 026, 2006. doi:10.1088/1126-6708/2006/05/026.

[126] P. Z. Skands. “Tuning Monte Carlo Generators: The Perugia Tunes”. *Phys. Rev. D*, 82: 074018, 2010. doi:10.1103/PhysRevD.82.074018.

[127] M. Czakon, P. Fiedler, and A. Mitov. “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$”. *Phys. Rev. Lett.*, 110: 252004, 2013. doi:10.1103/PhysRevLett.110.252004.

[128] M. Czakon and A. Mitov. “NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction”. *JHEP*, 1301: 080, 2013. doi:10.1007/JHEP01(2013)080.

[129] M. Czakon and A. Mitov. “NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels”. *JHEP*, 1212: 054, 2012. doi:10.1007/JHEP12(2012)054.

[130] P. Bärnreuther, M. Czakon, and A. Mitov. “Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$”. *Phys. Rev. Lett.*, 109: 132001, 2012. doi:10.1103/PhysRevLett.109.132001.

[131] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason. “Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation”. *Phys. Lett. B*, 710: 612–622, 2012. doi:10.1016/j.physletb.2012.03.013.

[132] M. Czakon and A. Mitov. “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders”. *Comput. Phys. Commun.*, 185: 2930, 2014. doi:10.1016/j.cpc.2014.06.021.

[133] N. Kidonakis. “Next-to-next-to-leading-order collinear and soft gluon corrrections for t-channel single top quark production”. *Phys. Rev. D*, 83: 091503, 2011. doi:10.1103/PhysRevD.83.091503.

[134] N. Kidonakis. “Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-”. *Phys. Rev. D*, 82: 054018, 2010. doi:10.1103/PhysRevD.82.054018.

[135] N. Kidonakis. “NNLL resummation for s-channel single top quark production”. *Phys. Rev. D*, 81: 054028, 2010. doi:10.1103/PhysRevD.81.054028.

[136] The NNPDF Collaboration. “Parton distributions for the LHC Run II”. *JHEP*, 1504: 040, 2015. doi:10.1007/JHEP04(2015)040.

[137] S. Catani, L. Cieri, G. Ferrera, D. de Florian, and M. Grazzini. “Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO”. *Phys. Rev. Lett.*, 103: 082001, 2009. doi:10.1103/PhysRevLett.103.082001.

[138] T. Sjöstrand, S. Mrenna, and P. Z. Skands. “A Brief Introduction to PYTHIA 8.1”. *Comput. Phys. Commun.*, 178: 852, 2008. doi:10.1016/j.cpc.2008.01.036.

[139] ATLAS Collaboration. “ATLAS Run 1 Pythia8 tunes”. Technical Report ATL-PHYS-PUB-2014-021, 2014. `http://cds.cern.ch/record/1966419`.

[140] R. D. Ball et al. “Parton distributions with LHC data”. *Nucl. Phys. B*, 867: 244–289, 2013. doi:10.1016/j.nuclphysb.2012.10.003.

[141] W. Beenakker, C. Borschensky, M. Krämer, A. Kulesza, and E. Laenen. “NNLL-fast: predictions for coloured supersymmetric particle production at the LHC with threshold and Coulomb resummation”. *JHEP*, 12: 133, 2016. doi:10.1007/JHEP12(2016)133.

[142] W. Beenakker, C. Borschensky, R. Heger, M. Krämer, A. Kulesza, et al. “NNLL resummation for stop pair-production at the LHC”. *JHEP*, 05: 153, 2016. doi:10.1007/JHEP05(2016)153.

[143] J. Butterworth et al. “PDF4LHC recommendations for LHC Run II”. *J. Phys. G*, 43: 023001, 2016. doi:10.1088/0954-3899/43/2/023001.

[144] ATLAS Collaboration. “Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment”. Technical Report ATLAS-CONF-2012-042, 2012. `http://cds.cern.ch/record/1435196`.

[145] ATLAS Collaboration. “Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2”. *Eur. Phys. J.*, C77(10): 673, 2017. doi:10.1140/epjc/s10052-017-5225-7.

[146] R. Fruehwirth. “Application of Kalman filtering to track and vertex fitting”. *Nucl. Instrum. Meth.*, A262: 444–450, 1987. doi:10.1016/0168-9002(87)90887-4.

[147] T. Cornelissen, M. Elsing, I. Gavrilenko, W. Liebig, E. Moyse, et al. “The new ATLAS track reconstruction (NEWT)”. *J. Phys. Conf. Ser.*, 119: 032014, 2008. doi:10.1088/1742-6596/119/3/032014.

[148] G. Piacquadio, K. Prokofiev, and A. Wildauer. “Primary vertex reconstruction in the ATLAS experiment at LHC”. *J. Phys. Conf. Ser.*, 119: 032033, 2008. doi:10.1088/1742-6596/119/3/032033.

[149] ATLAS Collaboration. “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC”. *Eur. Phys. J.*, C77(5): 332, 2017. doi:10.1140/epjc/s10052-017-4887-5.

[150] ATLAS Collaboration. “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV”. *Submitted to: Eur. Phys. J.*, 2019.

[151] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, et al. “Calorimeter Clustering Algorithms: Description and Performance”. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, 2008. `https://cds.cern.ch/record/1099735`.

[152] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using 2015-2016 LHC proton-proton collision data”. *Submitted to: JINST*, 2018.

[153] ATLAS Collaboration. “Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data”. Technical Report ATLAS-CONF-2016-024, 2016. `https://cds.cern.ch/record/2157687`.

[154] ATLAS Collaboration. “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV”. *Eur. Phys. J.*, C76(5): 292, 2016. doi:10.1140/epjc/s10052-016-4120-y.

[155] ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. *Eur. Phys. J.*, C77: 490, 2017. doi:10.1140/epjc/s10052-017-5004-5.

[156] M. Cacciari, G. P. Salam, and G. Soyez. “The anti-k_t jet clustering algorithm”. *JHEP*, 04: 063, 2008. doi:10.1088/1126-6708/2008/04/063.

[157] M. Cacciari, G. P. Salam, and G. Soyez. “The Catchment Area of Jets”. *JHEP*, 04: 005, 2008. doi:10.1088/1126-6708/2008/04/005.

[158] M. Cacciari and G. P. Salam. “Pileup subtraction using jet areas”. *Phys. Lett.*, B659: 119–126, 2008. doi:10.1016/j.physletb.2007.09.077.

[159] ATLAS Collaboration. “Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. *Phys. Rev.*, D96(7): 072002, 2017. doi:10.1103/PhysRevD.96.072002.

[160] ATLAS Collaboration. “Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector”. *Eur. Phys. J.*, C76(11): 581, 2016. doi:10.1140/epjc/s10052-016-4395-z.

[161] ATLAS Collaboration. “Expected performance of the ATLAS b-tagging algorithms in Run-2”. Technical Report ATL-PHYS-PUB-2015-022, 2015. `https://cds.cern.ch/record/2037697`.

[162] ATLAS Collaboration. “Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run”. Technical Report ATL-PHYS-PUB-2016-012, 2016. `https://cds.cern.ch/record/2160731`.

[163] ATLAS Collaboration. “Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV”. *JHEP*, 08: 089, 2018. doi:10.1007/JHEP08(2018)089.

[164] ATLAS Collaboration. “Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC”. Technical Report ATL-PHYS-PUB-2015-045, 2015. `http://cds.cern.ch/record/2064383`.

[165] ATLAS Collaboration. “Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV”. Technical Report ATLAS-CONF-2017-029, 2017. `https://cds.cern.ch/record/2261772`.

[166] ATLAS Collaboration. “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV”. *Eur. Phys. J. C*, 78(arXiv:1802.08168. 11): 903. 66 p, 2018.

[167] ATLAS Collaboration. “$E_{\mathrm{T}}^{\mathrm{miss}}$ performance in the ATLAS detector using 2015-2016 LHC p-p collisions”. Technical Report ATLAS-CONF-2018-023, 2018. `http://cds.cern.ch/record/2625233`.

[168] C. G. Lester and D. J. Summers. “Measuring masses of semi-invisibly decaying particles pair produced at hadron colliders”. *Phys. Lett. B*, 463: 99–103, 1999. doi:10.1016/S0370-2693(99)00945-4.

[169] A. J. Barr, B. Gripaios, and C. G. Lester. “Transverse masses and kinematic constraints: from the boundary to the crease”. *JHEP*, 0911: 096, 2009. doi:10.1088/1126-6708/2009/11/096.

[170] P. Konar, K. Kong, K. T. Matchev, and M. Park. “Dark Matter Particle Spectroscopy at the LHC: Generalizing M(T2) to Asymmetric Event Topologies”. *JHEP*, 1004: 086, 2010. doi:10.1007/JHEP04(2010)086.

[171] Y. Bai, H.-C. Cheng, J. Gallicchio, and J. Gu. “Stop the Top Background of the Stop Search”. *JHEP*, 1207: 110, 2012. doi:10.1007/JHEP07(2012)110.

[172] C. G. Lester and B. Nachman. “Bisection-based asymmetric M_{T2} computation: a higher precision calculator than existing symmetric methods”. *JHEP*, 1503: 100, 2015. doi:10.1007/JHEP03(2015)100.

[173] ATLAS Collaboration. “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”. *JHEP*, 11: 118, 2014. doi:10.1007/JHEP11(2014)118.

[174] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[175] Goodfellow, Ian and Bengio, Yoshua and Courville, Aaron. *Deep Learning*. MIT Press, 2006.

[176] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. “What is the best multi-stage architecture for object recognition?”. In *ICCV*, pages 2146–2153. IEEE, 2009.

[177] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. “Learning representations by back-propagating errors”. *Nature*, 323: 533–, 1986.

[178] A. Graves. *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer, 2011.

[179] S. Hochreiter. *Untersuchungen zu dynamischen neuronalen Netzen*. PhD thesis, Institut für Informatik, Lehrstuhl Prof. Brauer, Technische Universität München, 1991.

[180] Y. Bengio, P. Simard, and P. Frasconi. “Learning long-term dependencies with gradient descent is difficult”. *IEEE Transactions on Neural Networks*, 5(2): 157–166, 1994.

[181] S. Hochreiter and J. Schmidhuber. “Long short-term memory”. *Neural computation*, 9: 1735–80, 1997. doi:10.1162/neco.1997.9.8.1735.

[182] P. Y. Simard and D. a. Steinkraus. “Best practices for convolutional neural networks applied to visual document analysis”. Institute of Electrical and Electronics Engineers, Inc., 2003.

[183] B. Polyak. “Some methods of speeding up the convergence of iteration methods”. *Ussr Computational Mathematics and Mathematical Physics*, 4: 1–17, 1964. doi:10.1016/0041-5553(64)90137-5.

[184] J. Duchi, E. Hazan, and Y. Singer. “Adaptive subgradient methods for online learning and stochastic optimization”. *Journal of Machine Learning Research*, 12(Jul): 2121–2159, 2011.

[185] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization”. *CoRR*, abs/1412.6980, 2014.

[186] M. Baak, G. J. Besjes, D. Côté, A. Koutsman, J. Lorenz, et al. “HistFitter software framework for statistical data analysis”. *The European Physical Journal C*, 75(4): 153, 2015. doi:10.1140/epjc/s10052-015-3327-7.

[187] The ROOT Collaboration. “HistFactory: A tool for creating statistical models for use with RooFit and RooStats”. Technical Report CERN-OPEN-2012-016, 2012. `https://cds.cern.ch/record/1456844`.

[188] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. “Asymptotic formulae for likelihood-based tests of new physics”. *Eur. Phys. J.*, C71: 1554, 2011. doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z.

[189] A. L. Read. “Presentation of search results: The CL(s) technique”. *J. Phys.*, G28: 2693–2704, 2002. doi:10.1088/0954-3899/28/10/313.

[190] ATLAS Collaboration. “Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV”. Technical Report ATL-PHYS-PUB-2015-015, 2015. `https://cds.cern.ch/record/2037613`.

[191] ATLAS Collaboration. “A method for the construction of strongly reduced representations of ATLAS experimental uncertainties and the application thereof to the jet energy scale”. Technical Report ATL-PHYS-PUB-2015-014, 2015. `https://cds.cern.ch/record/2037436`.

[192] ATLAS Collaboration. “Measurement of b-tagging Efficiency of c-jets in$t\bar{t}$ Events Using a Likelihood Approach with the ATLAS Detector”. Technical Report ATLAS-CONF-2018-001, 2018. `https://cds.cern.ch/record/2306649`.

[193] ATLAS Collaboration. “Calibration of light-flavour jet b-tagging rates on ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV”. Technical Report ATLAS-CONF-2018-006, 2018. `https://cds.cern.ch/record/2314418`.

[194] M. Bahr et al. “Herwig++ Physics and Manual”. *Eur. Phys. J.*, C58: 639–707, 2008. doi:10.1140/epjc/s10052-008-0798-9.

[195] J. Bellm et al. “Herwig 7.1 Release Note”. arXiv.org e-print archive, `arXiv:hep-ph/1705.06919`, 2017.

[196] L. Moneta, K. Belasco, K. S. Cranmer, S. Kreiss, A. Lazzaro, et al. “The RooStats Project”. *PoS*, ACAT2010: 057, 2010. doi:10.22323/1.093.0057.

[197] ATLAS Collaboration. “Search for direct top squark pair production in the 3-body decay mode with a final state containing one lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$TeV pp collision data with the ATLAS detector”. Technical Report ATLAS-CONF-2019-017, 2019. `http://cds.cern.ch/record/2676594`.

[198] F. Chollet et al. “Keras”. `https://keras.io`, 2015.

[199] M. Abadi et al. “Tensorflow: Large-scale machine learning on heterogeneous systems”. `https://www.tensorflow.org/`, 2015.

[200] X. Glorot and Y. Bengio. “Understanding the difficulty of training deep feedforward neural networks.”. In *AISTATS*, volume 9 of *JMLR Proceedings*, pages 249–256. JMLR.org, 2010.

[201] S. Ioffe and C. Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. *CoRR*, abs/1502.03167, 2015.

[202] R. D. Cousins, J. T. Linnemann, and J. Tucker. “Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process”. *Nuclear Instruments and Methods in Physics Research A*, 595: 480–501, 2008. doi:10.1016/j.nima.2008.07.086.

[203] CMS Collaboration. “Searches for new phenomena in events with jets and high values of the M_{T2} variable, including signatures with disappearing tracks, in proton-proton collisions at $\sqrt{s} = 13$ TeV”. Technical Report CMS-PAS-SUS-19-005, 2019. `https://cds.cern.ch/record/2668105`.

List of Figures

2.1 One-loop corrections to M_h^2. For a fermion f (left) and a scalar particle S (right) [28]. 12
2.2 Leading order Feynman diagrams for likely production mechanisms via the strong interaction from gluon-gluon and gluon-quark fusion at hadron colliders [28]. $\tilde{g}$ and $\tilde{g}^*$ represent the gluino and its conjugate, and similarly the squark and its conjugate are denoted as $\tilde{q}$ and $\tilde{q}^*$. 17
2.3 Simplified model topology of direct top squark pair production. The top squark directly decays to a top quark and the neutralino LSP. $\tilde{\chi}_1^0$ is predominantly bino-like. By courtesy of the ATLAS supersymmetry physics analysis group. 19
2.4 Simplified model topology of direct top squark pair production in a kinematically compressed phase space, referred to as 3-body decay scenario. The decay structure is characterised by $m(W) + m(b) < \Delta m < m(t)$. The top quark is produced off-shell and thus the top squarks directly decay to a b quark, W boson and the neutralino LSP. By courtesy of the ATLAS supersymmetry physics analysis group. 19
2.5 Simplified model topology of direct top squark pair production in a highly compressed phase space. The mass difference between the supersymmetric particles is so small ($\Delta m < m(W)+m(b)$), that the top quark as well as the W boson can only be produced off-shell and thus the top squarks decay via a 4-body decay mode. By courtesy of the ATLAS supersymmetry physics analysis group. 19
2.6 Illustration of the preferred top squark decay modes, assuming that the top squarks decay into a bino-like neutralino. The phase space of interest is spanned by the masses of the top squark ($m(\tilde{t}_1)$) and the lightest neutralino ($m(\tilde{\chi}_1^0)$) [55]. 20
2.7 Summary of the individual ATLAS searches for top squark pair production based on 36.1 fb^{-1} of pp collision data at a centre-of-mass energy 13 TeV [64]. Exclusion limits at 95 % confidence level are shown in the $m(\tilde{t}_1)$ versus $m(\tilde{\chi}_1^0)$ mass plane. The dashed and solid lines show the expected and observed limits, respectively. For the derived limits, statistical as well as systematic uncertainties except for the theoretical signal cross section uncertainty are included. Different decay modes are considered separately with a branching ratio of 100 %, respectively: $\tilde{t}_1 \to t\tilde{\chi}_1^0$, $\tilde{t}_1 \to bW\tilde{\chi}_1^0$, $\tilde{t}_1 \to bff'\tilde{\chi}_1^0$ and $\tilde{t}_1 \to c\tilde{\chi}_1^0$. The latter two decay modes are superimposed. 21

3.1 CERN accelerator complex 23
3.2 Integrated luminosity of stable beams for pp collisions delivered to ATLAS over the period of twelve months [74]. 26

3.3 Mean number of interactions per bunch crossing for Run 1 (a) and Run 2 (b) of pp collision data at 7, 8 and 13 TeV centre-of-mass energy. Data recorded by the ATLAS detector during stable beams is shown, and the total integrated luminosity and the mean number of interactions per bunch crossing $< \mu >$ per year are given in the figures [74,75]. 26
3.4 ATLAS detector layout [79]. 27
3.5 Cross-sectional view of one quarter of the ATLAS detector showing tracks and showers from different charged and uncharged particles [80]. 28
3.6 Layout of the inner detector system including the insertable B-Layer (IBL) implemented during the first long shut down, pixel and silicon microstrip tracker and the transition radiation tracker [82]. 30
3.7 Assembly of the electromagnetic and hadronic calorimeter system in the ATLAS experiment [83]. 32
3.8 Schematic layout of the ATLAS muon spectrometer consisting of high precision MDT and CSC chambers as well as RPC and TGC trigger chambers [84]. 34
3.9 Conceptual structure of the ATLAS trigger system showing the two distinct levels, hardware based L1 and software based HLT triggers [85]. 35

4.1 Parton distribution functions of the gluon, valence quarks and various sea quarks of the proton obtained in the NNLO NNPDF3.0 global analysis [93]. The left plot corresponds to hadronic scales ($\mu^2 = 10$ GeV^2) and the right plot corresponds to higher scales as present at the LHC ($\mu^2 = 10^4$ GeV^2) [5]. 39
4.2 Feynman diagrams of top quark pair production and decay. 41
4.3 Feynman diagrams for the production of single top quarks. 41
4.4 Illustration of the construction of am_{T2}, in order to discriminate dileptonic $t\bar{t}$ events with one lost lepton [173]. 55
4.5 Illustration of the construction of m_{T2}^{τ}, used to discriminate dileptonic $t\bar{t}$ events with a hadronically decaying τ lepton [173]. 55
4.6 Distributions of the discriminating variables after the preselection as described in Table 8.1 except for the threshold on m_{T} which is loosened to $m_{\mathrm{T}} \geq 30$ GeV. The object definitions from the `bWN-fullSet` analysis are used. The cross section of the SUSY signal is magnified in order to stress the discriminating power of the variable. 56

5.1 Basic structure of a multilayer perceptron, which consists of an input layer, an individual amount of hidden layers and an output layer. The various layers in turn are build up from single neurons, representing for example input variables of the data set or predicted outputs. 59
5.2 Examples of activation functions h used to transform activations a in a dedicated layer. 59
5.3 Basic concept of a recurrent neural network. In addition to standard connections between various layers, cyclic connections in hidden layers are allowed. As a consequence, a RNN is able to process not only the information from the current input neuron, but also from previous inputs. This means that sequential data as well as inputs of variable length can be handled. For simplicity the bias neurons are omitted. 61
5.4 Unfolding of a single neuron of a hidden layer in a RNN. The unfolded neuron does not necessarily need a connection to the output neuron at each sequence. Instead, only a single output neuron at the end of the sequence is also viable. 61

5.5 Stochastic gradient descent. The ellipses represent contours in the weight space. By iteratively updating the weights with respect to the actual gradient of the loss function, the trajectory is moving towards the minimum. . . 64

5.6 Loss diagnostics. 67

5.7 Confusion matrix. Signal and background events can be predicted correctly or misclassified, resulting in four categories, true positives, false positives, true negatives and false negatives. 67

5.8 Receiver operating characteristic. The ROC can be derived from the distribution of the output activation of signal and background events. 68

7.1 Relative contributions of the individual SM backgrounds after the preselection. 76

7.2 Kinematic distributions of key observables after the preselection is applied. The $t\bar{t}$ background prediction is scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross sections. The hatched band around the total SM prediction and the hatched band in the data/SM ratio include statistical and systematic uncertainties, except the $t\bar{t}$ modelling uncertainties, which are determined only for the VR and SR, respectively. Overflows are included in the last bin. 77

7.3 Kinematic distributions of the most important variables of the analysis after the preselection. Only simulated events are shown. For the distribution of m_{T}, the cut on the shown quantity is omitted in the event preselection in order to stress the discriminating power of this observable. Only statistical uncertainties are shown in the distributions. Overflows are included in the last bin. The distribution for signal events is scaled by multiples of its nominal cross section, to illustrate the distinct kinematics compared to the SM background. 78

7.4 Relative contributions of the individual SM backgrounds in the signal region. 79

7.5 Kinematic distributions of the variables used to define the signal region. The selection on the displayed quantity is omitted as indicated by the vertical line, except for the distribution of the m_{T} variable, because a tight requirement on m_{T} is already applied at event preselection level. The direction of the arrow on top of the vertical line indicates the region of accepted events in the signal region. Only statistical uncertainties are included. The category 'Others' in the legend represents the sum of minor SM backgrounds that contribute less than 5 % of the total SM background. The lower panel shows the ratio of signal to SM background in each bin respectively. 80

7.6 Relative contributions of the individual SM backgrounds in the control region. 82

7.7 Kinematic distributions of key variables for events that pass the control region selection. The $t\bar{t}$ events are scaled by the normalisation factor obtained in the CR. Residual SM background processes are normalised with the respective theoretical cross section. The category 'Others' in the legend represents the sum of minor SM backgrounds that contribute less than 5 % of the total SM background. The lower panel shows the ratio of data to the SM prediction. The uncertainty band includes statistical and experimental systematic uncertainties. $t\bar{t}$ modelling uncertainties are only determined for the VR and SR and hence not included in the total uncertainty. The last bins include the overflow. 83

7.8 Relative contributions of the individual SM backgrounds in the validation region. 84

7.9 Kinematic distributions of key variables in the validation region. The $t\bar{t}$ events are scaled by the normalisation factor obtained in the CR. Residual SM background processes are normalised with the respective theoretical cross section. The category 'Others' in the legend represents the sum of minor SM backgrounds that contribute less than 5 % of the total SM background. The lower panel shows the ratio of data to the SM prediction. The uncertainty band includes statistical and systematic uncertainties. The last bins include the overflow. 85

7.10 Comparison between the observed data (n_{obs}) and the predicted SM background (n_{exp}) events in the validation and the signal region. The shaded area around the SM expectation includes all uncertainties. The bottom panel shows the difference between observed and the predicted events divided by the total uncertainty (σ_{tot}). 88

7.11 Kinematic distribution of relevant variables in the signal region. The requirement on the plotted quantity is omitted, but indicated by the vertical lines, except for the m_{T} distribution. The predicted $t\bar{t}$ background process is scaled with the normalisation factor obtained from the control region. Furthermore, the benchmark signal model is shown. The label 'Others' in the legend represents minor SM backgrounds that contribute less than 5% of the total background. All uncertainties are included and represented by the shaded area around the total SM expectation. The last bin contains overflows. 89

7.12 Kinematic distribution of am_{T2} for the shape-fit scenario. The event selections corresponding to the bins of the shape-fit are applied. $t\bar{t}$ events are scaled with the normalisation factor obtained from the control region. Statistical and systematic uncertainties are included. The last bin contains overflows. Additionally, the benchmark signal model is overlaid for comparison. The bottom panel shows the difference between observed and predicted events divided by the total uncertainty (σ_{tot}). 90

7.13 Expected (black dashed) and observed (red solid) exclusion limit at 95 % confidence level in the phase space spanned by $m(\tilde{\chi}_1^0)$ versus $m(\tilde{t}_1)$ for direct top squark pair production. Either $\tilde{t}_1 \to t\tilde{\chi}_1^0$, $\tilde{t}_1 \to bW\tilde{\chi}_1^0$, or $\tilde{t}_1 \to bff'\tilde{\chi}_1^0$ decay modes with a branching ratio of 100 % are assumed. The excluded regions from previous publications [55,173] are illustrated with the grey and blue shaded area, respectively. 91

7.14 Expected (black dashed) and observed (red solid) 95% excluded regions in the plane of $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$ versus $m(\tilde{t}_1)$ for direct top squark pair production assuming either $\tilde{t}_1 \to t\tilde{\chi}_1^0$, $\tilde{t}_1 \to bW\tilde{\chi}_1^0$, or $\tilde{t}_1 \to bff'\tilde{\chi}_1^0$ decay with a branching ratio of 100 %. The excluded regions from previous publications [55,173] are also illustrated with the grey shaded area. 92

8.1 Relative contributions of the individual SM backgrounds after the preselection. 94

8.2 Kinematic distributions of key observables after the preselection is applied. The $t\bar{t}$ background prediction is scaled by the normalisation factor obtained from a likelihood fit of the CR, while residual SM background processes are normalised with the respective theoretical cross section. The hatched area around the total SM prediction and the hatched band in the ratio include statistical and systematic uncertainties, except for the $t\bar{t}$ modelling uncertainties, which are defined only for the signal and validation region. Additionally, the expected distribution of the SUSY benchmark signal model is overlaid (blue dashed line). Overflows are included in the last bin. 95

8.3 Comparison of kinematic distributions between smeared particle-level signal events (“Smeared-truth”, shown as black dots on a black histogram) and fully reconstructed signal events (“Reco” shown as red dots on a red histogram) after preselection. Distributions are normalised to unity in order to investigate the shape of the distributions. 98

8.4 Conceptual layout of the machine learning architecture used to discriminate signal from SM background. In order to exploit the information of the whole jet collection, jet 4-vectors are processed in a recurrent neural network. The output vector of the jet-RNN is then passed as an input to the NN together with further discriminating variables. 99

8.5 Distributions of the linear correlation between two kinematic observables, respectively. Agreement between data and simulated background events is observed. Only statistical uncertainties are shown. Underflows and overflows are included in the first and last bins. 102

8.6 Distribution of the accuracy and loss of the NN as a function of the number of training epochs. 103

8.7 Distribution of the ML output score, referred to as $\mathrm{NN}_{\mathrm{bWN}}$. The events are separated into signal and background used for training and those used for testing, respectively. Distributions of training and test events are in good agreement, no overtraining is observed. Distributions are normalised to unity. 104

8.8 Receiver operating characteristic of the ML classifier. The events are separated into training and test set. Both curves are in good agreement, which indicates that the NN is not overtrained. The quantity in the legend denotes the area under the curve (AUC) and is a metric for the overall performance of the classification problem. 104

8.9 Confusion matrices of training set and test set. The values in the diagonal and off-diagonal terms represent the fraction of correctly and misidentified events of both classes. 105

8.10 Distribution of the mean of the sensitive input variable as a function of $\mathrm{NN}_{\mathrm{bWN}}$ for data and simulated SM events. Good agreement between observed and simulated events is observed. The error bands represent the standard deviation (RMS) of the respective variable in the corresponding bin. 106

8.11 Comparison of the distribution of the ML output score between particle-level and reconstructed events of the benchmark signal model. Both distributions agree well within uncertainty. In order to investigate differences of the shapes of the distributions, the distributions are normalised to unity. 107

8.12 Illustration of the ROC curves and working points, associated with a signal region, of the primary ML classifier and various alternative methods. 109

8.13 Distribution of the trained ML output score after the preselection. Values close to 1 indicate that the classified event is more signal-like while small values determine events as background-like. The $t\bar{t}$ background is not yet normalised in the CR and only statistical uncertainties are illustrated by the hatched band. 110

8.14 Breakdown of the individual SM contributions to the signal region. 110

8.15 Expected significance across the signal phase space spanned by $m(\tilde{t}_1)$ versus $m(\tilde{\chi}_1^0)$. The considered phase space spans over signal points characterised by 20 GeV $<$ Δm $<$ 200 GeV. The black contour line illustrates the threshold for $Z > 3$. 111

8.16 Distribution of m_T for simulated SM backgrounds as well as for three different signal points in the interval of $\mathrm{NN}_\mathrm{bWN} \in [0.65, 0.8)$ (a) and $\mathrm{NN}_\mathrm{bWN} \geq 0.8$ (b). At moderate values of the output scores (a), a significant amount of semi-leptonic $t\bar{t}$ processes saturates at $m_\mathrm{T} < 150$ GeV, whereas such events are suppressed in the region characterised by large values of NN_bWN (b). . . 111

8.17 Expected significance Z for rejecting the background+signal hypothesis for two different signal points as a function of bins included to the maximum likelihood fit. Z significantly increases if multiple bins are added to the fit procedure. Proper systematic uncertainties accounting for $t\bar{t}$ radiation, fragmentation and hard scattering are applied respectively, as well as systematic uncertainties for flavor tagging, lepton identification, isolation and reconstruction, and pileup are included. $E_\mathrm{T}^\mathrm{miss}$ uncertainties as well as uncertainties on jet energy scales and resolutions are also applied. Signal acceptance uncertainties and uncertainties on the cross section of the minor SM backgrounds are included as well. 112

8.18 Distribution of NN_bWN in the range from 0.65 to 1.0 for SM background as well as for signal points with three different mass splittings. An additional requirement on $m_\mathrm{T} > 150$ GeV is applied on the first three bins ranging from 0.65 to 0.8. Only the statistical uncertainty of the simulated events is shown in the shaded area around the total SM expectation. The $t\bar{t}$ background is not yet normalised to the CR. 112

8.19 Relative contributions of the individual SM backgrounds in the control region. 114

8.20 Signal contamination in the control region illustrated in the phase space spanned by $m(\tilde{t}_1)$ versus $m(\tilde{\chi}_1^0)$. The black contour line illustrates the threshold for a signal fraction of 5%. Signal points where the contamination is above 5% are already excluded by previous analyses. 115

8.21 Kinematic distributions of selected observables in the control region. The $t\bar{t}$ background prediction is scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross section. The hatched band around the total SM prediction and the hatched band in the data/SM ratio include statistical and systematic uncertainties, except for the $t\bar{t}$ modelling uncertainties, which are only defined for the signal and validation region. Overflows are included in the last bin. 116

8.22 Relative contributions of the individual SM backgrounds in the validation region. 117

8.23 Signal contamination in the validation region illustrated in the phase space spanned by $m(\tilde{t}_1)$ versus $m(\tilde{\chi}_1^0)$. The black contour line illustrates the threshold for a signal fraction of 10%. Signal points where the contamination is above 10% are already excluded by previous analyses. 118

8.24 Kinematic distributions of selected observables in the validation region. The $t\bar{t}$ background processes are scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross section. The shaded band around the total SM prediction as well as the shaded band in the data/SM ratio include statistical and systematic uncertainties. Overflows are included in the last bin. 119

8.25 Comparison of observed events with the SM prediction in the VR and SR. The shaded area around the total SM background prediction includes all uncertainties. The bottom panel shows the significance of the observed data events given the number of predicted SM background events. 123

8.26 Kinematic distributions of selected observables in the signal region. The $t\bar{t}$ background processes are scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross section. The shaded band around the total SM prediction as well as the shaded band in the data/SM ratio include statistical and systematic uncertainties. Overflows are included in the last bin. 124

8.27 Distribution of $\mathrm{NN_{bWN}}$. The variable bin widths correspond to the definitions of the control, validation and signal regions. The shaded area around the total SM prediction includes all uncertainties. The expected distribution from a representative signal benchmark model is overlaid. The bottom panel shows the significance of the observed data events given the number of predicted SM background events. 126

8.28 Expected (black dashed) and observed (red solid) excluded regions in the plane of $m(\tilde{\chi}_1^0)$ versus $m(\tilde{t}_1)$ at 95 % confidence level for direct top squark pair production. Either $\tilde{t}_1 \to t\tilde{\chi}_1^0$, $\tilde{t}_1 \to bW\tilde{\chi}_1^0$, or $\tilde{t}_1 \to bff'\tilde{\chi}_1^0$ decay modes are assumed with a branching ratio of 100 %. The grey shaded area illustrates the excluded region from a previous publication [58], including also the results obtained from the analysis described in Chapter 7. 127

8.29 Expected (black dashed) and observed (red solid) excluded regions in the plane of $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$ versus $m(\tilde{t}_1)$ at 95 % confidence level for direct top squark pair production. Either $\tilde{t}_1 \to t\tilde{\chi}_1^0$, $\tilde{t}_1 \to bW\tilde{\chi}_1^0$, or $\tilde{t}_1 \to bff'\tilde{\chi}_1^0$ decay modes are assumed with a branching ratio of 100 %. The grey shaded area illustrates excluded regions from a previous publication [58], including also the results obtained from the analysis described in Chapter 7. 128

9.1 Comparison of the expected (green dashed) excluded region obtained from a shape-fit in am_{T2} with the expected (red dashed) and observed (red solid) excluded regions derived from a shape-fit in $\mathrm{NN_{bWN}}$ in the plane of $m(\tilde{t}_1)$ versus $m(\tilde{\chi}_1^0)$ for direct top squark pair production. The limits correspond to 95 % confidence level. Either $\tilde{t}_1 \to t\tilde{\chi}_1^0$, $\tilde{t}_1 \to bW\tilde{\chi}_1^0$, or $\tilde{t}_1 \to bff'\tilde{\chi}_1^0$ decay modes with a branching ratio of 100 % are assumed. The excluded region from the previous publication [55] is illustrated as the grey shaded region. 130

9.2 Comparison of the expected (green dashed) excluded region obtained from a shape-fit in am_{T2} with the expected (red dashed) and observed (red solid) excluded regions derived from a shape-fit in $\mathrm{NN}_{\mathrm{bWN}}$ in the plane of $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$ versus $m(\tilde{t}_1)$ for direct top squark pair production. The limits correspond to 95 % confidence level. Either $\tilde{t}_1 \rightarrow t\tilde{\chi}_1^0$, $\tilde{t}_1 \rightarrow bW\tilde{\chi}_1^0$, or $\tilde{t}_1 \rightarrow bff'\tilde{\chi}_1^0$ decay modes with a branching ratio of 100 % are assumed. The excluded region from the previous publication [55] is illustrated as the grey shaded region. . 131

9.3 Expected (red dashed) and observed (black solid) 95% exclusion limit for top squark pair production [203] in the plane spanned by $m(\tilde{t}_1)$ and $m(\tilde{\chi}_1^0)$. Thin red dashed lines indicate the $\pm 1, 2\,\sigma_{\mathrm{experiment}}$ uncertainty in the expected limit. The thin black lines represent the $\pm 1\,\sigma_{\mathrm{theory}}$ uncertainty based on the theoretical uncertainties on the signal cross section. 132

A.1 Kinematic distributions of key observables after the preselection is applied. The $t\bar{t}$ background prediction is scaled by the normalisation factor obtained from a likelihood fit of the CR, while residual SM background processes are normalised with the respective theoretical cross-section. The hatched area around the total SM prediction and the hatched band in the ratio include statistical and systematic uncertainties, except for the $t\bar{t}$ modelling uncertainties, which are defined only for the signal and validation region. Additionally, the expected distribution of the SUSY benchmark signal model is overlaid (blue dashed line). Overflows are included in the last bin. 167

A.2 Comparison of kinematic distributions between smeared particle-level signal events (“Smeared-truth”, shown as black dots on a black histogram) and fully reconstructed signal events (“Reco” shown as red dots on a red histogram) after preselection. Distributions are normalised to unity in order to investigate the shape of the distributions. 169

A.2 Distributions of the linear correlation between two kinematic observables, respectively. Agreement between data and simulated background events is observed. Only statistical uncertainties are shown. Underflows and overflows are included in the first and last bins. 179

A.2 Distribution of the mean of the sensitive input variable as a function of $\mathrm{NN}_{\mathrm{bWN}}$ for data and simulated SM events. Good agreement between observed and simulated events is observed. The error bands represent the standard deviation (RMS) of the respective variable in the corresponding bin. 182

A.3 Kinematic distributions of selected observables in the control region. The $t\bar{t}$ background prediction is scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross section. The hatched band around the total SM prediction and the hatched band in the data/SM ratio include statistical and systematic uncertainties, except for the $t\bar{t}$ modelling uncertainties, which are only defined for the signal and validation region. Overflows are included in the last bin. 186

A.4 Kinematic distributions of selected observables in the validation region. The $t\bar{t}$ background processes are scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross section. The shaded band around the total SM prediction as well as the shaded band in the data/SM ratio include statistical and systematic uncertainties. Overflows are included in the last bin. 189
A.5 Kinematic distributions of selected observables in the signal region. The $t\bar{t}$ background processes are scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross section. The shaded band around the total SM prediction as well as the shaded band in the data/SM ratio include statistical and systematic uncertainties. Overflows are included in the last bin. 192

List of Tables

2.1 The three fundamental forces contained in the SM. Their associated mediators, together with their spin and mass are shown [5]. 4

2.2 Basic properties of the twelve fundamental fermions separated into leptons and quarks. Masses are rounded, precises values can be found in [5]. 4

2.3 Chiral supermultiplets in the MSSM. 14

2.4 Gauge supermultiplets in the MSSM. 15

4.1 Summary of the simulated samples used in analysis `bWN-subSet`. 43

4.2 Summary of the simulated samples in analysis `bWN-fullSet`. 43

4.3 Electron selection requirements for analysis `bWN-subSet` and `bWN-fullSet`, respectively. 46

4.4 Muon selection requirements for analysis `bWN-subSet` and `bWN-fullSet`, respectively. 48

4.5 Jet selection requirements, including requirements for b-tagged jets. 50

7.1 Summary of the event preselection criteria. 76

7.2 Overview of the signal region selection. In addition, the event preselection as defined in Table 7.1 is applied. The last row shows the bin intervals for the shape-fit scenario. 79

7.3 Number of expected events in the signal region for 36.1 fb^{-1}. Only statistical uncertainties are shown. 79

7.4 Comparison of the expected significance for various signal points in the single-bin and shape-fit scenario. Statistical and systematics uncertainties are included in the computation. 81

7.5 Overview of the event selection of the signal region, the control region to estimate the normalisation of the $t\bar{t}$ background, and the validation region to verify the estimation method. The event preselection as defined in Table 7.1 is applied to all regions. 82

7.6 Observed and predicted SM events in the control region. The individual uncertainties are correlated and do not add up quadratically to the total uncertainty. The lower section lists the number of SM events obtained from simulation. 82

7.7 Observed and predicted SM events in the validation region. The individual uncertainties are correlated and do not add up quadratically to the total uncertainty. Only statistical uncertainties are included in the simulated SM backgrounds listed in the lower part of the table. 84

7.8 Summary of the dominant systematic uncertainties on the total predicted number of SM background events for the validation and signal region. Individual uncertainties may be correlated, and hence do not necessarily add up in quadrature to the total uncertainty. Systematic uncertainties which contribute less than 1 % are omitted in the list. The percentages show the size of the uncertainty relative to the total predicted SM background. The systematic uncertainties are ordered by the size of their relative contribution to the total predicted SM background in the SR. 86

7.9 Summary of the dominant systematic uncertainties on the total predicted number of SM background events for the five bins in the shape-fit. Individual uncertainties may be correlated, and hence do not necessarily add up in quadrature to the total uncertainty. Systematic uncertainties which contribute less than 1 % are omitted in the list. The percentages show the size of the uncertainty relative to the total predicted SM background. The systematic uncertainties are ordered by their relative contribution to the total predicted SM background in the most sensitive bin (Bin 1) with respect to the SUSY signal benchmark. 87

7.10 Number of observed events in the signal region together with the predicted number of SM background events. All types of uncertainties are included. In addition, the normalisation factor ($\mu_{t\bar{t}}$) for the prediction of $t\bar{t}$ processes as obtained in the fit is listed, as well the expected ($S^{95}_{\text{exp.}}$) and observed ($S^{95}_{\text{obs.}}$) 95 % upper limit on the number of beyond-SM events. 88

7.11 Number of observed events and predicted SM background events for 36.1 fb^{-1}, in each signal region contributing to the shape-fit. Individual uncertainties can be correlated and do not necessarily add up quadratically to the total background uncertainty. 89

8.1 Summary of the event preselection criteria. 94

8.2 Discriminating variables applied as inputs to the NN. In addition, the output vector of the jet-RNN to process the arbitrary-length jet collection is also applied as an input. 100

8.3 Architecture and parameters of the neural network. 100

8.4 Comparison of the discovery significance of NN_{bWN} with alternative cut-based and multi-variate approaches. 108

8.5 Overview of the signal region definition. A single-bin SR is defined for a potential data excess (Discovery), while a shape-fit is used to enhance the exclusion sensitivity (Exclusion). Square brackets indicate the interval in the NN_{bWN} score. An additional requirement of $m_{\text{T}} > 150$ GeV is applied to bins marked by *, in order to suppress potential contamination of semi-leptonic $t\bar{t}$ events. 109

8.6 Expected simulated events in the signal region for 139 fb^{-1}. Only the statistical uncertainties are given. 110

8.7 Expected events in the corresponding bin of the shape-fit configuration for 139 fb^{-1}. Only the statistical uncertainty of the simulated events is given. The $t\bar{t}$ background is not yet normalised to the CR. 113

8.8 Summary of the event selection for the signal region and the associated control and validation region. The event preselection as defined in Table 8.1 is applied to all regions, but not explicitly shown. 114

8.9 Observed and predicted SM events in the control region. The individual uncertainties in the upper table are correlated and do not add up quadratically to the total uncertainty. The lower section lists the number of SM events obtained from simulation. 114

8.10 Observed and predicted SM events in the validation region. The individual uncertainties in the upper list are correlated and do not add up quadratically to the total uncertainty. The lower section lists the number of SM events obtained from simulation. 117

8.11 Summary of the dominant experimental and theoretical systematic uncertainties on the total predicted SM background in the validation region and the signal region, respectively. Individual uncertainties may be correlated, and hence do not necessarily add up in quadrature to the total uncertainty. Systematic uncertainties which contribute less than 1 % to the total SM prediction are omitted in the list. The percentages show the size of the uncertainty relative to the total predicted SM background. The systematic effects are ordered by the size of their relative contribution to the total predicted SM background in the SR. 120

8.12 Summary of the dominant systematic uncertainties on the total predicted number of SM background events for the bins included in the shape-fit. Individual uncertainties may be correlated, and hence do not necessarily add up in quadrature to the total uncertainty. In order to avoid complexity, the various components for jet energy scale and resolution, $E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution, as well as systematic effects affecting the flavour tagging efficiency are summarised accordingly. Systematic uncertainties which contribute less than 1 % to the total SM prediction are omitted in the list. The percentages show the size of the uncertainty relative to the total predicted SM background. The sources of systematic uncertainties are ordered by their relative contribution to the total predicted SM background in the most sensitive bin (Bin 10) with respect to the SUSY signal benchmark. 122

8.13 Number of observed events in the signal region together with the number of predicted SM background events and their total uncertainties. In addition, the normalisation factor ($\mu_{t\bar{t}}$) for the prediction of the $t\bar{t}$ background is listed, and the expected (S^{95}_{exp}) and observed (S^{95}_{obs}) 95 % confidence level upper limits on the number of beyond-SM events. 123

8.14 Observed data events in each bin of the shape-fit configuration together with the predicted number of total SM background events and their uncertainties. The bin i (i = 1–10) corresponds to the i-th bin (from left to right) of the ML classifier output score used in the shape-fit. Contributions of the individual SM backgrounds are also given. Individual uncertainties may be correlated, and hence do not necessarily add up in quadrature to the total uncertainty. 125

A.1 Overview of an alternative signal region selection. In addition, the event preselection as defined in Table 8.1 is applied. 182

A.2 Overview of an alternative signal region selection. In addition, the event preselection as defined in Table 8.1 is applied. Optimisation of the requirements of the other input variables used to train $\mathrm{NN}_{\mathrm{bWN}}$ had no effect on the discovery significance. 183

A.3 Discriminating variables applied as inputs to train an alternative neural network. 183

A.4 Architecture and parameters of the neural network. 184

List of Abbreviations

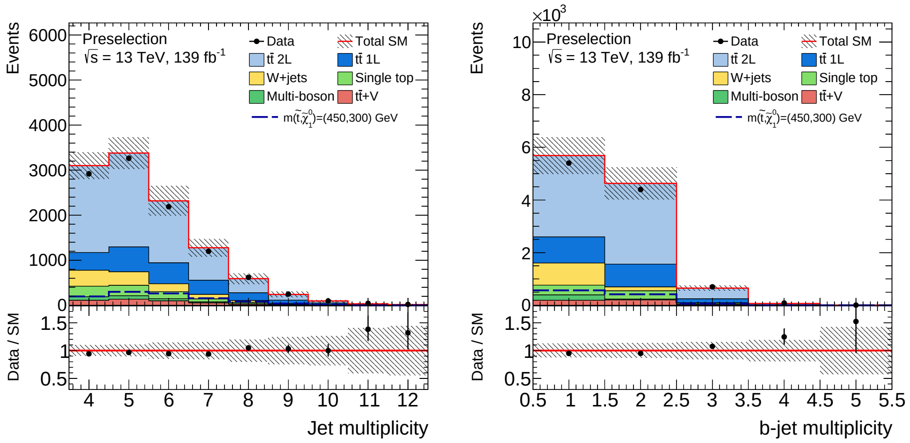
AUC	Area under ROC curve	68
BDT	Boosted Decision Tree	49
BSM	Theory beyond the standard model	11
CMB	Cosmological microwave background	12
CR	Control region	37
CSC	Cathode strip chamber	33
CTP	Central trigger processor	34
EMEC	Electromagnetic end-cap calorimeter	31
EWK	Electroweak	7
FCal	Forward calorimeter	32
GUT	Grand unified theory	12
HEC	Hadronic end-cap calorimeter	32
HLT	High-level trigger	34
IBL	Insertable B-Layer	29
ID	Inner detector	26
IP	Interaction point	24
JVT	Jet vertex tagger	49
L1	Level-1 trigger	34
ΛCDM	Lambda cold dark matter	12
LEP	Large Electron Positron Collider	23
LHC	Large Hadron Collider	23
LO	Leading order	39
LSP	Lightest supersymmetric particle	15
LSTM	Long short-term memory	62
MDT	Muon drift tube	33
ML	Machine learning	xviii
MS	Muon spectrometer	27
MSSM	Minimal supersymmetric standard model	13
NLO	Next-to-leading order	39
NN	Neural network	57
NNLL	Next-to-next-to-leading logarithm	42
NNLO	Next-to-next-to-leading order	39
PDF	Parton distribution function	39
QCD	Quantum chromodynamics	3
QED	Quantum electrodynamics	3
ReLU	Rectified linear unit	59
RNN	Recurrent neural network	57
ROC	Receiver operating characteristic	68
RPC	Resistive plate chamber	33
SCT	Silicon microstrip tracker	29

SGD	Stochastic gradient descent	64
SM	Standard model of particle physics	3
SPS	Super Proton Synchrotron	24
SR	Signal region	37
SUSY	Supersymmetry	13
TGC	Thin gap chamber	33
TRT	Transition radiation tracker	29
VR	Validation region	37

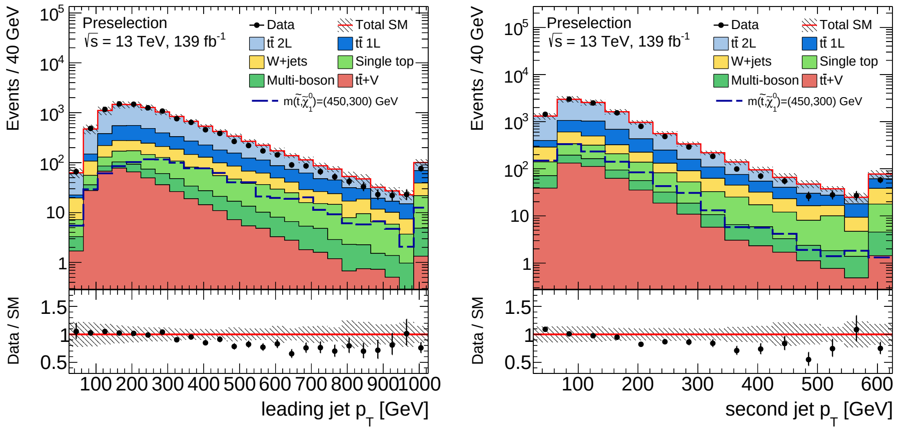
Appendix A

Supplement to `bWN-fullSet`

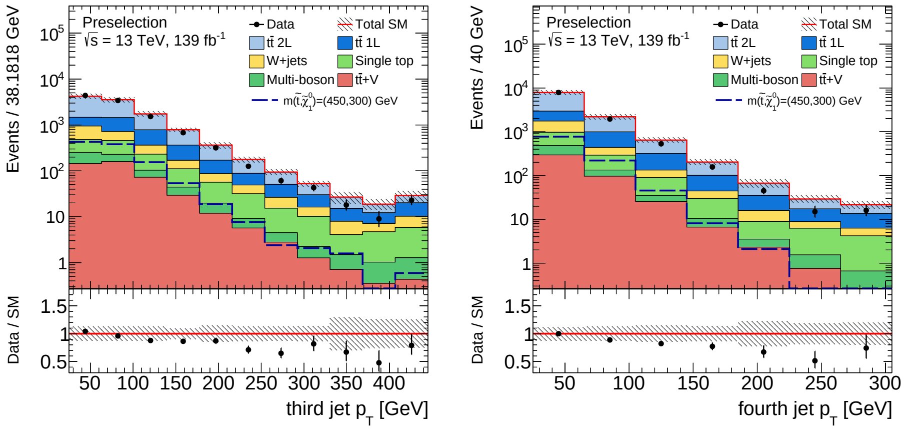
A.1 Additional kinematic distributions after the preselection



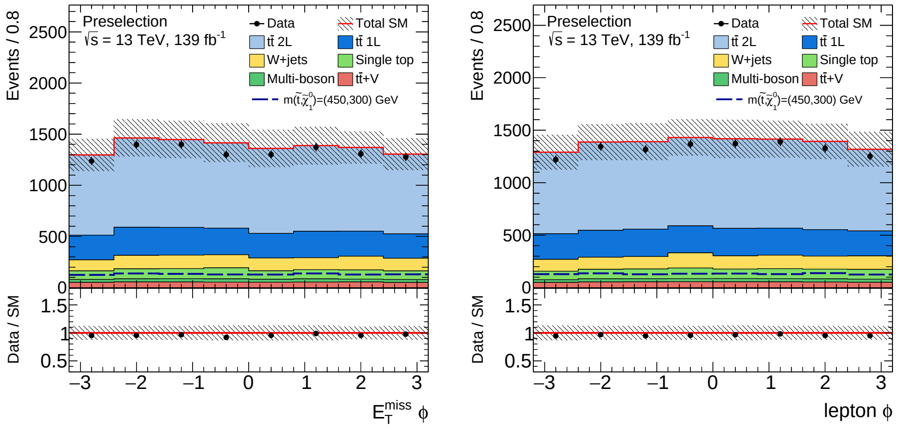
(a) Jet multiplicity (b) *b*-jet multiplicity



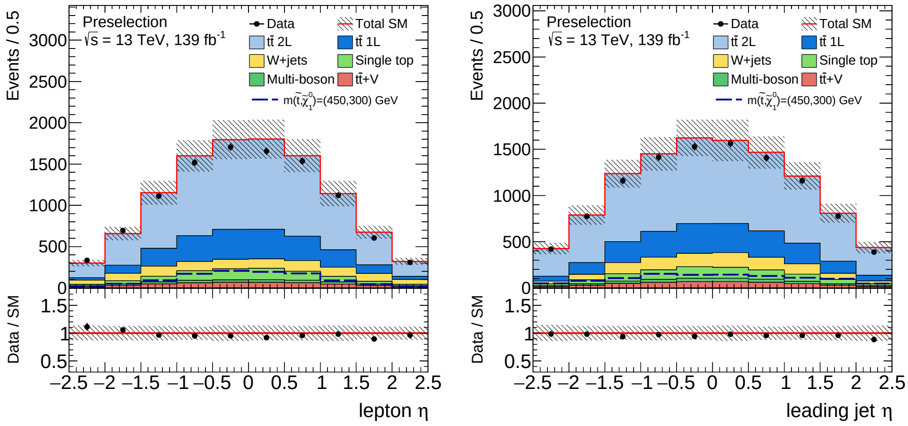
(c) Leading jet p_{T} (d) 2nd leading jet p_{T}



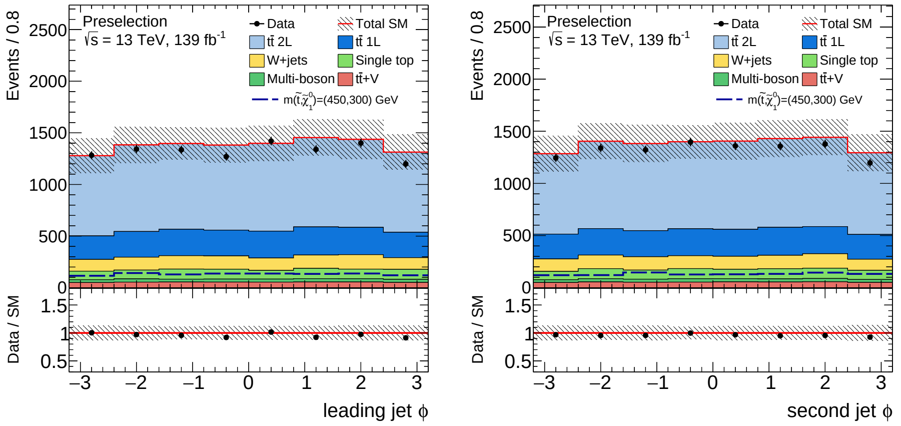
(e) 3rd leading jet p_{T} (f) 4th leading jet p_{T}



(g) $\phi(E_{\mathrm{T}}^{\mathrm{miss}})$ (h) $\phi(\ell)$



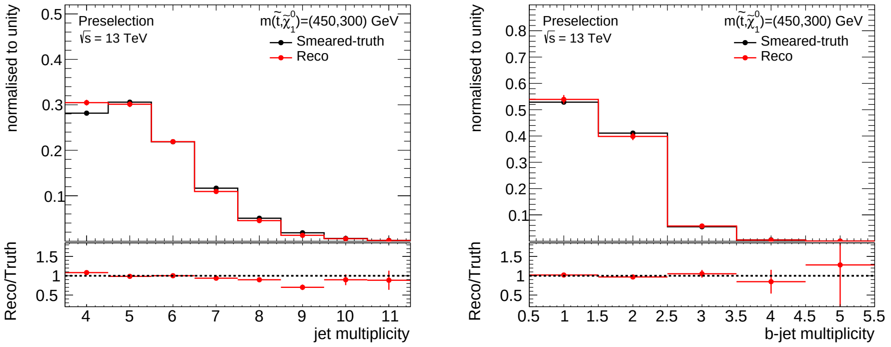
(i) $\eta(\ell)$ (j) Leading jet η



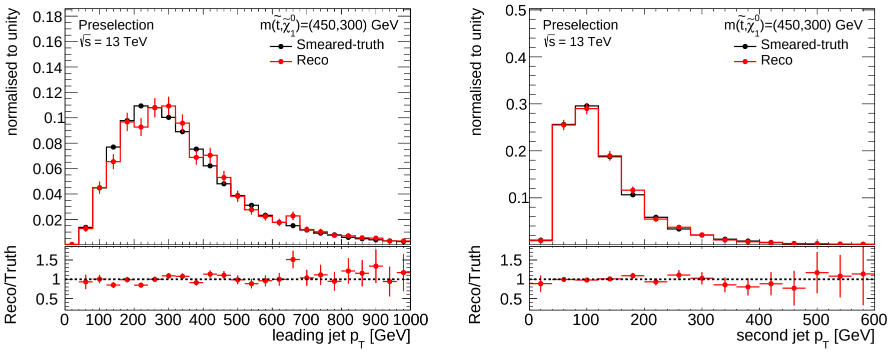
(k) Leading jet ϕ (l) 2^{nd} leading jet ϕ

Figure A.1: Kinematic distributions of key observables after the preselection is applied. The $t\bar{t}$ background prediction is scaled by the normalisation factor obtained from a likelihood fit of the CR, while residual SM background processes are normalised with the respective theoretical cross-section. The hatched area around the total SM prediction and the hatched band in the ratio include statistical and systematic uncertainties, except for the $t\bar{t}$ modelling uncertainties, which are defined only for the signal and validation region. Additionally, the expected distribution of the SUSY benchmark signal model is overlaid (blue dashed line). Overflows are included in the last bin.

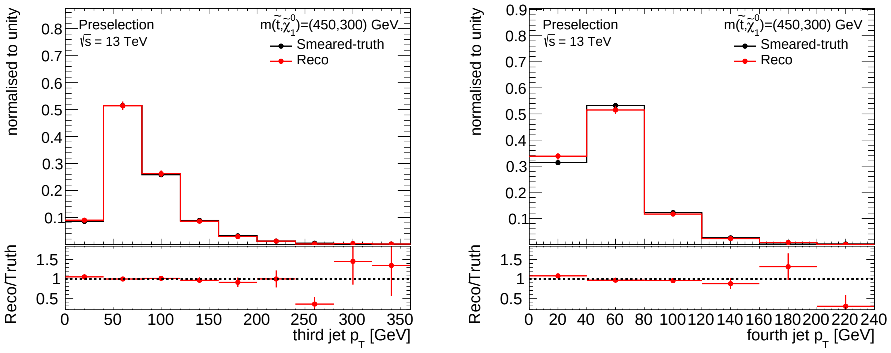
A.2 Comparison of particle-level and detector-level events



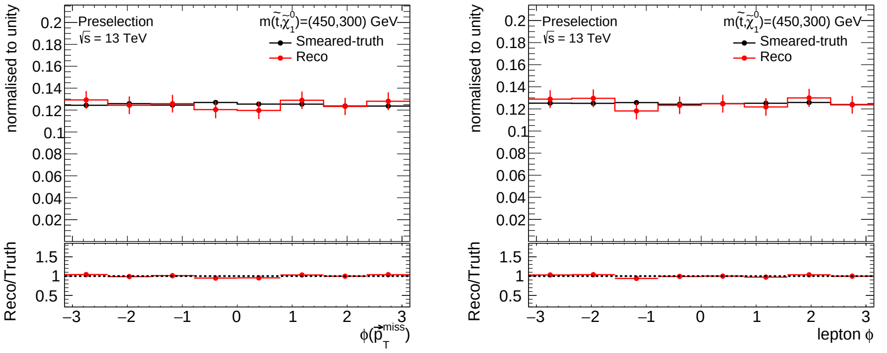
(a) Jet multiplicity (b) b-jet multiplicity



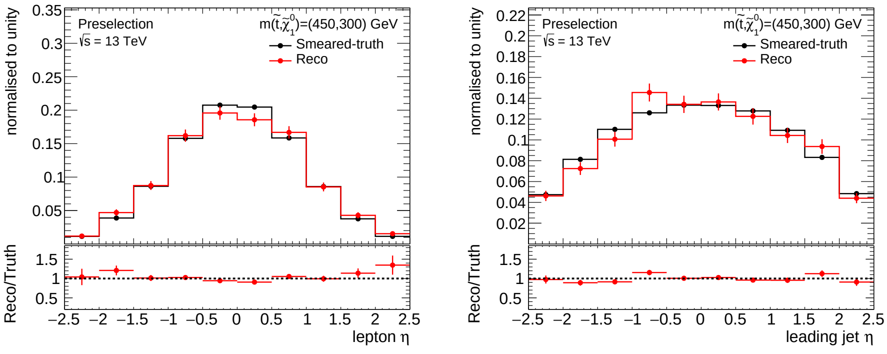
(c) Leading jet p_{T} (d) 2nd leading jet p_{T}



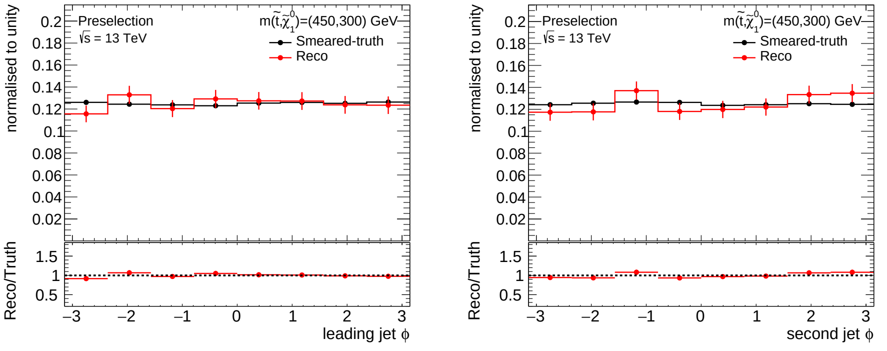
(e) 3rd leading jet p_{T} (f) 4th leading jet p_{T}



(g) $\phi(E_{\mathrm{T}}^{\mathrm{miss}})$ (h) $\phi(\ell)$



(i) $\eta(\ell)$ (j) Leading jet η

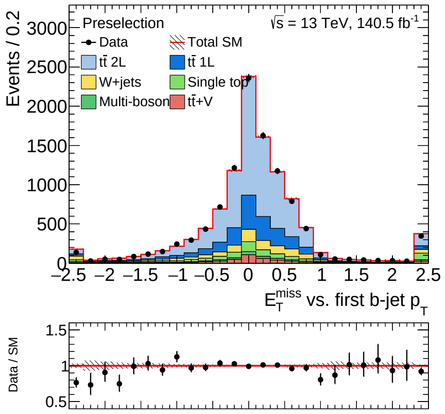


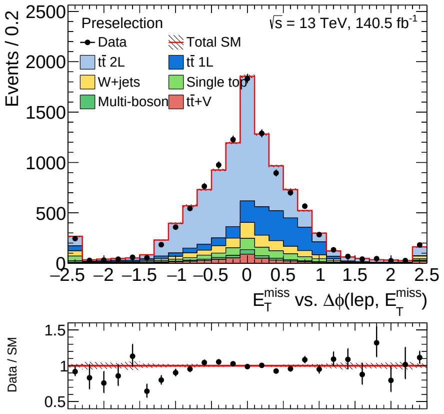
(k) Leading jet ϕ (l) 2^{nd} leading jet ϕ

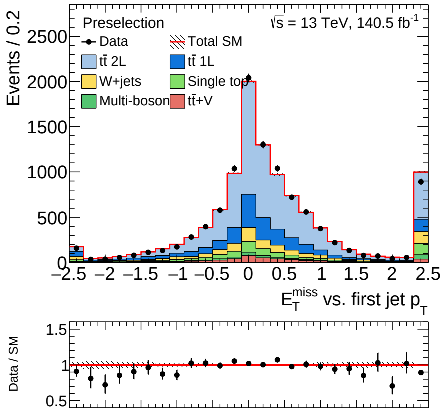
Figure A.2: Comparison of kinematic distributions between smeared particle-level signal events (“Smeared-truth”, shown as black dots on a black histogram) and fully reconstructed signal events (“Reco” shown as red dots on a red histogram) after preselection. Distributions are normalised to unity in order to investigate the shape of the distributions.

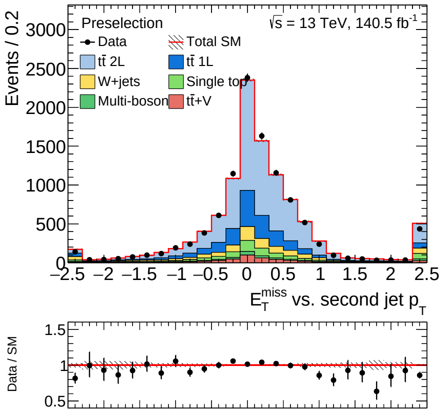
A.3 Linear correlation of input variables

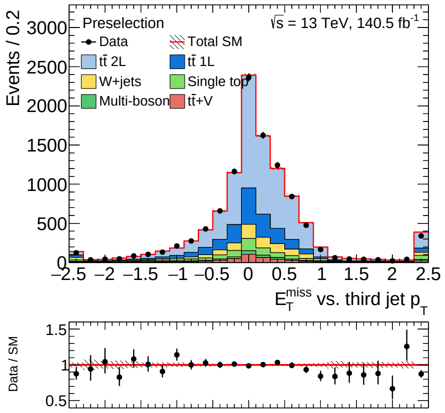
In this Section, the linear correlation defined as $(x-\langle x\rangle)\cdot(y-\langle y\rangle)/(\mathrm{RMS}_x\cdot\mathrm{RMS}_y)$ is shown for the all input variables of the NN except for the η and ϕ distributions of the jets and the leading lepton. Despite fluctuations in statistically limited bins, data and simulated background events agree within uncertainty.

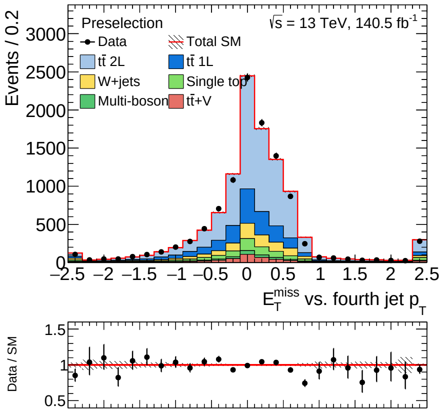

Preselection
$\sqrt{s}$ = 13 TeV, 140.5 fb^{-1}
Data
Total SM
$t\bar{t}$ 2L
$t\bar{t}$ 1L
W+jets
Single top
Multi-boson
$t\bar{t}$+V
Events / 0.2
0
500
1000
1500
2000
2500
3000
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
E_T^{miss} vs. first b-jet p_T
Data / SM
0.5
1
1.5

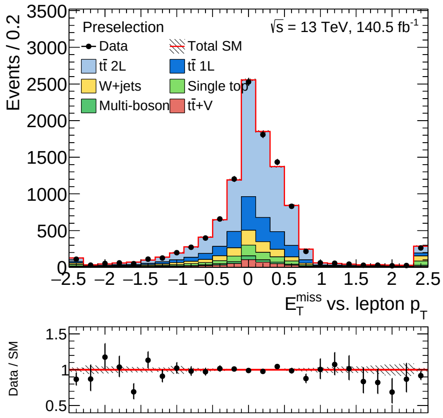

Preselection
$\sqrt{s}$ = 13 TeV, 140.5 fb^{-1}
Data
Total SM
$t\bar{t}$ 2L
$t\bar{t}$ 1L
W+jets
Single top
Multi-boson
$t\bar{t}$+V
Events / 0.2
0
500
1000
1500
2000
2500
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
E_T^{miss} vs. $\Delta\phi$(lep, E_T^{miss})
Data / SM
0.5
1
1.5

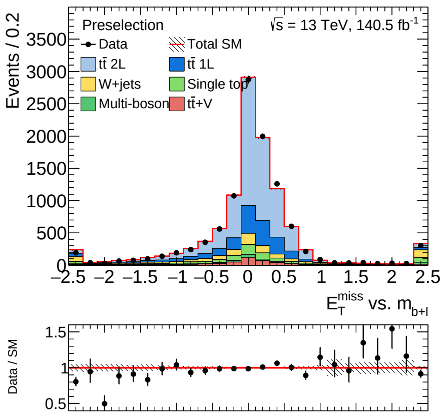

Preselection
$\sqrt{s}$ = 13 TeV, 140.5 fb^{-1}
Data
Total SM
$t\bar{t}$ 2L
$t\bar{t}$ 1L
W+jets
Single top
Multi-boson
$t\bar{t}$+V
Events / 0.2
0
500
1000
1500
2000
2500
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
E_T^{miss} vs. first jet p_T
Data / SM
0.5
1
1.5

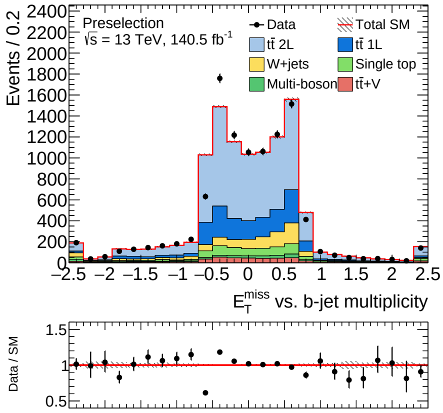

Preselection
$\sqrt{s}$ = 13 TeV, 140.5 fb^{-1}
Data
Total SM
$t\bar{t}$ 2L
$t\bar{t}$ 1L
W+jets
Single top
Multi-boson
$t\bar{t}$+V
Events / 0.2
0
500
1000
1500
2000
2500
3000
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
E_T^{miss} vs. second jet p_T
Data / SM
0.5
1
1.5

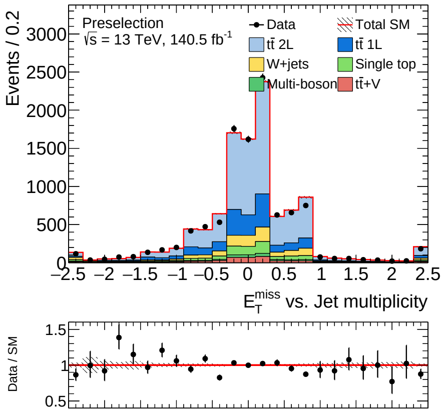

Preselection
$\sqrt{s}$ = 13 TeV, 140.5 fb^{-1}
Data
Total SM
$t\bar{t}$ 2L
$t\bar{t}$ 1L
W+jets
Single top
Multi-boson
$t\bar{t}$+V
Events / 0.2
0
500
1000
1500
2000
2500
3000
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
E_T^{miss} vs. third jet p_T
Data / SM
0.5
1
1.5

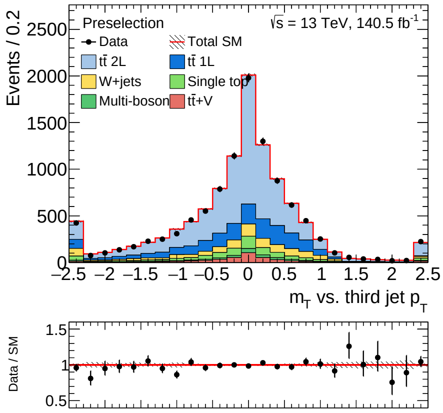

Preselection
$\sqrt{s}$ = 13 TeV, 140.5 fb^{-1}
Data
Total SM
$t\bar{t}$ 2L
$t\bar{t}$ 1L
W+jets
Single top
Multi-boson
$t\bar{t}$+V
Events / 0.2
0
500
1000
1500
2000
2500
3000
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
E_T^{miss} vs. fourth jet p_T
Data / SM
0.5
1
1.5

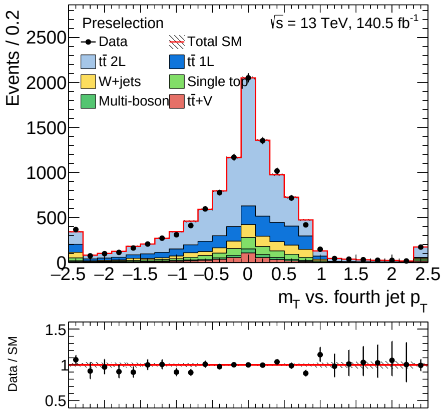

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
E_T^{miss} vs. lepton p_T
Data / SM

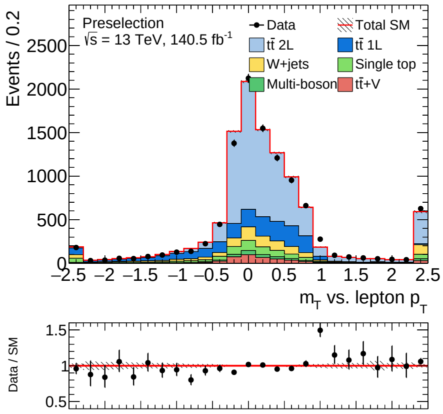

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
E_T^{miss} vs. m_{b+l}
Data / SM

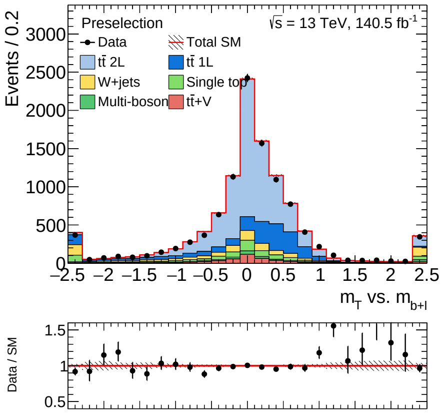

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
E_T^{miss} vs. b-jet multiplicity
Data / SM

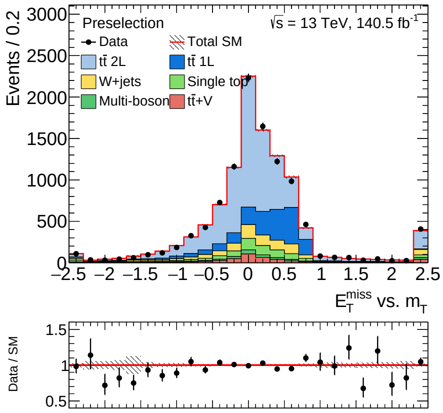

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
E_T^{miss} vs. Jet multiplicity
Data / SM

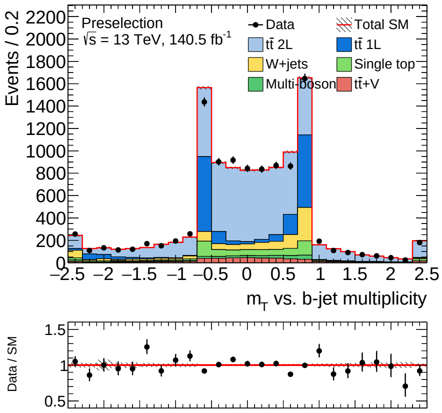

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
m_T vs. third jet p_T
Data / SM

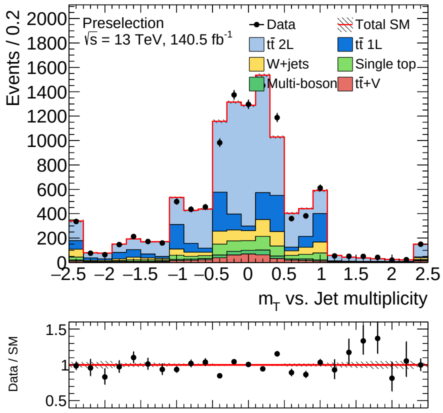

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
m_T vs. fourth jet p_T
Data / SM

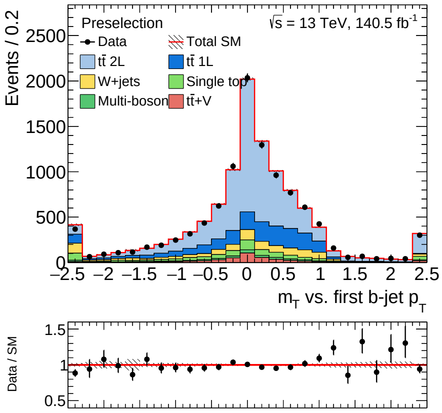

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
m_T vs. lepton p_T
Data / SM

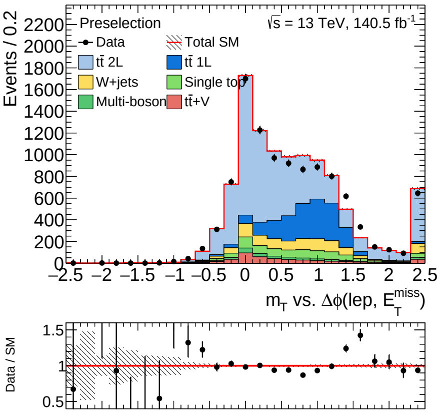

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
m_T vs. m_b+l
Data / SM

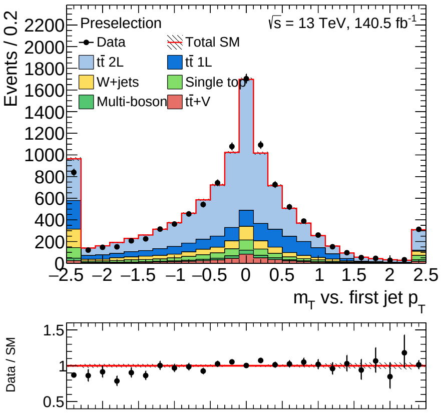

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
E_T^miss vs. m_T
Data / SM

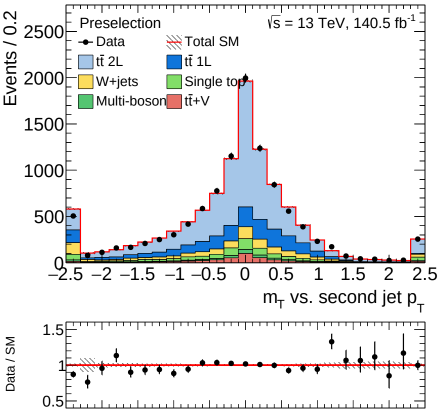

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
m_T vs. b-jet multiplicity
Data / SM

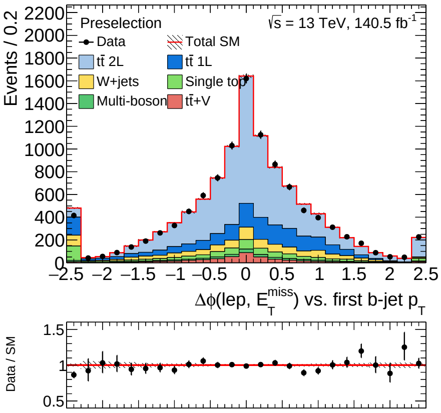

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
m_T vs. Jet multiplicity
Data / SM

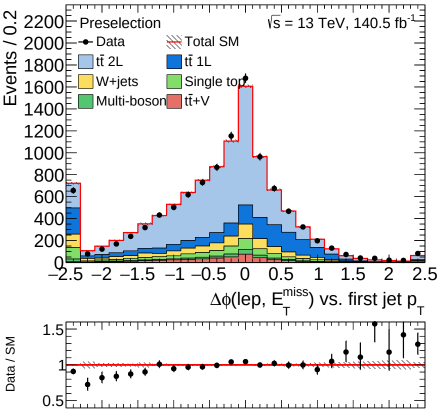

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
m_T vs. first b-jet p_T
Data / SM

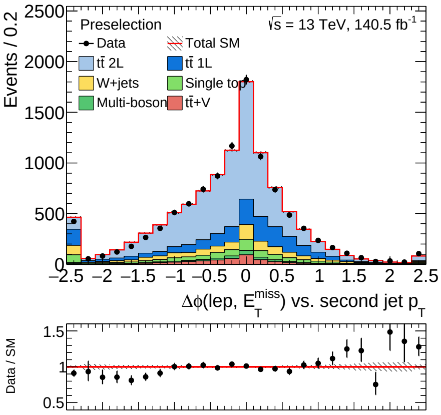

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
m_T vs. Δφ(lep, E_T^miss)
Data / SM
0.5
1
1.5

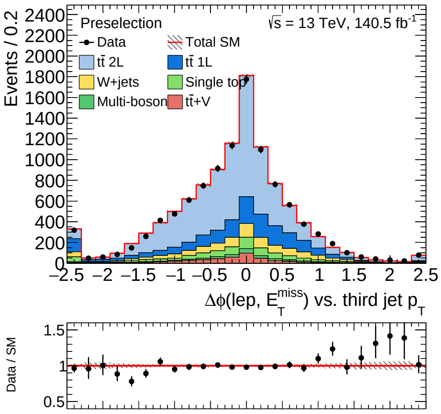

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
m_T vs. first jet p_T
Data / SM
0.5
1
1.5

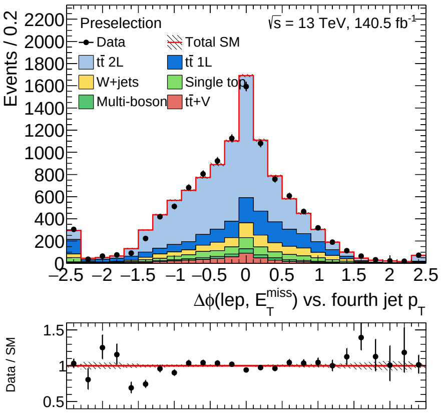

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
0
500
1000
1500
2000
2500
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
m_T vs. second jet p_T
Data / SM
0.5
1
1.5

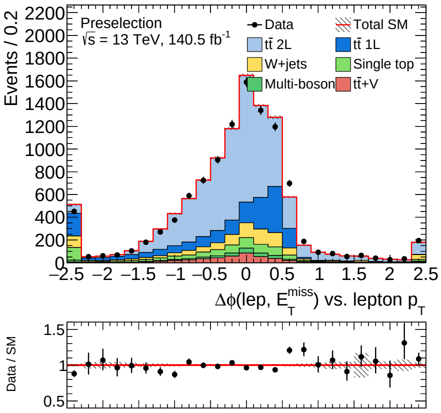

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
Δφ(lep, E_T^miss) vs. first b-jet p_T
Data / SM
0.5
1
1.5

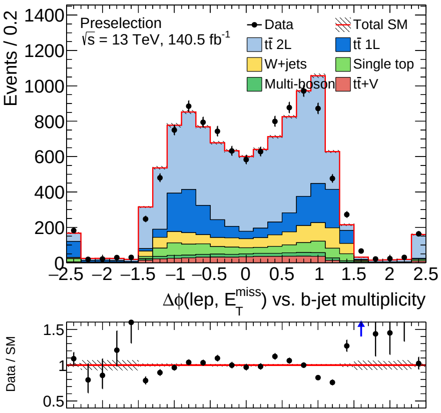

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
Δφ(lep, E_T^miss) vs. first jet p_T
Data / SM
0.5
1
1.5

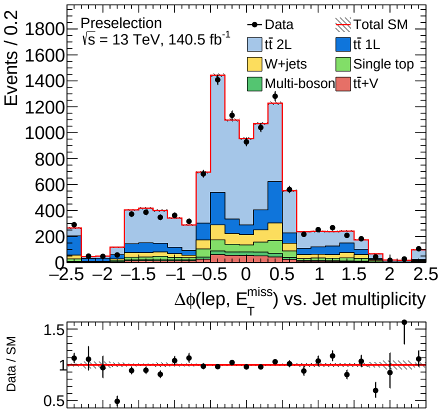

Preselection
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
0
500
1000
1500
2000
2500
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
Δφ(lep, E_T^miss) vs. second jet p_T
Data / SM
0.5
1
1.5

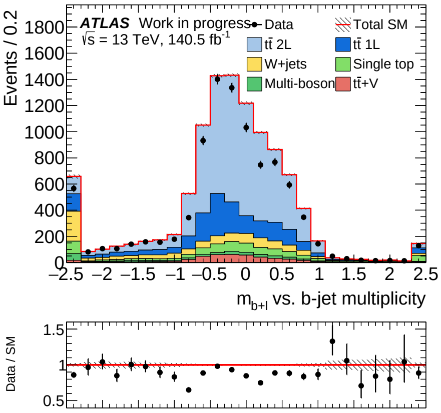

Preselection
$\sqrt{s}$ = 13 TeV, 140.5 fb^{-1}
Data
Total SM
$t\bar{t}$ 2L
$t\bar{t}$ 1L
W+jets
Single top
Multi-boson
$t\bar{t}$+V
Events / 0.2
$\Delta\phi(\mathrm{lep}, E_T^{\mathrm{miss}})$ vs. third jet p_T
Data / SM

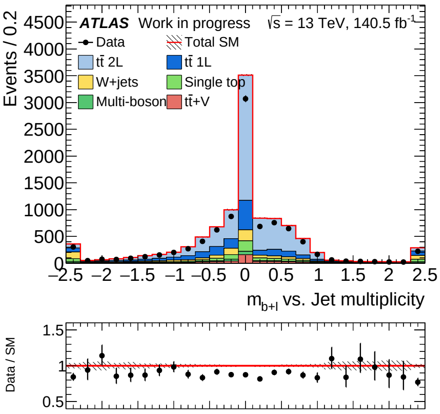

Preselection
$\sqrt{s}$ = 13 TeV, 140.5 fb^{-1}
Data
Total SM
$t\bar{t}$ 2L
$t\bar{t}$ 1L
W+jets
Single top
Multi-boson
$t\bar{t}$+V
Events / 0.2
$\Delta\phi(\mathrm{lep}, E_T^{\mathrm{miss}})$ vs. fourth jet p_T
Data / SM

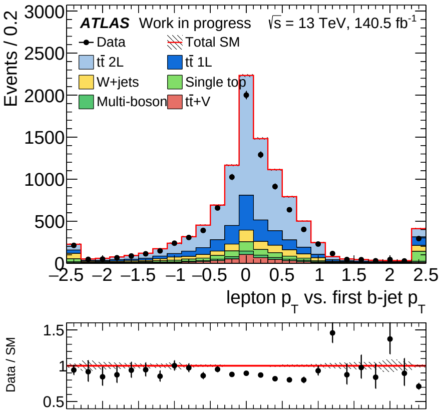

Preselection
$\sqrt{s}$ = 13 TeV, 140.5 fb^{-1}
Data
Total SM
$t\bar{t}$ 2L
$t\bar{t}$ 1L
W+jets
Single top
Multi-boson
$t\bar{t}$+V
Events / 0.2
$\Delta\phi(\mathrm{lep}, E_T^{\mathrm{miss}})$ vs. lepton p_T
Data / SM

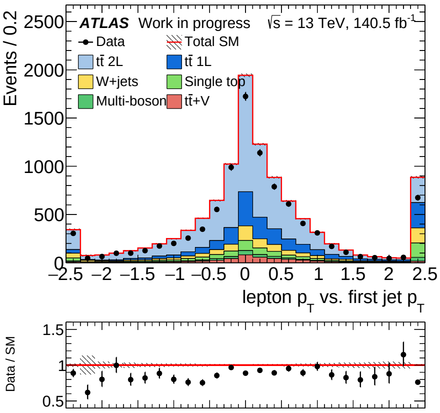

Preselection
$\sqrt{s}$ = 13 TeV, 140.5 fb^{-1}
Data
Total SM
$t\bar{t}$ 2L
$t\bar{t}$ 1L
W+jets
Single top
Multi-boson
$t\bar{t}$+V
Events / 0.2
$\Delta\phi(\mathrm{lep}, E_T^{\mathrm{miss}})$ vs. b-jet multiplicity
Data / SM

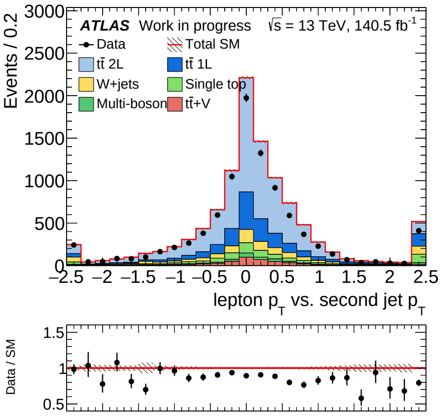

Preselection
$\sqrt{s}$ = 13 TeV, 140.5 fb^{-1}
Data
Total SM
$t\bar{t}$ 2L
$t\bar{t}$ 1L
W+jets
Single top
Multi-boson
$t\bar{t}$+V
Events / 0.2
$\Delta\phi(\mathrm{lep}, E_T^{\mathrm{miss}})$ vs. Jet multiplicity
Data / SM

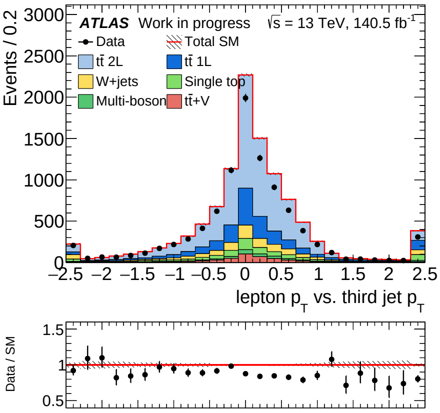

ATLAS Work in progress
$\sqrt{s}$ = 13 TeV, 140.5 fb^{-1}
Data
Total SM
$t\bar{t}$ 2L
$t\bar{t}$ 1L
W+jets
Single top
Multi-boson
$t\bar{t}$+V
Events / 0.2
m_{b+l} vs. b-jet multiplicity
Data / SM

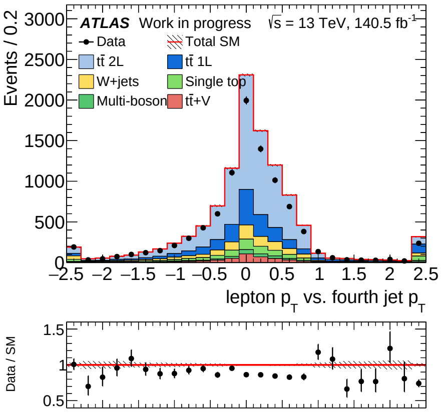

ATLAS Work in progress
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
m_{b+l} vs. Jet multiplicity
Data / SM

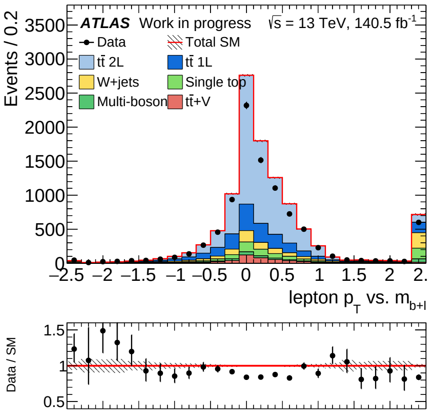

ATLAS Work in progress
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
lepton p_T vs. first b-jet p_T
Data / SM

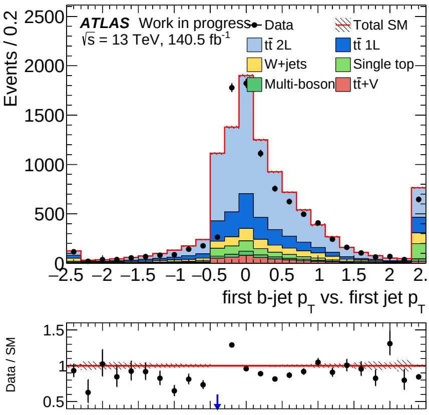

ATLAS Work in progress
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
lepton p_T vs. first jet p_T
Data / SM

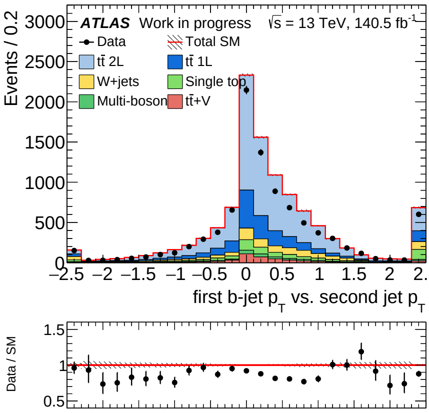

ATLAS Work in progress
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
lepton p_T vs. second jet p_T
Data / SM

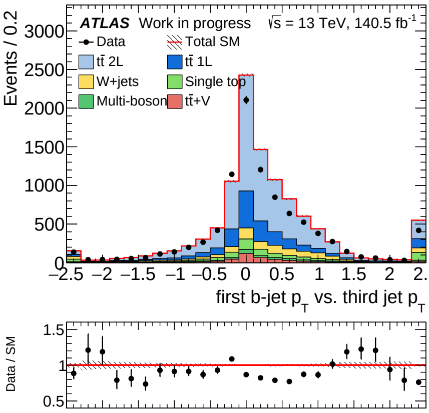

ATLAS Work in progress
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
lepton p_T vs. third jet p_T
Data / SM

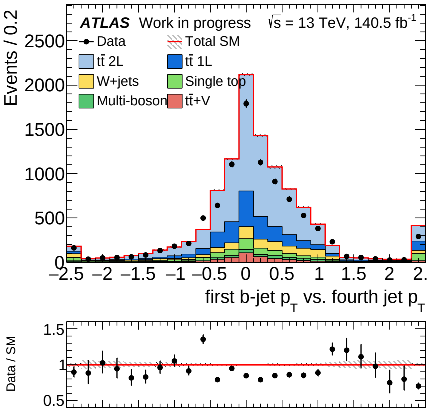

ATLAS Work in progress
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
lepton p_T vs. fourth jet p_T
Data / SM

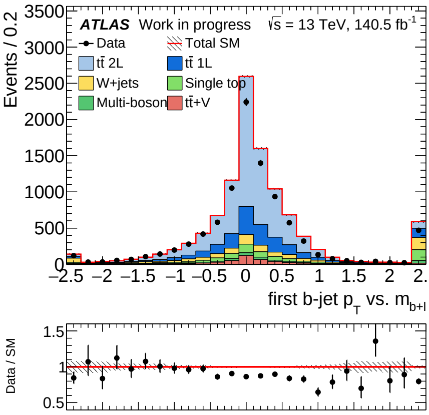

ATLAS Work in progress
√s = 13 TeV, 140.5 fb^{-1}
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
lepton p$_T$ vs. m$_{b+l}$
Data / SM
0
500
1000
1500
2000
2500
3000
3500
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5
0.5
1
1.5

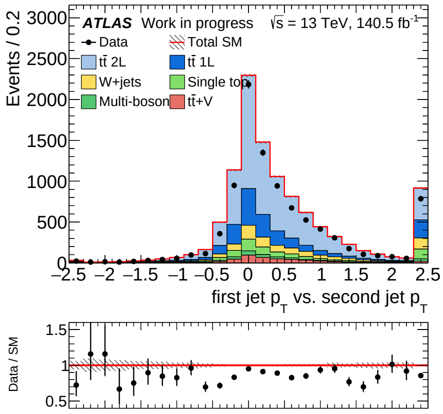

ATLAS Work in progress
√s = 13 TeV, 140.5 fb^{-1}
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
first b-jet p$_T$ vs. first jet p$_T$
Data / SM
0
500
1000
1500
2000
2500
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5
0.5
1
1.5

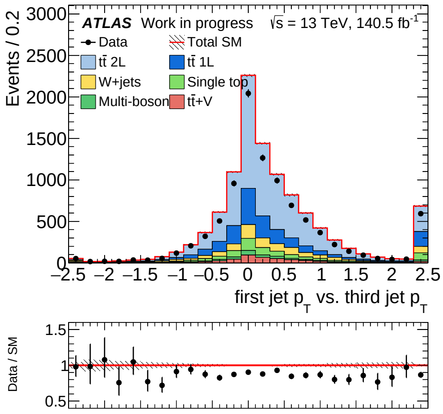

ATLAS Work in progress
√s = 13 TeV, 140.5 fb^{-1}
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
first b-jet p$_T$ vs. second jet p$_T$
Data / SM
0
500
1000
1500
2000
2500
3000
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5
0.5
1
1.5

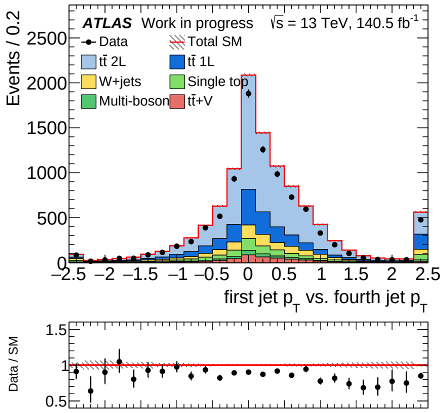

ATLAS Work in progress
√s = 13 TeV, 140.5 fb^{-1}
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
first b-jet p$_T$ vs. third jet p$_T$
Data / SM
0
500
1000
1500
2000
2500
3000
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5
0.5
1
1.5

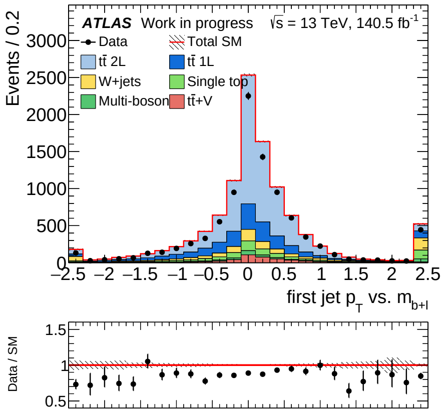

ATLAS Work in progress
√s = 13 TeV, 140.5 fb^{-1}
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
first b-jet p$_T$ vs. fourth jet p$_T$
Data / SM
0
500
1000
1500
2000
2500
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5
0.5
1
1.5

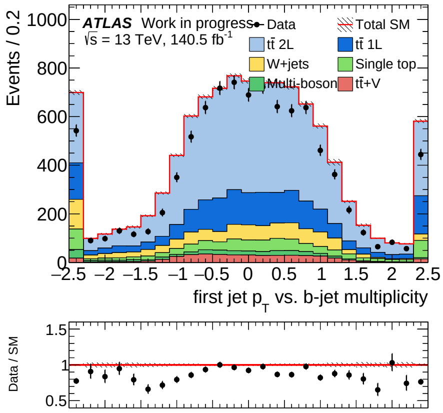

ATLAS Work in progress
√s = 13 TeV, 140.5 fb^{-1}
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
first b-jet p$_T$ vs. m$_{b+l}$
Data / SM
0
500
1000
1500
2000
2500
3000
3500
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5
0.5
1
1.5

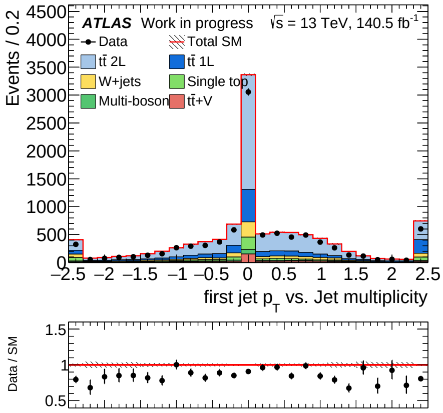

ATLAS Work in progress
√s = 13 TeV, 140.5 fb-1
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
first jet pT vs. second jet pT
Data / SM

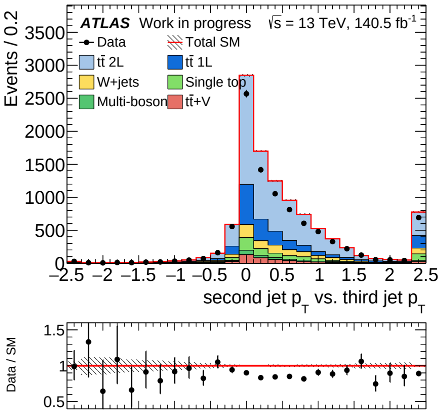

ATLAS Work in progress
√s = 13 TeV, 140.5 fb-1
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
first jet pT vs. third jet pT
Data / SM

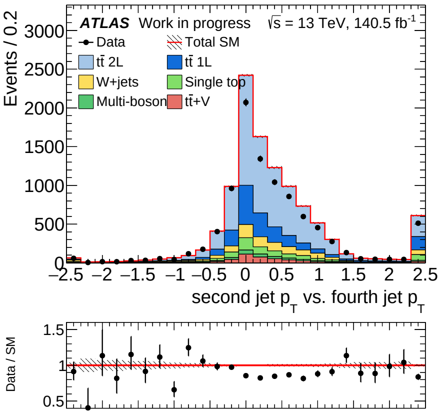

ATLAS Work in progress
√s = 13 TeV, 140.5 fb-1
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
first jet pT vs. fourth jet pT
Data / SM

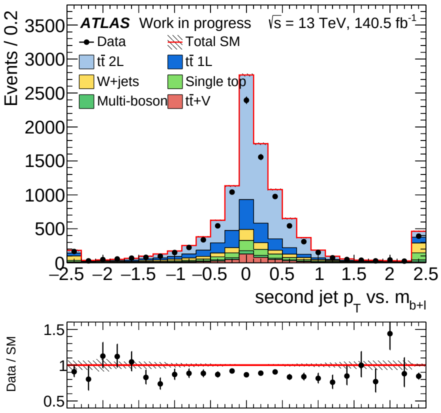

ATLAS Work in progress
√s = 13 TeV, 140.5 fb-1
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
first jet pT vs. mb+l
Data / SM

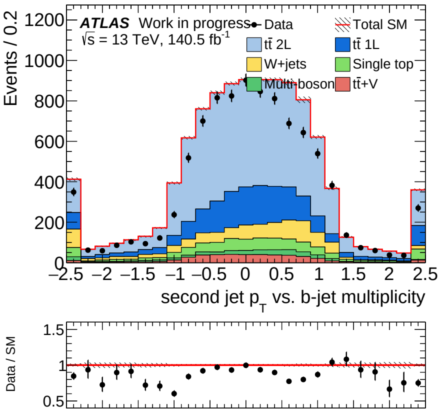

ATLAS Work in progress
√s = 13 TeV, 140.5 fb-1
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
first jet pT vs. b-jet multiplicity
Data / SM

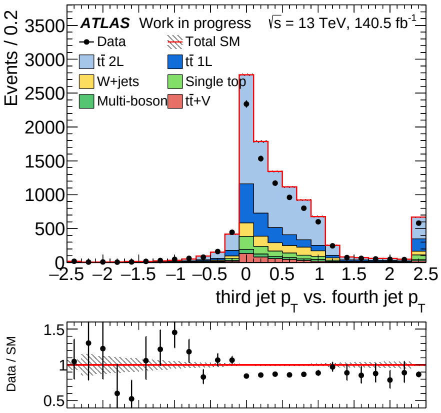

ATLAS Work in progress
√s = 13 TeV, 140.5 fb-1
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
first jet pT vs. Jet multiplicity
Data / SM

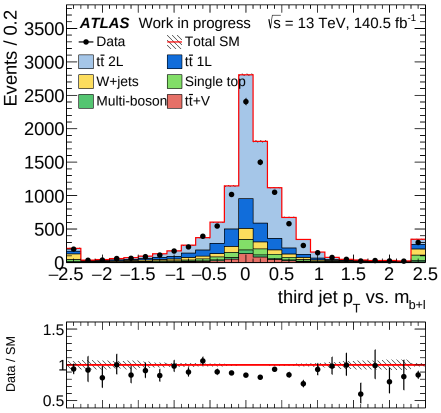

ATLAS Work in progress
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
second jet p_T vs. third jet p_T
Data / SM


ATLAS Work in progress
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
second jet p_T vs. fourth jet p_T
Data / SM


ATLAS Work in progress
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
second jet p_T vs. m_{b+l}
Data / SM


ATLAS Work in progress
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
second jet p_T vs. b-jet multiplicity
Data / SM


ATLAS Work in progress
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
third jet p_T vs. fourth jet p_T
Data / SM


ATLAS Work in progress
√s = 13 TeV, 140.5 fb⁻¹
Data
Total SM
tt̄ 2L
tt̄ 1L
W+jets
Single top
Multi-boson
tt̄+V
Events / 0.2
third jet p_T vs. m_{b+l}
Data / SM

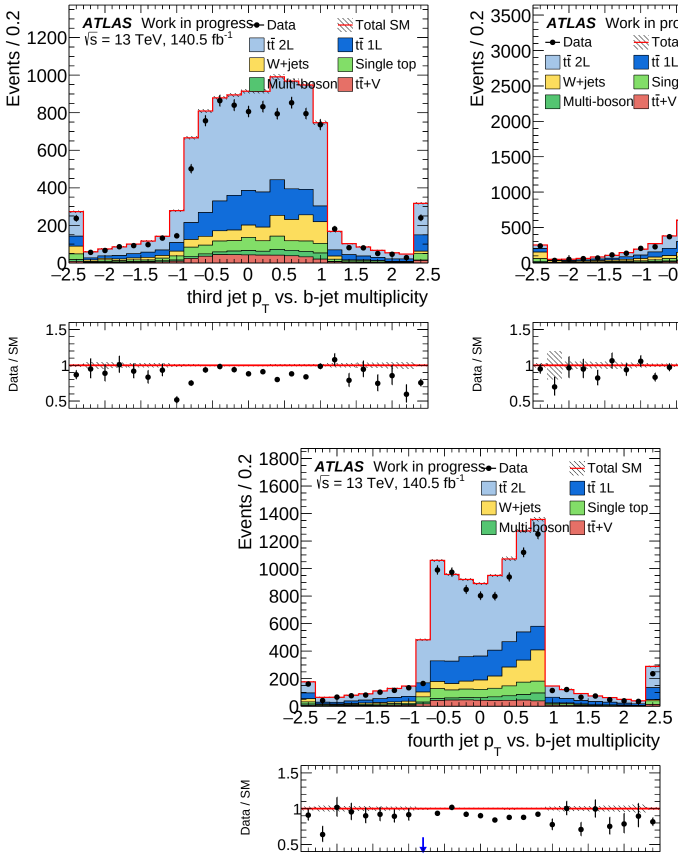
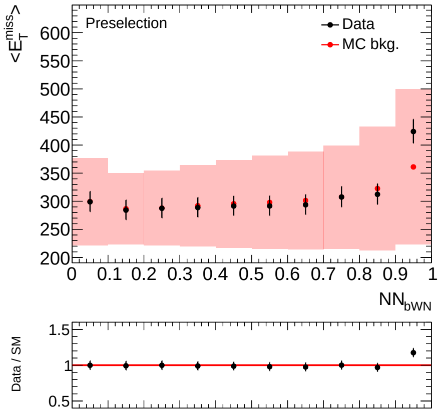
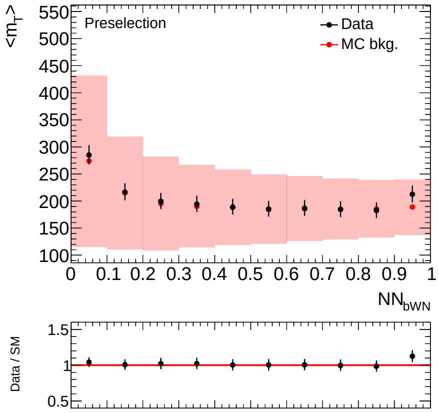


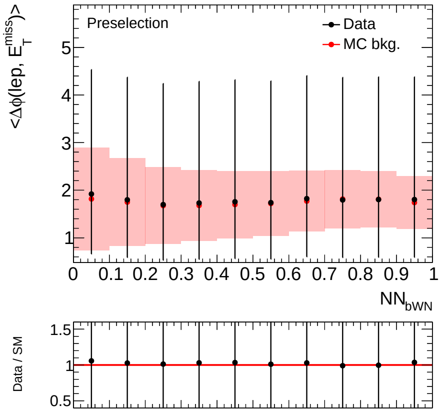
Figure A.2: Distributions of the linear correlation between two kinematic observables, respectively. Agreement between data and simulated background events is observed. Only statistical uncertainties are shown. Underflows and overflows are included in the first and last bins.

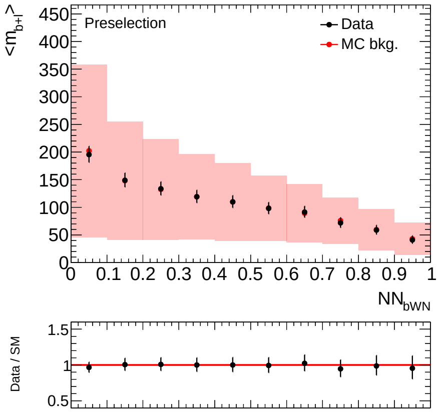
A.4 Profile plots

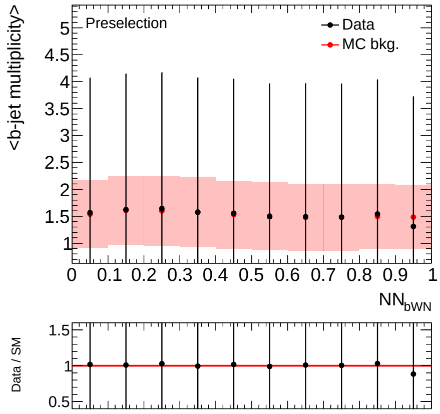
The dependence of the average of the input variables to the classifier output score is compared to data and the SM expectation. Overall, good agreement between data and simulated background events is observed. The error bands represent the standard deviation (RMS) of the distribution in the corresponding bin.

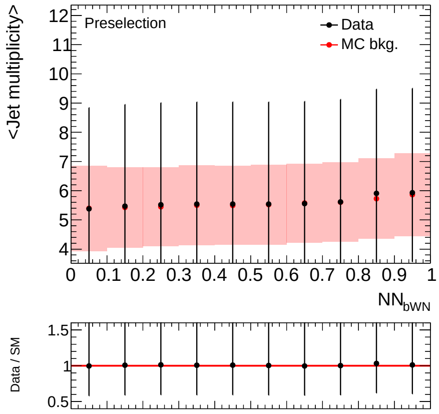


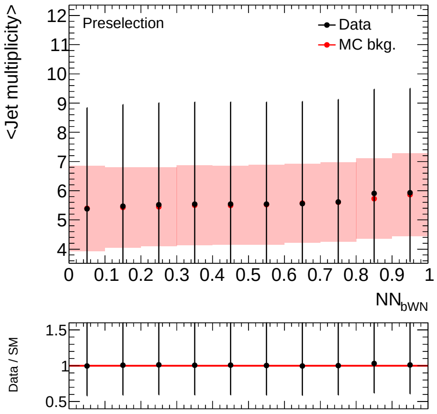


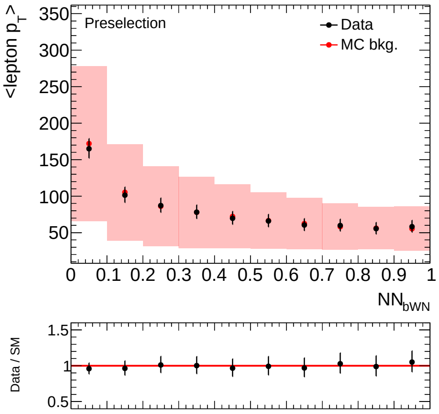


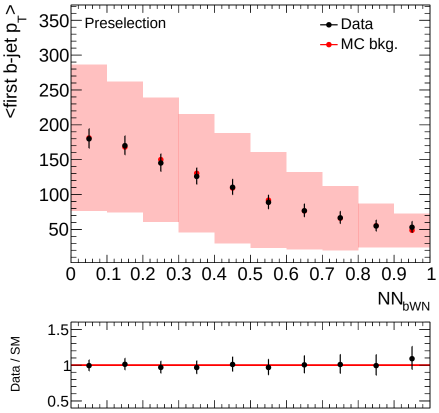


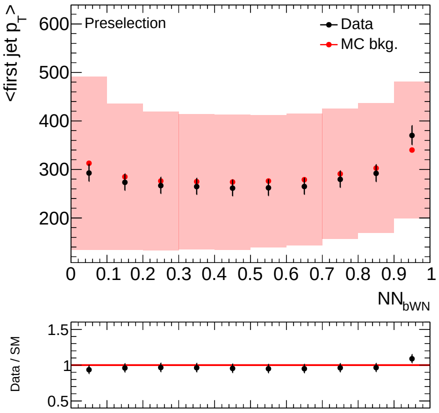


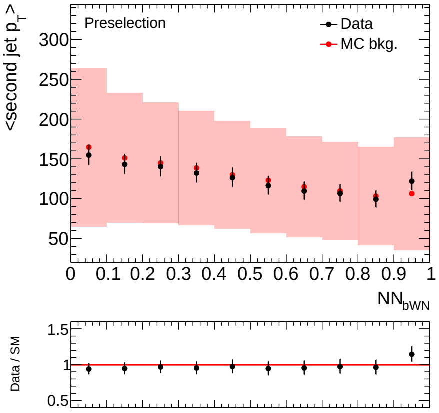


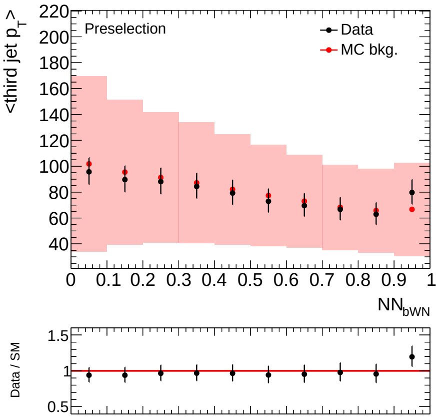

<Jet multiplicity>
Preselection
Data
MC bkg.
4
5
6
7
8
9
10
11
12
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
NN_{bWN}
Data / SM
0.5
1
1.5


<lepton p_T>
Preselection
Data
MC bkg.
50
100
150
200
250
300
350
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
NN_{bWN}
Data / SM
0.5
1
1.5


<first b-jet p_T>
Preselection
Data
MC bkg.
50
100
150
200
250
300
350
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
NN_{bWN}
Data / SM
0.5
1
1.5


<first jet p_T>
Preselection
Data
MC bkg.
200
300
400
500
600
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
NN_{bWN}
Data / SM
0.5
1
1.5


<second jet p_T>
Preselection
Data
MC bkg.
50
100
150
200
250
300
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
NN_{bWN}
Data / SM
0.5
1
1.5


<third jet p_T>
Preselection
Data
MC bkg.
40
60
80
100
120
140
160
180
200
220
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
NN_{bWN}
Data / SM
0.5
1
1.5

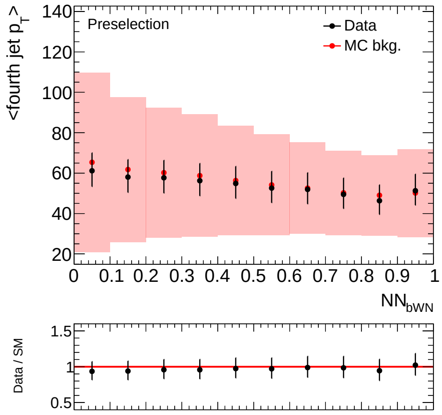


Figure A.2: Distribution of the mean of the sensitive input variable as a function of $\mathrm{NN}_{\mathrm{bWN}}$ for data and simulated SM events. Good agreement between observed and simulated events is observed. The error bands represent the standard deviation (RMS) of the respective variable in the corresponding bin.

A.5 Further details on the alternative analysis techniques

Further details about the alternative analysis strategies which are compared in Section 8.2.1 are given in the following. The alternative analysis methods incorporate either different kinematic variables for signal region definitions or different input variables are used to train the alternative multivariate algorithms.

`bWN-subSet (re-optimised)`

This analysis uses the identical set of discriminating variables as `bWN-subSet`, but for the signal region definition the dedicated thresholds have been optimised for the larger integrated luminosity.

	Signal region definition
$E_{\mathrm{T}}^{\mathrm{miss}}$	$> 300\,\mathrm{GeV}$
m_{T}	$> 130\,\mathrm{GeV}$
am_{T2}	$< 100\,\mathrm{GeV}$
$\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$	< 2

Table A.1: Overview of an alternative signal region selection. In addition, the event preselection as defined in Table 8.1 is applied.

`Cut-based`

In this alternative analysis the same 12 input variables as described for final $\mathrm{NN}_{\mathrm{bWN}}$ approach (see Table 8.2), but the signal region is based on particular requirements on the different variables.

`NN (alt. input)`

The signal region in this method is based on the threshold on the output score of a neural network at which the discovery significance is maximal. The network architecture is the same as listed in Table 8.3 and the input variables of the network are listed in Table A.3.

	Signal region definition
$E_{\mathrm{T}}^{\mathrm{miss}}$	$> 400\,\mathrm{GeV}$
m_{T}	$> 150\,\mathrm{GeV}$
$\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$	< 2
m_{bl}	$< 80\,\mathrm{GeV}$
leading b-jet p_{T}	$< 80\,\mathrm{GeV}$
leading jet p_{T}	$> 300\,\mathrm{GeV}$

Table A.2: Overview of an alternative signal region selection. In addition, the event preselection as defined in Table 8.1 is applied. Optimisation of the requirements of the other input variables used to train $\mathrm{NN}_{\mathrm{bWN}}$ had no effect on the discovery significance.

Input variables	Description
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse energy
m_{T}	Transverse mass
$\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$	Azimuthal angle between the direction $E_{\mathrm{T}}^{\mathrm{miss}}$ and the lepton
m_{bl}	Invariant mass of the leading b-tagged jet and the lepton
$p_{\mathrm{T}}(\ell)$	Transverse momentum of the lepton
$\eta(\ell)$	Pseupdorapidity of the lepton
leading jet p_{T}	Transverse momentum of the leading jet
leading jet η	Pseupdorapidity of the leading jet
leading b-jet p_{T}	Transverse momentum of the leading b-tagged jet
leading b-jet η	Pseupdorapidity of the leading b-jet
n_{jet}	Jet multiplicity
$n_{b-\mathrm{jet}}$	b-jet multiplicity

Table A.3: Discriminating variables applied as inputs to train an alternative neural network.

`NN incl. 4 jets`

The signal region in this method is based on the threshold on the output score of a neural network at which the discovery significance is maximal. The network architecture is the same as listed in Table 8.3. In addition to the 12 input variables listed in Table 8.2, the 4-vectors (p_{T}, η, ϕ, E) of the four leading jets are applied as input variables but the jet-RNN is not used as an input to the neural network.

`BDT`

For this method, a Boosted Decision Tree has been trained to classify signal and background. The signal region in this method is based on the threshold on the output score of the BDT at which the discovery significance is maximal. The exact same input variables listed in Table 8.2 are used to train the BDT, the jet-RNN is omitted. The parameters of the BDT are listed in Table A.4.

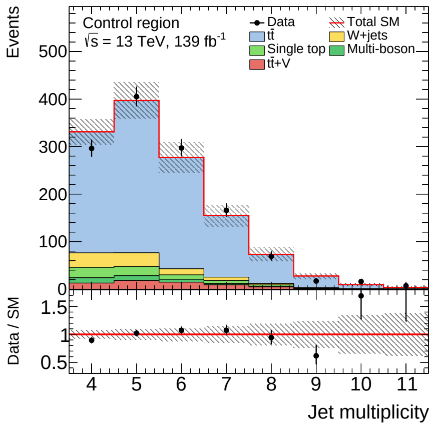
`NN w/o jet-RNN`

The signal region in this method is based on the threshold on the output score of a neural network at which the discovery significance is maximal. The network architecture is the same as listed in Table 8.3. The exact same input variables listed in Table 8.2 are used to train the neural network but the jet-RNN is not used as an input to the neural network.

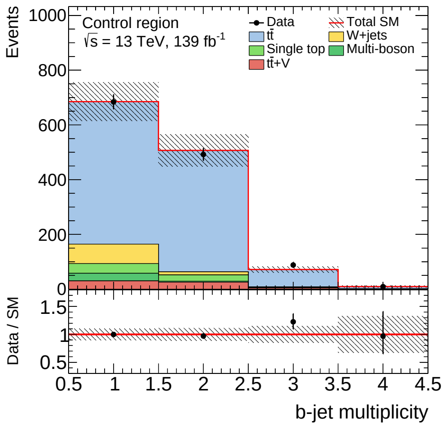
Architecture	Parameter set
Number of trees	1000
Maximum depth of trees	3
Learning rate	0.1
Regularisation	$L2$ ($\lambda = 1$)

Table A.4: Architecture and parameters of the neural network.

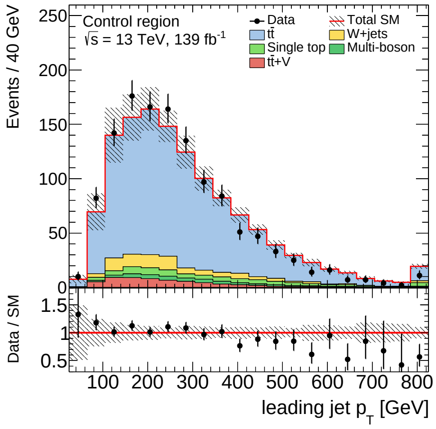
A.6 Kinematic distributions of input variables in the control region



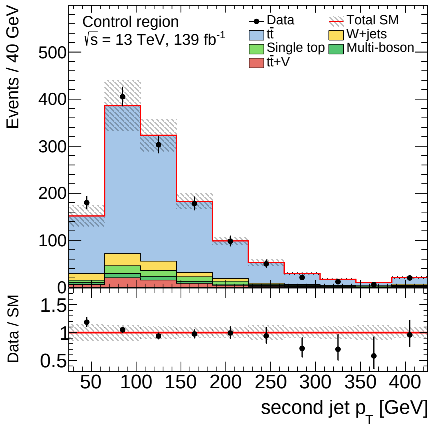
(a) Jet multiplicity



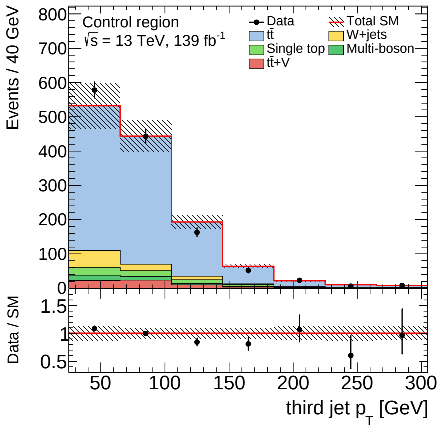
(b) b-jet multiplicity



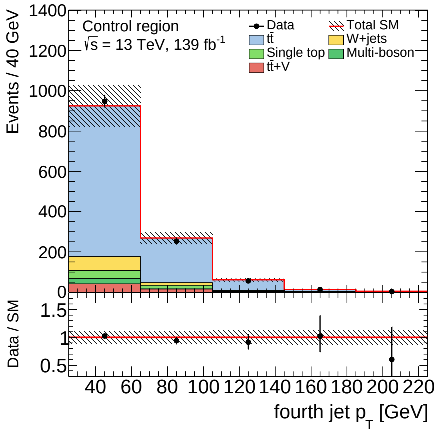
(c) Leading jet p_{T}



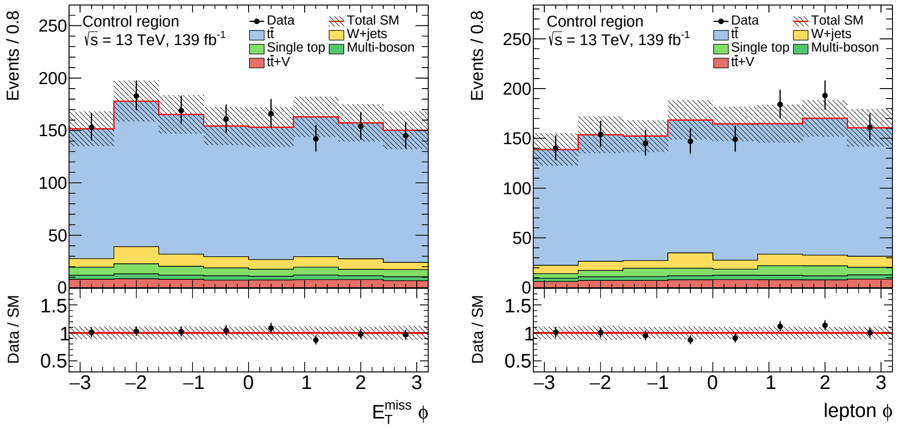
(d) 2^{nd} leading jet p_{T}



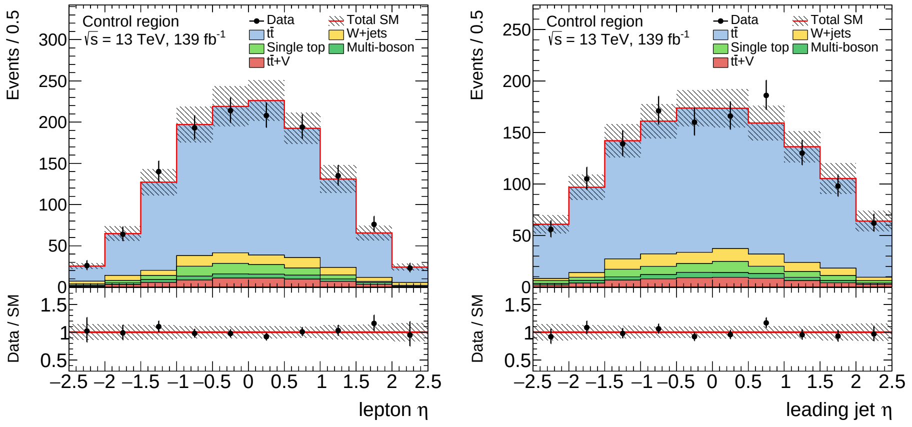
(e) 3^{rd} leading jet p_{T}



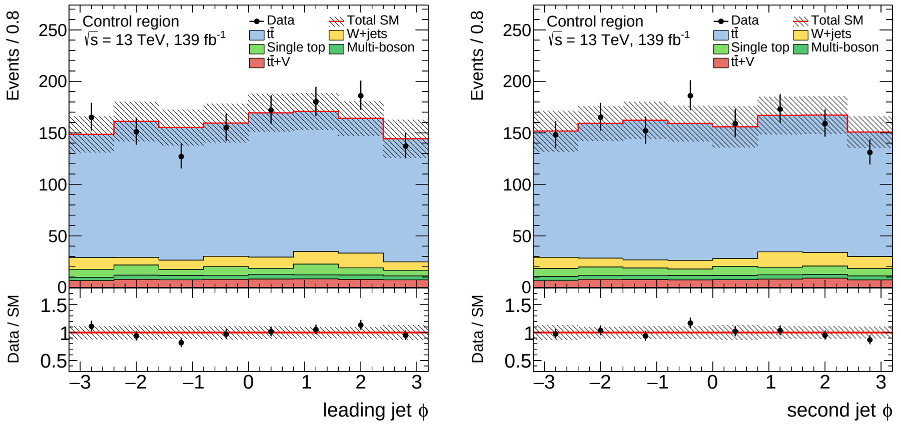
(f) 4^{th} leading jet p_{T}



(g) $\phi(E_{\mathrm{T}}^{\mathrm{miss}})$ (h) $\phi(\ell)$



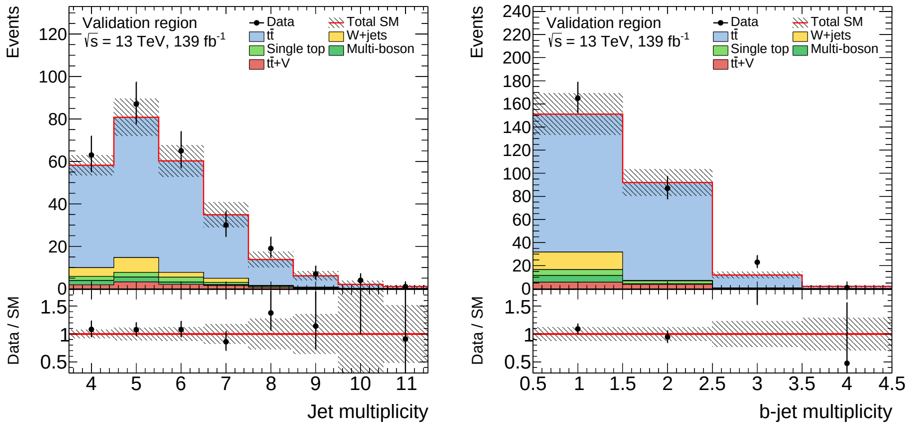
(i) $\eta(\ell)$ (j) Leading jet η



(k) Leading jet ϕ (l) 2^{nd} leading jet ϕ

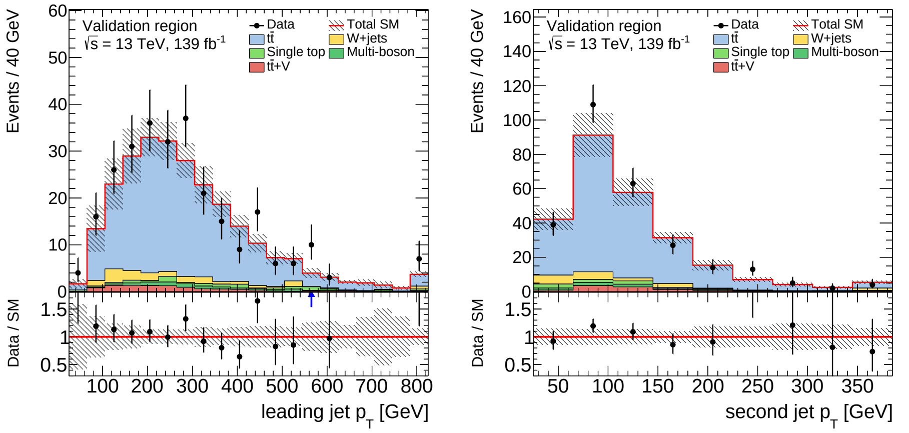
Figure A.3: Kinematic distributions of selected observables in the control region. The $t\bar{t}$ background prediction is scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross section. The hatched band around the total SM prediction and the hatched band in the data/SM ratio include statistical and systematic uncertainties, except for the $t\bar{t}$ modelling uncertainties, which are only defined for the signal and validation region. Overflows are included in the last bin.

A.7 Kinematic distributions of input variables in the validation region



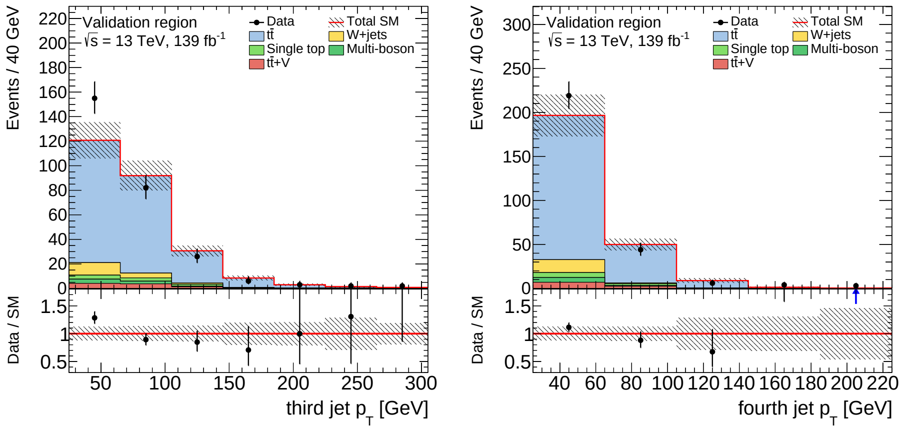
(a) Jet multiplicity

(b) b-jet multiplicity



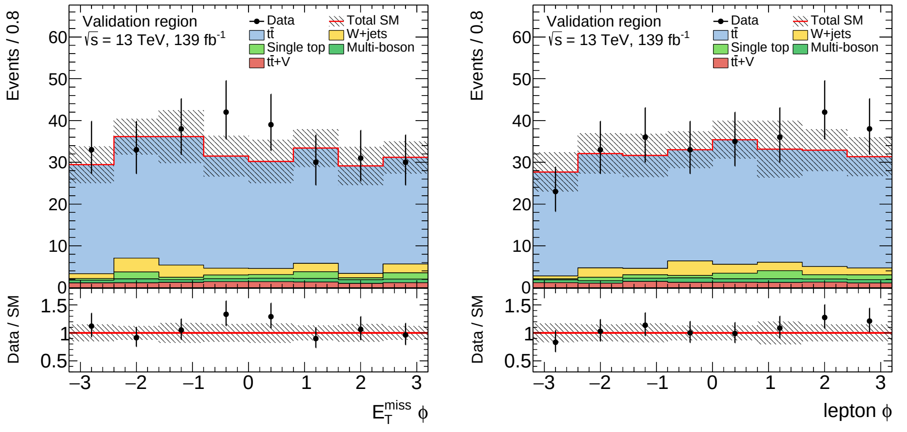
(c) Leading jet p_{T}

(d) 2^{nd} leading jet p_{T}

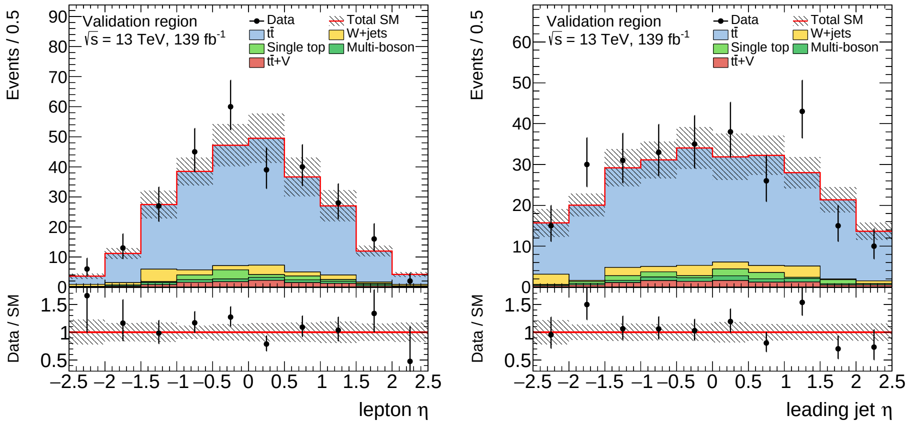


(e) 3^{rd} leading jet p_{T}

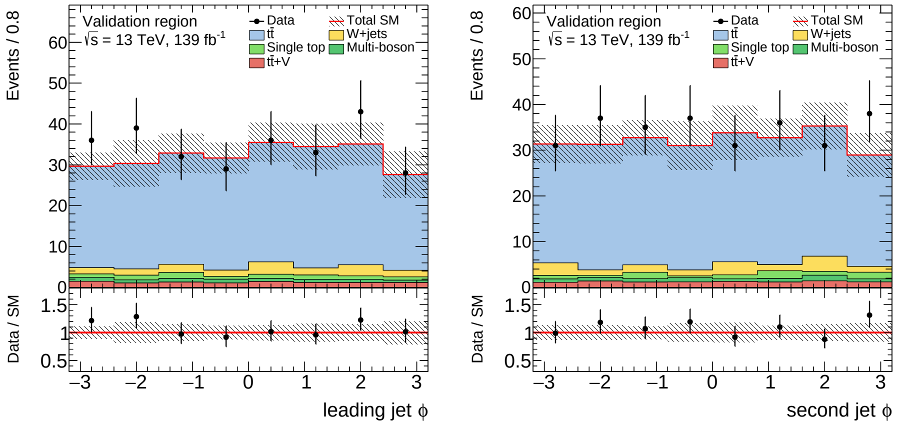
(f) 4^{th} leading jet p_{T}



(g) $\phi(E_{\mathrm{T}}^{\mathrm{miss}})$ (h) $\phi(\ell)$



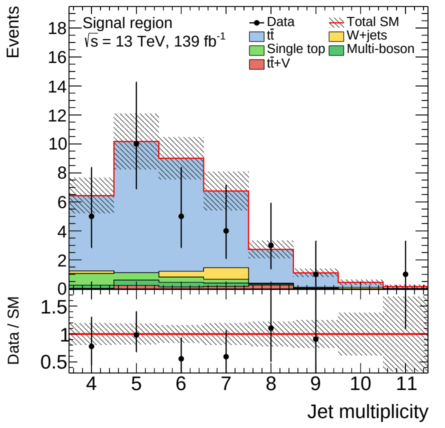
(i) $\eta(\ell)$ (j) Leading jet η



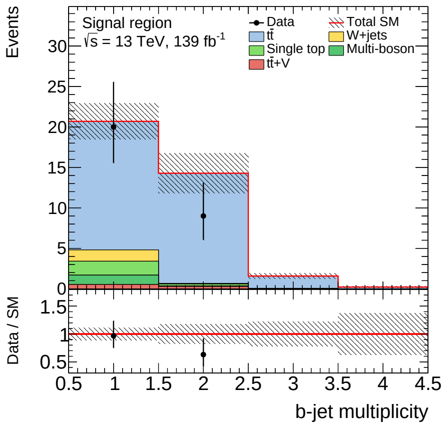
(k) Leading jet ϕ (l) 2^{nd} leading jet ϕ

Figure A.4: Kinematic distributions of selected observables in the validation region. The $t\bar{t}$ background processes are scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross section. The shaded band around the total SM prediction as well as the shaded band in the data/SM ratio include statistical and systematic uncertainties. Overflows are included in the last bin.

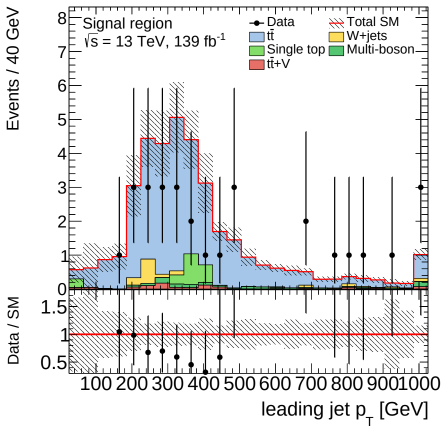
A.8 Kinematic distributions of input variables in the signal region



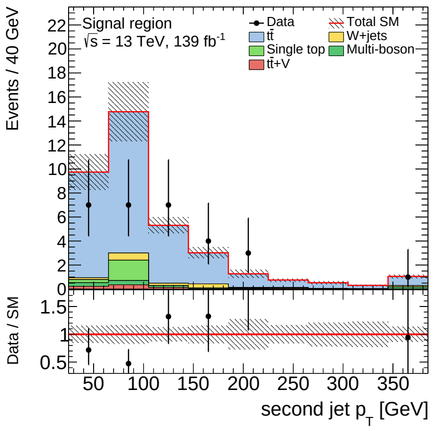
(a) Jet multiplicity



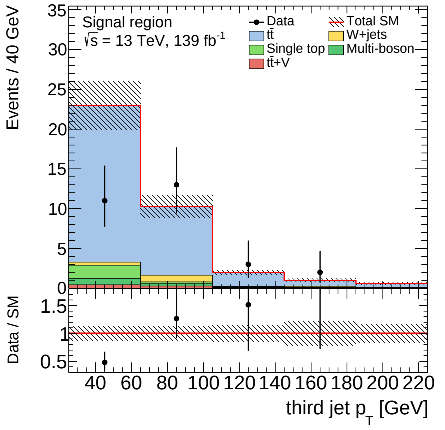
(b) b-jet multiplicity



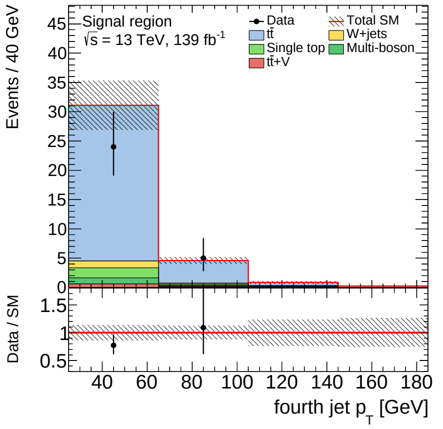
(c) Leading jet p_{T}



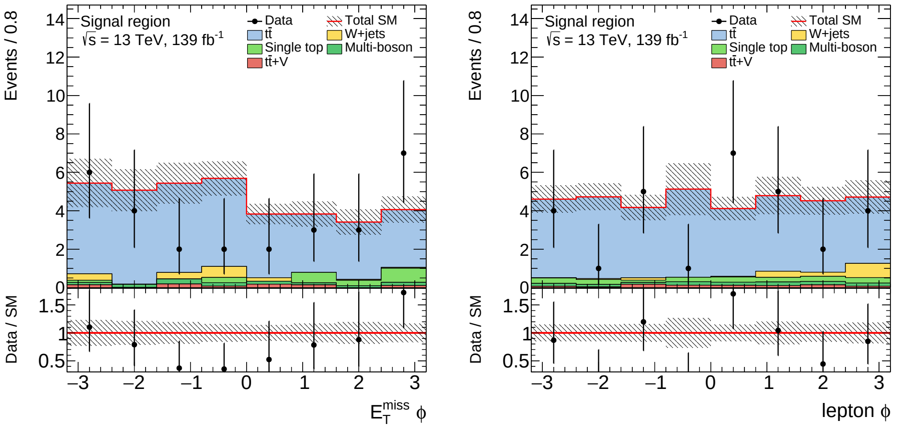
(d) 2^{nd} leading jet p_{T}



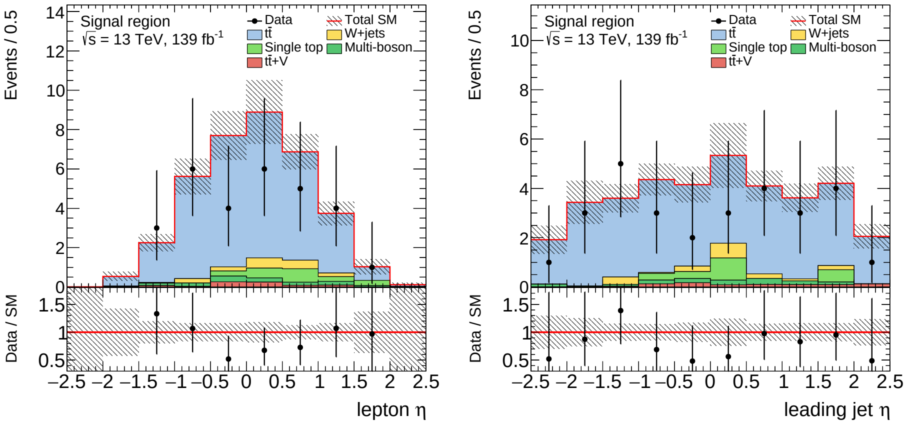
(e) 3^{rd} leading jet p_{T}



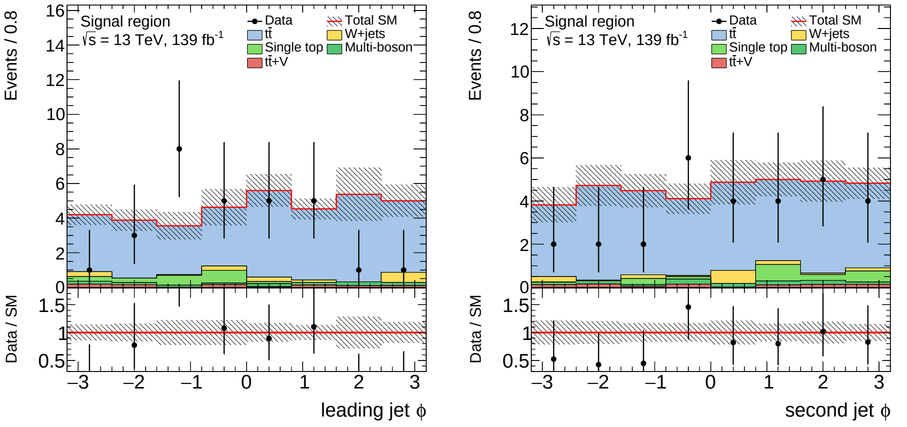
(f) 4^{th} leading jet p_{T}



(g) $\phi(E_{\mathrm{T}}^{\mathrm{miss}})$ (h) $\phi(\ell)$



(i) $\eta(\ell)$ (j) Leading jet η



(k) Leading jet ϕ (l) 2^{nd} leading jet ϕ

Figure A.5: Kinematic distributions of selected observables in the signal region. The $t\bar{t}$ background processes are scaled by the normalisation factor obtained from a likelihood fit of the CR, while all other SM background processes are normalised with the respective theoretical cross section. The shaded band around the total SM prediction as well as the shaded band in the data/SM ratio include statistical and systematic uncertainties. Overflows are included in the last bin.