UiO : University of Oslo

Knut Oddvar Høie Vadla

Search for production of charginos and neutralinos in dilepton final states with the ATLAS detector at the LHC

Thesis submitted for the degree of Philosophiae Doctor

Department of Physics
Faculty of Mathematics and Natural Sciences

2022

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2563

ISSN 1501-7710

Print production: Graphics Center, University of Oslo.

Abstract

Searches for supersymmetric particles, in particular charginos ($\tilde{\chi}_1^{\pm}$) and neutralinos ($\tilde{\chi}_1^0$ and $\tilde{\chi}_2^0$), have been carried out using data from proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. The data were collected by the ATLAS experiment during the LHC Run 2.

The first search considers $\tilde{\chi}_1^{\pm}\ \tilde{\chi}_1^{\mp}$ pair-production with decays via sleptons and sneutrinos into final states with two leptons (e, μ) and missing transverse energy from two $\tilde{\chi}_1^0$s and two neutrinos. In the simplified models considered in this thesis, $\tilde{\chi}_1^0$ is taken to be the lightest supersymmetric particle and a dark matter candidate. Sensitivity studies are presented for 6 and $10\,\text{fb}^{-1}$, in addition to observed exclusion limits in the mass-plane of $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_1^0$ with $13.3\,\text{fb}^{-1}$. The exclusion reach for the $\tilde{\chi}_1^{\pm}$ mass increases up to 620 GeV (for a massless $\tilde{\chi}_1^0$), compared to 480 GeV with the Run 1 data at $\sqrt{s} = 8\,\text{TeV}$, whereas the exclusion reach for the $\tilde{\chi}_1^0$ mass increases from 180 GeV in Run 1 up to 260 GeV with these early data of Run 2.

The second search considers production of a mass-degenerate $\tilde{\chi}_1^{\pm}\ \tilde{\chi}_2^0$ pair with decays via a W and a Z boson into two leptons, two jets and missing transverse energy from two $\tilde{\chi}_1^0$s. Sensitivity studies are presented for $36.5\,\text{fb}^{-1}$, and observed exclusion limits in the plane of the $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$ masses are obtained with $36.1\,\text{fb}^{-1}$. The exclusion reach for the $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ mass increases from 420 GeV with Run 1 data to 580 GeV (for a massless $\tilde{\chi}_1^0$) with the 2015+2016 dataset of Run 2, whereas the exclusion reach for the $\tilde{\chi}_1^0$ mass increases from 150 GeV to 230 GeV w.r.t. the Run 1 limits.

The third search is a continuation of the previous one, but redefines and optimizes the signal regions for the full Run 2 dataset. Sensitivity studies and background estimation are presented for $139\,\text{fb}^{-1}$, with expected exclusion limits in the mass-plane of $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$. The expected mass exclusion reach increases from 580 to 750 GeV for $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ and from 230 to 320 GeV for $\tilde{\chi}_1^0$ w.r.t. the expected exclusion limits obtained for the analysis with $36.1\,\text{fb}^{-1}$. The complete ATLAS publication, which incorporates the results just mentioned, goes beyond the expectations and excludes $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ masses up to 820 GeV and $\tilde{\chi}_1^0$ masses up to 380 GeV.

The fourth and final search investigates whether machine learning, or multivariate analysis, with the XGBoost algorithm can improve the sensitivity to the signal model of the two previous searches, with the integrated luminosity of the full Run 2 dataset. On the one hand, the single-bin XGBoost signal regions increase the expected exclusion reach from 600 to 700 GeV for $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ and from 150 to 350 GeV for $\tilde{\chi}_1^0$, compared to conventional single-bin discovery regions. On the other hand, the signal sensitivity of the XGBoost signal regions are comparable to the conventional multi-bin signal regions, suggesting that there may be more sensitivity to be gained by also considering multi-bin shape-fits for the XGBoost output.

Acknowledgements

First of all I would like to thank my main supervisor Prof. Farid Ould-Saada for giving me the opportunity to research some of the most interesting topics I can think of, and for having supported me through both ups and downs. I have really appreciated the times I have spent with you and your family at dinner parties at Nesodden and at the "Oslo house" in Saint-Genis, summertime and "weeks" at CERN, and even watching Tour de France live in the Jura mountains. Those are good memories.

To my second supervisor, Eirik Gramstad, thank you for all your help with the nitty-gritty details of SUSY analysis, such as figuring out how to configure SUSYTools, how to debug failed grid jobs and for inside tips on which p-tags to choose for derivation samples. I have also very much enjoyed your good humor and your engagement in the Running Couplings, which included introduction slides to all the couplings, you running for multiple teams simultaneously and providing excessive statistics and plots to analyze the performance the Monday after Holmenkollstafetten.

I also want to give a thanks to James Catmore for sparking my interest in machine learning and for providing a code base that I built upon for the ML part of this thesis. And for graciously lending me his clothes after I made the terrible decision to fly via Charles de Gaulle on my way to CERN once.

It has been a privilege being a part of the high-energy physics community at CERN. I would like to thank the members of the ATLAS Collaboration that I have had the pleasure of working with, in particular Ben Hooberman, Jonathan Long, Matt Zhang, Rupert Tombs, Sarah Williams, Jason Oliver, Abhishek Sharma, Arka Santra and the rest of the SUSY 2 leptons + jets working group.

I am also grateful for having been a part of the Strategic Dark Matter Initiative (SDI) at UiO, which provided funding for my position. I learned a lot from attending the dark matter lunches and seminars.

The people of the HEP group at UiO have meant a lot during my time as a PhD student. In particular, I want to mention my fellow PhD students Simen, Eli, Jeriek, Even, Kristian, Andreas, Eirik H., Vanja, Carl Andreas, Henrik, Gert and Victor. Thank you for the enjoyable office chats, lunch breaks, schools, conferences and other less productive activities (did anyone say quiz?). I would also like to extend a nod to our sporty colleagues across the mountain, Inga and Steffen. And a special thanks to Oda, for being so supportive and patient with me.

My new colleagues at the Norwegian Police ICT Services also deserve thanks for being patient, while I have spent ages finishing up the thesis.

Finally, I want to thank my family and friends for always being so supportive and encouraging.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Large Hadron Collider (LHC) had its first successfull operational period of particle collisions between November 2009 and February 2013, referred to as Run 1. In March 2010, the record-breaking center-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ was finally reached in proton-proton collisions. This was followed by a period of data-taking for physics analysis during most of 2011. At the end of 2011, the collision energy was raised again, to $\sqrt{s} = 8\,\text{TeV}$, with another period of data-taking for physics analysis, which ended in early 2013.

During the LHC Run 1, the ATLAS experiment collected about $4.7\,\text{fb}^{-1}$ of integrated luminosity from proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$, and about $20\,\text{fb}^{-1}$ of integrated luminosity at $\sqrt{s} = 8\,\text{TeV}$, which were of sufficiently high quality to be used for physics analysis. In 2012, using the data from Run 1, the ATLAS and CMS experiments were able to declare discovery of the final missing piece of the Standard Model (SM) of particle physics, the Higgs boson, almost 50 years after it was first theorized in 1964.

The LHC Run 1 was followed by a two year long shutdown in preparation for Run 2. In the Spring of 2015, the LHC started the new run of proton-proton collisions at yet another record-breaking energy level of $\sqrt{s} = 13\,\text{TeV}$. This became the operational collision energy for protons throughout the whole of Run 2, which came to a close at the end of 2018. The work of this thesis makes use of the data collected by the ATLAS experiment during the LHC Run 2.

The center-of-mass energy of the colliding initial-state particles corresponds to an upper limit on the mass that can be produced for any final-state particle. I.e., the increase in collision energy from Run 1 to Run 2 thereby increased the upper limit on the particle masses that could be probed in the collision events. This fact sparked hope of discovering new particles which have been proposed in theories that go beyond the Standard Model. One of the most popular theoretical models beyond the SM is Supersymmetry, which adds a so-called superpartner to all the particles in the SM. Since such superpartners have not yet been discovered, they must be heavier than their SM counterparts. Could any of these hypothetical particles be present in the extended mass-range accessible by the LHC in Run 2?

In this thesis, I present the work I have done as part of working groups in the ATLAS experiment searching for Supersymmetry in various production mechanisms and final-states. The searches have been conducted within the framework of simplified models, with only two free mass parameters in each model. My main focus has been on production of charginos and neutralinos in electroweak interactions, with a signature of two leptons, hadronic jets and missing transverse energy in the final-state. The lightest neutralino, which in this thesis is assumed to be the lightest supersymmetric particle, is a prime

example of a so-called weakly interacting massive particle (WIMP), and has been considered among the most promising candidates for dark matter.

My work has been centered on the task of finding event and particle selections that optimize the sensitivity to the signal models. We refer to these selections as signal regions. I consider two different approaches to defining the signal regions. In the three first analyses of the thesis, I make use of the conventional approach of applying so-called "rectangular" cuts to kinematic variables, such as the transverse momenta of the visible particles and on the missing transverse energy, in order to separate signal from background. In the fourth and final analysis, I present a feasibility study which instead makes use of a machine learning approach, in the form of gradient boosted decision trees, to learn more refined, multivariate decision boundaries between signal and background.

The thesis is presented in two main parts. The first part, chapter 1 through 4, introduces the theoretical motivation and contents of Supersymmetry (chapter 2), the facilities and kinematics of particle collisions at the LHC (chapter 3), and the experimental data which is collected and reconstructed by the ATLAS experiment (chapter 4). The second part, chapter 5 through 9, first outlines the general ingredients and workflow of the physics analyses to come (chapter 5), before presenting my main research contributions. Chapter 6 presents my first analysis, where I estimated the expected sensitivity for chargino pair production with decays via sleptons and sneutrinos into two leptons and missing transverse energy from two neutrinos and two lightest neutralinos in the final-state, with the amount of data we expected to have in time for the summer conferences of 2016. In chapter 7, I move on to another signal model, namely production of a mass-degenerate chargino-neutralino pair with decays via a W and Z boson into final-states with two leptons, two jets and missing transverse energy from two lightest neutralinos, using the ATLAS dataset from 2015 and 2016, where my work was on defining and optimizing signal regions. Chapter 8 deals with the same signal model as chapter 7, but in this case the signal regions are redefined and optimized for the full Run 2 dataset. Finally, in chapter 9, I investigate whether machine learning in the form of binary classification with the XGBoost algorithm can improve the sensitivity to the signal model considered in chapter 7 and 8, with the integrated luminosity of the full Run 2 dataset.

Part I

Foundation

Chapter 2

Supersymmetry

2.1 The Standard Model

In order to understand Supersymmetry (SUSY), we first need to know the Standard Model (SM) of particle physics, which sums up our current knowledge about the elementary particles and how they interact; and we need to know its limitations, which motivate the search for a more complete theory that can explain more of what we observe.

2.1.1 The elementary particles

The elementary particles of the SM are neatly illustrated in Figure 2.1. We can broadly put these particles into two categories: (i) the fermions, or *matter particles*, and (ii) the bosons, or *force carriers*.

The fermions, which are characterized by having the spin quantum number $s = 1/2$, come in three so-called generations, where the second and third generation particles are basically increasingly heavier copies of the particles in the first generation[1]. The first generation fermions consist of two quarks: called up (u) and down (d), and two leptons: the electron (e) and the electron neutrino

[1] We do not know if this holds for the neutrinos, since their masses are still unknown.

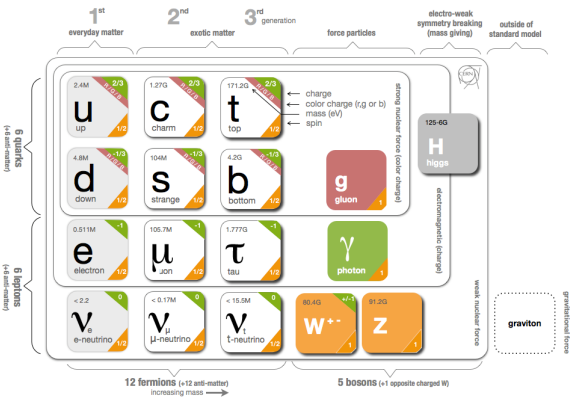


Figure 2.1: A schematic overview of the elementary particles that we currently know of, and which are the building blocks of the Standard Model of particle physics [1].

(ν_e). The up- and down quarks make up the protons (uud) and neutrons (udd) in the atomic nuclei, whereas the electrons complete the atoms by forming electron clouds around the nuclei. The final first-generation fermion, the electron neutrino, is, e.g., emitted in decays of radioactive nuclei. The electron, the electron neutrino and the up quark are stable, because of lepton and baryon number conservation in the SM, and can thereby not decay further to other particles.

The neutrinos have only got one type of charge, namely weak isospin, T_3 (the charge of the weak interacton, see Section 2.1.2), and they are almost massless[2]. The massive leptons, including the electron (e), the muon (μ) and the tau (τ), do in addition to the weak isospin also have electric charge. Finally, we have the quarks, which also hold a third, so-called *color*[3] charge.

In addition, all the fermions have anti-particles, with exactly the same mass and spin, but with all their internal charges conjugated (inverted).

The bosons, characterized by having an integer value of spin, do in the SM come in two forms: vector bosons with $s = 1$ and a scalar boson with $s = 0$. The vector bosons are the force mediating particles: the gluons (g) (eight different color-pair combinations) couple to color charged particles, the photon (γ) couples to electrically charged particles, whereas the $W^{\pm}$ and Z couple to weakly charged particles. Then, the final piece left in the SM is the only scalar particle we currently know of: the Higgs boson (H), which is the manifestation of the Higgs quantum field that causes most of the SM particles to acquire a mass.

2.1.2 The elementary forces

The SM is mathematically described in terms of quantum fields (one field for each type of particle), and gauge symmetry groups, which describe the transformations that the fields are allowed to undergo. The symmetry group of the SM can be expressed as a direct product of three independent gauge groups:

$$\mathrm{G_{SM}} = \mathrm{SU(3)}_c \times \mathrm{SU(2)_L} \times \mathrm{U(1)}_Y, \tag{2.1}$$

where $\mathrm{SU(3)}_c$ is the special unitary group of 3-dimensional color (c) transformations, $\mathrm{SU(2)}_L$ is the special unitary group of 2-dimensional weak isospin transformations (strictly between left-handed (L) particle [and right-handed anti-particle] states) and $\mathrm{U(1)}_Y$ is the unitary group of 1-dimensional weak hypercharge (Y) transformations.

The $\mathrm{SU(3)}_c$ group represents the *quantum chromodynamics* (QCD) [2, 3, 4, 5] or *strong* interaction between color charged quark states, as mediated by gluons. The gluons act on quark triplet states (3-dimensional vectors), corresponding

[2]The neutrinos are in fact assumed to be massless in the SM, in contrast to observations. The well established neutrino flavor oscillations require three distinct neutrino masses, as well as mixing between generations, but only upper bounds on their masses have been obtainable in measurements to date. See Section 2.2.4

[3]Not to be confused with the electromagnetic radiation with wavelengths in the range of the human visual perception. The term color in this context is only an analogy to the three colors red, green and blue which give a neutral result (white) when combined.

to the three color charges (*dimensions* of QCD), which are called *red* (r), *green* (g) and *blue* (b). The strong interaction exerts an attractive force between color states, which is responsible for binding the quarks inside baryons (combinations of rgb) and mesons (combinations of $r\bar{r}$, $g\bar{g}$ and $b\bar{b}$) together in *color neutral* states. This phenomenon is called *color confinement*. A consequence of this color interaction is so-called *asymptotic freedom*, where the individual color charged particles within a bound state (a hadron) are less affected by the strong interaction the closer they get to each other. On the other hand, when charged states are separated, e.g., in production of a quark-anti-quark pair, the attractive force between them increases with increasing distance. If the separating states are sufficiently energetic, new quarks may be produced from the high energy density between them, and contribute to form new color neutral hadrons. This is called *hadronization*, and produce what we often refer to as hadronic showers or jets.

The combined $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_Y$ group represents the *electroweak* gauge symmetry [6, 7, 8], which unifies the *electromagnetic* and *weak* interactions at energies above 246 GeV[4]. Above this electroweak unification scale, the $\mathrm{SU}(2)_\mathrm{L}$ group has three massless gauge bosons (generators of the symmetry) called W_1, W_2 and W_3, which act on weak isospin doublets (2-dimensional vectors) of neutrino and lepton pairs or up- and down-type quark pairs, and the $\mathrm{U}(1)_Y$ group has one massless gauge boson called B, which acts on single fermion states (1-dimensional/single particle states). The electric charge, Q, of the electromagnetic interaction, the weak isospin, T_3, of the weak interaction and the weak hypercharge, Y, of the unified electroweak interaction, are related by

$$Q = T_3 + \frac{1}{2}Y. \tag{2.2}$$

Below the unification scale, however, the electroweak symmetry is spontaneously broken, because of the shape of the Higgs potential[5] [9, 10, 11]. The spontaneously broken symmetry gives rise to a particular realization of the observable gauge boson states. The W_1 and W_2 states mix to form the $W^\pm$ bosons, and at the same time acquire masses via the *Higgs mechanism* (spontaneous symmetry breaking). Similarly, the W_3 and B states mix to form the γ and Z bosons, where only the Z boson acquires a mass and the γ remains massless. We have then recovered the $W^\pm$ and Z bosons we know from the weak interaction, and the γ we know from the electromagnetic interaction, or *quantum electrodynamics* (QED) [12, 13, 14], as the full renormalizable gauge theory is called.

The weak interaction has the peculiar feature of being a vector minus axial-vector interaction, which in consequence only acts on left-handed particles (and right-handed anti-particles), and thereby maximally violates conservation of parity (P, effect on the physical laws by inversion of the spatial coordinates).

[4] The number 246 GeV corresponds to the measured vacuum expectation value, $v = (G_F\sqrt{2})^{-1/2}$, of the Higgs field, where G_F is the Fermi effective coupling constant.

[5] The so-called Mexican hat or wine bottle potential.

This is in stark contrast to QED and QCD, which both conserve parity. The combination of charge conjugation and parity transformation (CP), almost restore symmetry (because C-symmetry is also maximally violated), but not fully. The CP-symmetry is found to be slightly violated due to a complex phase that enters the CKM[6]-matrix [15, 16], which describes all possible flavor changing quark interactions. However, it turns out that by adding a time reversal (T, which is minimally violated) to the CP-transformation, the combined CPT-symmetry is conserved by the weak interaction, and hence by all interactions in the SM.

2.2 Shortcomings of the Standard Model

The SM has been immensely successful in explaining the particles we know of today and their interactions (with the exception of gravity). The vast number of particle experiments that have been conducted over the years have continued to confirm the predictions of the SM to ever greater precision. However, there are some aspects of nature where the SM comes up short.

2.2.1 Baryon asymmetry

As a consequence of our world being made up of baryonic matter, more matter than anti-matter must have been produced during the evolution of the universe. The process responsible for creating this imbalance between baryons and anti-baryons is called *baryogenesis*. Andrei Sakharov proposed in 1967 [17] that three conditions must be met in order for baryogenesis to take place: (i) baryon number conservation must be violated; (ii) charge(C)- and charge-parity(CP)-symmetry must be violated – that is, not only must there be an unequal rate of particle- and anti-particle production, but also an unequal rate of left-handed particle and right-handed anti-particle production, and vice versa; and (iii) interactions must be out of thermal equilibrium – that is, baryogenesis must happen at a lower rate compared to the expansion of the universe. The baryon and lepton numbers are accidental symmetries of the SM formulation, but this is only true at the classical level. It has been shown that quantum effects of the weak interaction can violate baryon number non-perturbatively [18]. However, experimental results have put strong limits on the size of baryon number violation within the SM, e.g., in terms of the life time of the proton [19].

2.2.2 Gravity

Whereas the SM has been successful in describing three out of the four fundamental forces (the electromagnetic, the weak and the strong interactions) as being mediated by gauge bosons (quanta of gauge fields or *force fields*), similar attempts to construct a corresponding quantized theory of gravity has so far not been able to obtain perturbative renormalizability, and as a consequence has

[6] Cabibbo, Kobayashi and Maskawa

not been able to provide physical predictions across energy scales. Experiments have not been able to guide the theoretical work much either, as gravity is so much weaker than the three other fundamental forces. The *Planck scale*, the energy scale at which the gravitational coupling is expected to be comparable to the other interactions, is around the order of 10^{19} GeV, or equivalently at a length scale of 10^{-35} m or a time scale of 10^{-43} s. This is way beyond what is accessible to particle colliders today, with the LHC operating at 1.3×10^4 GeV.

2.2.3 Accelerating expansion of the Universe

By comparing the magnitude-redshift of standard candles [20] (astronomical object with known luminosity) and/or the distance-redshift of standard rulers (astronomical objects with known size), among other methods, it has been established that the expansion rate of the universe is increasing with distance. The most common explanation for this phenomenon suggests that there is a cosmological constant or vacuum energy (negative pressure) that gives rise to a repulsive force, resulting in an accelerated expansion of space. Because of its unknown origin (and being a counterpart to the unknown *dark matter* in the universe), it has been coined *dark energy*, and makes up about 69% of the total energy-matter content of the universe [21].

2.2.4 Neutrino oscillations and masses

According to the Standard Model, all of the three neutrino flavors are massless, and experiments have to date excluded neutrino masses above 1.1 eV [19]. However, experiments have also shown that the neutrinos mix and undergo flavor oscillations as function of the ratio of distance traveled to their respective energies. The Super-Kamiokande experiment first confirmed oscillations of atmospheric neutrinos in 1998 [22], whereas the Sudbury Neutrino Observatory measured oscillations from solar neutrinos in 2001 [23]. Neutrino flavor oscillations can only occur if the neutrinos have non-zero masses, and the observable flavor (e, μ, τ) eigenstates are mixtures of the three distinct mass ('1', '2', '3') eigenstates. The neutrino sector of the Standard Model therefore needs to be extended in order to account for this phenomenon.

2.2.5 The hierarchy problem

Quantum loop corrections to the Higgs boson mass (the only scalar particle in the SM) are quadratically divergent in the momenta running in the loops. Since we need to account for all momenta where the SM is still valid, which to the best of our knowledge extends up to the Planck scale of 10^{18} GeV, we get corrections to the Higgs mass that are much larger than the Higgs mass itself. In order to obtain the measured Higgs mass of 125 GeV, there needs to be a fine-tuned cancellation of these corrections of size 10^{16} GeV, which without involving new physics is regarded as highly unnatural. This huge discrepancy between what

we expect and what we measure is what is called the *hierarchy problem* of the SM [24].

2.2.6 Dark matter

Many different experiments suggest that there must be a significant amount of matter in the universe that we so far have not been able to observe directly. All indications of this unseen matter have come through indirect, gravitational effects on astrophysical observations, including a wide range of phenomena and scales, such as: the rotation curves of galaxies in galaxy clusters [25], the velocity dispersion of stars inside individual galaxies [26], gravitational lensing [27], the power spectrum of the Cosmic Microwave Background (CMB) [21], the large-scale structure formation of the universe [28], and more. The fact that we have not been able to observe it directly yet, tells us something about its identity:

- it must be electrically neutral, since it does not interact with light or any other electromagnetic radiation;
- it cannot have color charge, since we otherwise would have detected it binding to normal atoms, resulting in anomalous heavy nuclei;
- it must have a non-zero mass in order to exert a gravitational pull on normal matter;
- it must be cold, i.e., move with non-relativistic speeds, in order to conform with the observed large-scale structure of the universe;
- it must be stable, meaning it cannot decay into lighter SM-particles (at least on the time scale of the age of the universe); and
- it can at most have a weak interaction to normal matter.

These features rule out all the particles in the SM as possible candidates for dark matter, including the neutrinos. The observations of the CMB by the Planck collaboration estimate that dark matter makes up 25.9% of the mass-energy content of the universe [21], which is more than five times the amount of normal, baryonic matter.

2.2.7 Gauge-coupling unification

Reminiscent of how James Clerk Maxwell found that the electric and magnetic forces could be unified in one combined, electromagetic force, and the work of Glashow, Weinberg and Salam which showed that the electromagnetic $\mathrm{U}(1)_Y$ and weak $\mathrm{SU}(2)_\mathrm{L}$ gauge symmetries could at some higher energy scale be unified in an electroweak $\mathrm{SU}(2)_\mathrm{L}\times\mathrm{U}(1)_Y$ gauge symmetry, several theoretical arguments have also been made in favor of an even larger symmetry, unifying the electroweak with the strong gauge interaction. Theories that exhibit such a unification of all the SM gauge couplings, at some higher energy scale, have become known as *Grand Unified Theories* (GUTs). The simplest example of a GUT is the

$\mathrm{SU}(5) \supset \mathrm{SU}(3)_c \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_Y$ gauge group, as proposed in 1974 by Georgi and Salam [29], which is supposed to be broken down to the SM gauge group at the so-called *grand unification scale* ($\sim 10^{16}$ GeV). In this model, the lepton and quark fields are combined into single, irreducible representations, which give rise to interactions that do not conserve the baryon and lepton numbers individually, but instead keep the combined $B - L$ quantum number a conserved quantity. One of the consequences is that the proton is allowed to decay, a process that has not yet been observed.

2.3 Introducing fermion-boson symmetry

So far, we have briefly touched upon some of the successes and shortcomings of the SM, in terms of the natural phenomena it is able to describe, and some of the cases where it fails to do so. A multitude of theories have been developed *beyond the SM* (BSM), in order to try and explain one or more of these missing pieces. Among the most studied BSM theories is *supersymmetry*, which adds a symmetry between fermions and bosons, and as a consequence is able to provide explanations to several of the outstanding questions in the SM. In the following section we will introduce some of the formalism behind this new symmetry [30, 31, 32].

Quantum field theory (QFT) generalizes quantum mechanics to the realm of special relativity (SR), where the relative speeds of particles are non-negligible compared to the speed of light. The four-dimensional spacetime of SR has a Minkowski metric, where a point on the Minkowski manifold is denoted by the coordinates $x^\mu = (t, x, y, z)$.

One of the postulates of SR is that the laws of physics are invariant under linear transformations between reference frames,

$$x^\mu \to x'^\mu = \Lambda^\mu{}_\nu x^\nu, \tag{2.3}$$

that consist of boosts and rotations. These are called Lorentz transformations and they make up the *Lorentz group* of transformations, L.

A special subgroup of L called the *proper orthocronous Lorentz group*, $L_+^\uparrow$, only includes proper boosts and rotations (no space or time reflections) and makes sure that time runs forward, therefore requiring that $\det\Lambda = 1$ and $\Lambda^0{}_0 \geq 1$.

By adding translation to the set of transformations in the Lorentz group,

$$x^\mu \to x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu, \tag{2.4}$$

we can obtain the *restricted Poincaré group*, $P_+^\uparrow$, which is a semidirect product of the proper orthoncronous Lorentz group and the translation group, $\mathrm{T}(1,3)$, where the latter only translates the 3-dimensional space coordinates, and leaves the time coordinate unaffected. The product (combination) of a 4-dimensional operator with a 3-dimensional operator, is expressed mathematically in terms of the semidirect product:

$$P_+^\uparrow = L_+^\uparrow \rtimes \mathrm{T}(1,3). \tag{2.5}$$

The Poincaré group captures all symmetries of the Minkowski spacetime.

It can be shown that $L_+^\uparrow$ is homomorphic (structure preserving) to the 2-dimensional special[7], unitary[8] groups $\mathrm{SU}(2) \times \mathrm{SU}(2)$ (from the structures of the boost and rotation generators),

$$L_+^\uparrow \cong \mathrm{SU}(2) \times \mathrm{SU}(2), \tag{2.6}$$

and furthermore to $\mathrm{SL}(2,\mathbb{C})$, the special group of linear transformations acting on 2-dimensional complex numbers. The latter homomorphism means that each element $\Lambda \in L_+^\uparrow$ can be assigned to two elements $M \in \mathrm{SL}(2,\mathbb{C})$. We can express this as two $\mathrm{SL}(2,\mathbb{C})$ groups with a $\mathbb{Z}_2$ graded Lie algebra or *superalgebra*[9]:

$$L_+^\uparrow \cong \mathrm{SL}(2,\mathbb{C})/\mathbb{Z}_2, \tag{2.7}$$

where $\mathbb{Z}_2$ is the set of integers modulo 2, i.e., $\{0,\ 1\}$. This means that we can choose to work with 2-dimensional (complex) Weyl spinors instead of the 4-dimensional Dirac spinors.

It can be shown that $\mathrm{SL}(2,\mathbb{C})$ has the following two inequivalent fundamental representations: a *self-representation*,

$$\rho(M) = M, \tag{2.8}$$

and a *complex conjugate self-representation*,

$$\rho(M) = M^*. \tag{2.9}$$

We can use this to define two sets of Weyl spinors:

- *Left-handed* (self-representation), which transform as:

$$\psi'_A = M_A{}^B \psi_B, \quad A, B = 1, 2. \tag{2.10}$$

- *Right-handed* (complex conjugate self-representation), which transform as[10]:

$$\bar{\psi}'_{\dot{A}} = (M^*)_{\dot{A}}{}^{\dot{B}} \psi_{\dot{B}}, \quad \dot{A}, \dot{B} = 1, 2. \tag{2.11}$$

[7] Special: determinant $= +1$

[8] Unitary: preserving probability amplitudes

[9] A $\mathbb{Z}_2$ graded Lie algebra, or superalgebra, is a vector space L which is a sum of two vector spaces L_0 and L_1, $L = L_0 \oplus L_1$, with a binary operation, $\bullet : L \times L \to L$, such that $\forall x_i \in L$:

i $x_i \bullet x_j \in L_{i+j \bmod 2}$ (grading)

ii $x_i \bullet x_j = -(-1)^{ij} x_j \bullet x_i$ (supersymmetrization)

iii $x_i \bullet (x_j \bullet x_k)(-1)^{ik} + x_j \bullet (x_k \bullet x_i)(-1)^{ji} + x_k \bullet (x_i \bullet x_j)(-1)^{kj}$ (generalized Jacobi identity)

(see Reference [30]).

[10] The dots above the indices $\dot{A}$ and $\dot{B}$ are merely added to mark that they are distinct from the indices A and B.

A Dirac fermion can then be represented by a left- and a right-handed Weyl-spinor as follows,

$$\psi_a = \begin{pmatrix} \psi_A \\ \bar{\chi}^{\dot{A}} \end{pmatrix}, \tag{2.12}$$

where $\psi_A^\dagger \neq \bar{\chi}^{\dot{A}}$ and $a = 1, 2, 3, 4$. Similarly a Majorana fermion can be represented as

$$\psi_a = \begin{pmatrix} \psi_A \\ \bar{\psi}^{\dot{A}} \end{pmatrix}, \tag{2.13}$$

where $\psi_A^\dagger = \bar{\psi}^{\dot{A}}$.

In 1967, Coleman and Mandula [33] showed that any extension of the Poincaré group (external symmetries) that includes the SM gauge group (internal symmetries) is isomorphic to a direct product of the groups, i.e., all elements of one group will commute with all elements of the other group. Hence, no unification of external and internal symmetries is possible. This has become known as *the no-go theorem*.

However, in 1975, Haag, Lopuszanski and Sohnius [34] showed that one can get around the no-go theorem by using a graded Lie algebra instead. An example of such a $\mathbb{Z}_2$-graded Lie algebra is the so-called *superalgebra*, which is of the form

$$L = L_0 \oplus L_1, \tag{2.14}$$

where $L_0 = P_+^\uparrow$ and L_1 is a new vector space spanned by four Majorana spinor charges (generators) Q_a, with $a = 1, 2, 3, 4$, or alternatively, Q_A, with $A = 1, 2$, and $Q_{\dot{A}}^\dagger$, with $\dot{A} = 1, 2$. This is called an $N = 1$ supersymmetry. One can also add additional sets of spinor charges, Q_a^α, with $\alpha = 1, \ldots, N$, to get what we call $N > 1$ supersymmetry, which results in more complicated group structures.

The *irreducible representations* or *superfields* of the superalgebra can be labeled by (m, j), with m representing mass and j representing a generalized form of spin[11]. For a given value of j, there are $2j + 1$ states with $j_3 = -j, -j + 1, \ldots, j - 1, j$, obtained by acting with the spinor charges Q_a on the states $|m, j, j_3\rangle$.

The Majorana spinor charges work as ladder operators that either raise or lower the spin of the states they are acting on with half a unit, so that

$$Q|\text{boson}\rangle = |\text{fermion}\rangle, \tag{2.15}$$

$$Q|\text{fermion}\rangle = |\text{boson}\rangle. \tag{2.16}$$

[11]The generalized spin operator $J_k = S_k + \frac{1}{8m}\bar{Q}\gamma_k\gamma^5 Q$, with $k \in \{1, 2, 3\}$, satisfies the angular momentum algebra, and includes the standard quantum mechanical spin operator S_k, plus an additional term involving a product of the spinor charges Q and $\bar{Q}$. One can show that eigenstates of J_k, with a corresponding eigenvalue of j_3 (and j), are also eigenvalues of S_k, with $s_k = j_k$ (and $s = j$). This means that the eigenstates of J_k are also eigenstates of the second term in its expression with separate eigenvalues. Therefore, one can, by using the spinor charges as ladder operators, construct multiple eigenstates of J_k, each with the same generalized spin values j and j_3, but with different values of s and s_3.

This means that for every superfield labeled by (m, j), there is an equal number of boson and fermion states, or put differently, there is an equal number of fermionic and bosonic degrees of freedom (d.o.f.).

It is convenient to collect these states in two types of superfields:

- *Chiral (or scalar) superfield* ($j = 0$)*:*
  - one Weyl spinor with $s = \frac{1}{2}$ and $s_3 = \pm\frac{1}{2}$ *(2 fermionic d.o.f.)*
  - two scalars with $s = 0$ *(2 bosonic d.o.f.)*
- *Gauge (or vector) superfield* ($j = \frac{1}{2}$)*:*
  - one (massive) vector boson with $s = 1$ and $s_3 = -1, 0, 1$ *(3 bosonic d.o.f.)*
  - two (hermitian conjugate) Weyl-spinors, each with $s = \frac{1}{2}$ and $s_3 = \pm\frac{1}{2}$ *(4 fermionic d.o.f.)*
  - one scalar with $s = 0$ *(1 bosonic d.o.f.)*

2.4 The Minimal Supersymmetric Standard Model

Now that we have a framework in place, that naturally produces fermion-boson symmetric states, the next step is to incorporate the SM particles into this framework, and to identify the corresponding *superpartners* that arise alongside these [35].

Using one set with four Majorana spinor charges ($N = 1$ supersymmetry), we can construct the *Minimal Supersymmetric Standard Model* (MSSM). This refers to the minimal extension of the SM that includes the symmetry between bosons and fermions. That is, each SM particle gets its supersymmetric partner with the same quantum numbers, apart from differing by half a unit of spin (if SUSY is unbroken, also the masses of the supersymmetric partners are expected to be the same as their SM counterparts).

The simplest way to construct the SM particles is by using the states of the chiral ($j = 0$) and the gauge ($j = \frac{1}{2}$) superfields. Since the Q_a operators that generate the different spin states of a superfield (given by specific values of (m, j)) are Majorana-spinors, the particle-anti-particle Weyl-spinors from the same superfield cannot make up the left-handed and right-handed hermitian conjugate Weyl-spinors of a Dirac fermion. A Dirac fermion needs a distinct particle and anti-particle pair. Therefore, one chiral superfield can, e.g., provide the Weyl-spinor of the left-handed up-quark, u_{L}, with $s = \frac{1}{2}$, along with its scalar superpartner the *left-handed* up-squark, $\tilde{u}_{\mathrm{L}}$, with $s = 0$ (being a scalar, it does not have helicity, but is labled by L to signify that it is the superpartner of the left-handed up-quark). The right-handed up-quark then needs to come from another complex conjugate superfield, with a Weyl-spinor, $u_{\mathrm{R}}^{\dagger}$, along with its superpartner the *right-handed* up-squark, $\tilde{u}_{\mathrm{R}}^{*}$.

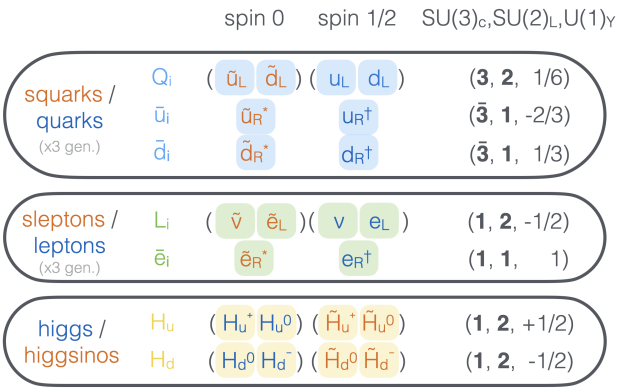


Figure 2.2: Chiral superfields in the MSSM. The table only explicitly shows the first generation quarks, squarks, leptons and sleptons, leaving out the two heavier generations (with the same structure) for brevity. The second column from the left denotes the $\mathrm{SU}(2)_\mathrm{L}$ doublets for left-handed particles and their superpartners in capital letters, and the corresponding singlets for the right-handed particles and their superpartners in lowercase letters. In the third and fourth columns, the superfields are arranged according to spin, where the SM particles are shown in blue text, whereas their superpartners are shown in orange. The final column denotes the dimensionality and gauge quantum numbers of the fields, where the first number tells whether the fields transform as color triplets or singlets under $\mathrm{SU}(3)_c$ transformations (the bar denotes triplets of anti-colors), the second number whether the fields transform as weak isospin doublets or singlets under $\mathrm{SU}(2)_\mathrm{L}$, and the last number tells the value of their weak hypercharge, Y, which is related to their electromagnetic charge and third component of weak isospin by Equation (2.2). Inspired by Table 1.1 in Reference [35].

Let us denote the *left-handed* superfield of the up-quark and -squark by $u = (u_\mathrm{L}, \tilde{u}_\mathrm{L})$, and the corresponding *right-handed* superfield by $\bar{u} = (u_\mathrm{R}^\dagger, \tilde{u}_\mathrm{R}^*)$. If we do the same for the down-quark and -squark, with $d = (d_\mathrm{L}, \tilde{d}_\mathrm{L})$ and $\bar{d} = (d_\mathrm{R}^\dagger, \tilde{d}_\mathrm{R}^*)$ for the *left-* and *right-handed* superfields, we can also collect the *left-handed* superfields in a weak isospin superfield $Q_1 = (u_\mathrm{L}, d_\mathrm{L})^T$, where the subscript denotes the quark/squark-generation. The rest of the fermions and sfermions can be constructed and grouped in the same manner, where the weak isospin doublets for the *left-handed* leptons are labeled $L_1 = (\nu_e, e_L)^T$, and similarly for the second and third generations.

In addition to SM particles and their superpartners, an extra Higgs doublet is needed in order to provide masses to both up- and down-type quarks and leptons in the weak isospin doublets. The fact that we need two Higgs (weak isospin) doublets[12], $H_u = (H_u^+, H_u^0)^T$ and $H_d = (H_d^0, H_d^-)^T$, results in five Higgs scalars

[12]The subscripts u and d signifies which Higgs doublet gives masses to the up- and down-type quarks and leptons.

Figure 2.3: Gauge superfields in the MSSM. The superfields of the SM gauge bosons are denoted by blue text, whereas their fermionic superpartners, the gauginos, are denoted by orange. The final column tells that the gluon and gluino transform as color octets under $\mathrm{SU}(3)_c$ and as a singlet under $\mathrm{SU}(2)_\mathrm{L}$, the W bosons and the winos transform as singlets under $\mathrm{SU}(3)_c$, but as triplets under $\mathrm{SU}(2)_\mathrm{L}$, whereas the B boson and the bino do not carry either color charge or weak isospin, and therefore transform as singlets under both $\mathrm{SU}(3)_c$ and $\mathrm{SU}(2)_\mathrm{L}$ transformations. All of these superfields have zero weak hypercharge, Y. Inspired by Table 1.2 in Reference [35].

in total: two neutral and CP-even[13] (h^0 and H^0), two charged ($H^\pm$) and one neutral and CP-odd[14] (A^0). The fermionic superpartners of the Higgs bosons are called *higgsinos*. As a consequence of the additional Higgs doublet, a new mass term, quadratic in the Higgs/higgsino doublet fields, arises in the superpotential of the MSSM. This term contains the Higgs/higgsino mass parameter, μ, which is the only new free parameter in the unbroken supersymmetric Lagrangian on top of the parameters of the SM.

The particles (fermions and *sfermions*[15], Higgs bosons and higgsinos) that are constructed from the components of chiral superfields, are listed in Figure 2.2.

The SM gauge bosons (spin-1) get fermionic superpartners called *gauginos* (spin-1/2), where, e.g., the superpartner of the gluon is called a gluino. The particles (gauge bosons and gauginos) in the MSSM, that are made from the components of gauge superfields, are given in Figure 2.3.

Figure 2.4 lists the gauge and mass eigenstates of the new particles added on top of the SM. These are the supersymmetric particles and the new Higgs boson states, where the latter arises from the fact that there are now two Higgs doublets, instead of the one present in the SM. The gauge eigenstates represent the pure flavor eigenstates of the gauge interactions. However, the observable particle states are in general mixtures of all gauge/flavor eigenstates that share the same conserved quantum numbers, and for squarks and sleptons,

[13] A CP-even scalar field (wave function), ϕ, remains identical under the combined transformation of charge conjugation ($\hat{C}$) and parity inversion ($\hat{P}$): $\hat{C}\hat{P}\phi = \phi$.

[14] A CP-odd scalar field (wave function), ϕ, changes sign under the combined transformation of charge conjugation ($\hat{C}$) and parity inversion ($\hat{P}$): $\hat{C}\hat{P}\phi = -\phi$.

[15] The supersymmetric fermions, which are superpartners of the SM bosons.

are of the same squark or slepton flavor. Below the weak scale, where the EW symmetry is broken down to the electromegnetic (EM) symmetry of QED, i.e., $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_Y \to \mathrm{U}(1)_\mathrm{EM}$, fields that share the same electromagnetic charge are allowed to mix, regardless of their T_3 and Y charges.

As a result, the (complex) field components of the two Higgs doublets mix among themselves, as long as they have the same electric charge, to form the five observable Higgs bosons mentioned above.

The mass states of the squarks and sleptons can be a mixture of their left- and right-handed components. In practice, the first and second generation squarks and sleptons can be considered the same as their gauge eigenstates, due to the negligible Yukawa couplings which make up the non-diagonal elements in the mass-mixing matrices. However, the sizable Yukawa couplings of the third generation squarks and sleptons cannot be ignored, which leads to two distinct mass eigenstates, labeled 1 and 2, where the mass state with label 1 is heavier than the mass state with label 2.

The observable mass states of the neutral, electroweak gauginos, the bino ($\tilde{B}^0$) and the wino ($\tilde{W}^0$), mix with the neutral higgsinos, $\tilde{H}_u^0$ and $\tilde{H}_d^0$, to form four *neutralino* states, $\tilde{\chi}_i^0$, with $i = 1, 2, 3, 4$, according to increasing mass. Similarly, the charged electroweak gauginos and the charged higgsinos mix to form the observable *chargino* mass states, $\tilde{\chi}_i^\pm$, with $i = 1, 2$. The neutralinos and charginos are sometimes collectively referred to as *electroweakinos*.

Finally, the gluino and the *gravitino*[16] do not have any other gauge eigenstates to mix with, so their observable mass eigenstates are the same as their gauge eigenstates.

In contrast with the SM, where all the renormalizable terms of the Lagrangian conserve the baryon (B) and lepton (L) numbers, the MSSM does contain renormalizable terms that violate these quantities. However, we know from experiments that B and L are conserved to high accuracy, which puts restrictions on how large the B- and L-violating terms in the MSSM can be. A common approach to deal with this tension, is to impose a $B - L$ symmetry on the terms in the Lagrangian, via a multiplicative quantum number called R-parity, which has to be conserved:

$$P_R = (-1)^{3(B-L)+2s}, \tag{2.17}$$

where the spin quantum number, s, has been added to obtain the attractive property that all SM particles have $P_R = +1$, while all sparticles have $P_R = -1$.

In consequence, R-parity conservation implies that the *lightest supersymmetric particle* (LSP) cannot decay further to SM particles, and must be stable. If the LSP is electrically neutral, it will at most have weak interaction, and can therefore be a good candidate for making up (maybe a part of) the dark matter. At colliders, sparticles would need to be produced in pairs, and decay to the LSP, which will escape detection, and therefore leave a significant amount of *missing energy* in the events.

[16] The superpartner of the hypothetical *graviton*, which mediates the gravitational interaction in quantum gravity.

	Spin	P_R	Gauge eigenstates	Mass eigenstates
higgs	0	+1	H_u^0 H_d^0 H_u^+ H_d^-	h^0 H^0 A^0 $H^\pm$
squarks	0	-1	$\tilde{u}_L$ $\tilde{u}_R$ $\tilde{d}_L$ $\tilde{d}_R$	(same)
	0	-1	$\tilde{s}_L$ $\tilde{s}_R$ $\tilde{c}_L$ $\tilde{c}_R$	(same)
	0	-1	$\tilde{t}_L$ $\tilde{t}_R$ $\tilde{b}_L$ $\tilde{b}_R$	$\tilde{t}_1$ $\tilde{t}_2$ $\tilde{b}_1$ $\tilde{b}_2$
sleptons	0	-1	$\tilde{e}_L$ $\tilde{e}_R$ $\tilde{\nu}_e$	(same)
	0	-1	$\tilde{\mu}_L$ $\tilde{\mu}_R$ $\tilde{\nu}_\mu$	(same)
	0	-1	$\tilde{\tau}_L$ $\tilde{\tau}_R$ $\tilde{\nu}_\tau$	$\tilde{\tau}_1$ $\tilde{\tau}_2$ $\tilde{\nu}_\tau$
neutralinos	1/2	-1	$\tilde{B}^0$ $\tilde{W}^0$ $\tilde{H}_u^0$ $\tilde{H}_d^0$	$\tilde{\chi}_1^0$ $\tilde{\chi}_2^0$ $\tilde{\chi}_3^0$ $\tilde{\chi}_4^0$
charginos	1/2	-1	$\tilde{W}^\pm$ $\tilde{H}_u^+$ $\tilde{H}_d^-$	$\tilde{\chi}_1^\pm$ $\tilde{\chi}_2^\pm$
gluino	1/2	-1	$\tilde{g}$	(same)
gravitino	3/2	-1	$\tilde{G}$	(same)

Figure 2.4: Gauge and mass eigenstates of the new particles added on top of the SM, along with their spin and R-parity quantum numbers. The mass eigenstates are mixtures (linear combinations) of their corresponding gauge eigenstates. Given the small Yukawa couplings of the first and second generation squarks and sleptons, we have not included mass mixing for these states in the table. The gluino and gravitino simply do not have other gauge eigenstates, with the same set of quantum numbers, to mix with. Inspired by Table 8.1 in Reference [35].

2.5 Supersymmetry breaking models

Since a superfield is labeled by a mass, m, the sparticles should have the same mass as their SM counterparts. If this were the case, we should have seen lots of sparticles by now. Therefore, supersymmetry must be broken at the energy scales currently accessible to us.

There exist several models for how supersymmetry can be spontaneously broken, which adds supersymmetry breaking terms to the Lagrangian. The SUSY breaking terms introduce more than 100 new, free parameters, which comprise masses and couplings for the sparticles.

In order to stay within the experimental constraints, e.g., to the extent CP is violated in nature, SUSY must be *softly* broken. In practice, this is done by adding *ad hoc* terms to the MSSM Lagriangian that produce the particle masses we need, but at the same time only break the supersymmetry in a minimal way. By applying a set of theoretically and experimentally well-motivated assumptions/constraints on the SUSY breaking terms in the Lagrangian, we can reduce the number of new, free parameters down to sets that are more tractable to explore. Examples of groups of models or mechanisms that can explain the soft SUSY breaking terms we have added explicitly, are *gauge-mediated SUSY*

breaking (GMSB) models (mediated by new, hidden sector gauge bosons entering loop corrections) and *Planck-scale-mediated SUSY breaking* (PMSB) models (mediated by gravity). We will only focus on the PMSB mechanism in what follows.

The idea is that the MSSM (visible sector) couples flavor blindly to some hidden sector, which contains a field, F, that develops a non-zero vacuum expectectation value (VEV), $\langle F \rangle \neq 0$, which in turn breaks supersymmetry in the hidden sector [35]. In this case, the coupling to the hidden sector is assumed to be mediated by gravitational strength interactions, including the effects of supergravity, which are assumed to become sizable at the Planck scale, $M_{\rm P} \sim 10^{18}\,\text{GeV}$. Then the soft mass terms in the visible sector should by dimensional analysis be roughly

$$m_{\rm soft} \sim \langle F \rangle / M_{\rm P}. \tag{2.18}$$

This is consistent with $m_{\rm soft} \to 0$, when either $\langle F \rangle \to 0$ or $M_{\rm P} \to \infty$, which corresponds to SUSY being unbroken or gravity becoming irrelevant, respectively.

For the soft masses to be of the order $m_{\rm soft} \sim 100\,\text{GeV}$, the mass scale associated with the breaking of SUSY in the hidden sector must be situated around $\sqrt{\langle F \rangle} \sim 10^{10}\,\text{GeV}$.

A popular simplification, motivated by gauge coupling unification, is to assume a *minimal* form for the normalization of kinetic terms and gauge interactions in the (non-renormalizable) Lagrangian at the Planck scale. By assuming, at the Planck scale, a common mass for the gauginos, $m_{1/2}$, a common mass for the scalars (sleptons and squarks), m_0, a common coefficient for the (trilinear) Yukawa couplings (two sleptons or two squarks coupling to a Higgs/higgsino field), A_0, and a common coefficient for the (bilinear, i.e., two-scalar particle, including Higgs/higgsino field) couplings, B_0, we have reduced the number of free parameters from the soft breaking terms down to four. In total, by also counting the Higgs/higgsino mass parameter from the superpotential, μ, of the unbroken MSSM, we are left with only five free parameters in total:

$$m_{1/2}, m_0, A_0, B_0, \mu. \tag{2.19}$$

A popular re-parametrization of B_0 and μ, is to instead express these in terms the ratio of the vacuum expectation values of the H_u^0 and H_d^0 fields, $\langle H_u^0 \rangle / \langle H_d^0 \rangle = v_u/v_d = \tan\beta$, and the sign of μ, giving the alternative parametrization:

$$m_{1/2}, m_0, A_0, \tan\beta, \text{sign}(\mu). \tag{2.20}$$

These parameters will then need to be evolved down to the energy scale in question by use of the renormalization group equation (RGE), to obtain predicted values at lower energy. One may in practice choose to start the RGE running from the grand unification (GUT) scale, at $m_{\rm GUT} \sim 10^{16}\,\text{GeV}$, rather than the Planck scale, at $m_{\rm P} \sim 10^{19}\,\text{GeV}$, because the unification of the gauge couplings (at the GUT scale) makes us more confident in what is happening between the weak and the GUT scale, than we are at what happens above the GUT scale.

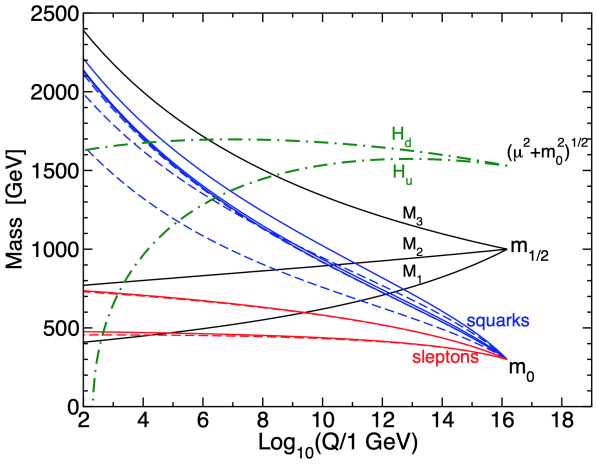


Figure 2.5: Example of how the masses of the sparticles can be realized at the weak scale (left end of the plot), from particular choices of a common scalar mass, m_0, a common gaugino mass, $m_{1/2}$, and the Higgs/higgsino mass parameter μ from the superpotential, at the GUT (or Planck) scale (right end of the plot). From Figure 8.4 in Reference [35].

This kind of reduction has been referred to as *minimal SUperGRAvity* (mSUGRA) or the *Constrained Minimal Supersymmetric Standard Model* (CMSSM), and has become a benchmark model for both theoretical and experimental studies.

Figure 2.5 shows a sketch of how the mass-hierarchy of the MSSM may look like at the weak scale, after the masses have been evolved using the RGE from the mSUGRA/CMSSM assumptions at the GUT scale, $m_{\text{GUT}} = 1.5 \times 10^{16}$ GeV. This model suggests that gaugino masses of the bino ($\tilde{B}$ with mass M_1) and the wino ($\tilde{W}$ with mass M_2), with $M_1 < M_2$, can both be realized with masses below 1 TeV. The sleptons are also predicted to be among the lightest sparticles, and could be found in between the $\tilde{B}$ and $\tilde{W}$ masses. The gluino mass, M_3, and the masses of the squarks are in this model believed to be quite a bit heavier than the electroweakinos and the sleptons.

The mass-hierarchy at the weak scale is, however, heavily dependent on the initial choices of m_0, $m_{1/2}$ (setting the initial condition for the RGE running of M_1 and M_2) and μ. When $\tilde{\chi}_1^0$ is taken to be the LSP, the gaugino-higgsino compositions and mass-splittings of the neutralinos, $\tilde{\chi}_i^0$, with $i = 1, 2, 3, 4$, and charginos, $\tilde{\chi}_j^\pm$, with $j = 1, 2$, will often be considered w.r.t. three different scenarios:

- $M_1 < M_2, |\mu|$**:** $\tilde{\chi}_1^0$ is bino-like, while $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ are both wino-like and about twice as heavy as the LSP. The experimental signature is typically not expected to be kinematically dominated by the mass-splitting between the LSP and the NLSP (non-compressed mass spectrum).

- $|\mu| < M_1, M_2$**:** $\tilde{\chi}_1^0$, $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ are all higgsino-like, and have comparable masses. The experimental signature is expected to be dominated by the mass-splitting between the LSP and the NLSP, where the NLSP decays to the LSP via soft (low-p_T) SM particles (compressed mass spectrum).

- $M_2 \ll M_1, |\mu|$**:** $\tilde{\chi}_1^0$ and $\tilde{\chi}_1^\pm$ are all wino-like, and are expected to have very similar masses. The high degree of mass-degeneracy may cause the NLSP to not decay promptly, i.e., giving it a sizable life-time, and produce a displaced vertex (highly compressed mass spectrum).

2.6 Simplified models

Even though the five parameters of mSUGRA/CMSSM are a lot more manageable than the full MSSM, the 5-dimensional phase space still poses a challenging task to explore and interpret. Therefore, so-called simplified models [36], where we reduce the number of free parameters even further, and typically down to two or three, have become popular at the LHC.

The idea is to cover as many event topologies as possible, by looking at a large number of specific and well-defined processes and final states. The simplified models are often inspired by more complete models, such as mSUGRA/CMSSM, but assume a simplified mass-hierarchy, where only a few particles will be able to contribute. We can typically summarize the simplified model with a single Feynman diagram, where the only free parameters are the mass of the pair-produced sparticles and the mass of the LSP, and the decay to the LSP happens with 100% branching fraction via the decay chain specified in the diagram. Figure 2.6 shows a typical example of a diagram illustrating a simplified model.

This simplicity comes at the cost of providing a very limited phenomenology in a single model. The strength of the simplified models comes from considering multiple models together. One can for example consider additional decay chains of the same pair-produced sparticles by combining them, and scale the branching fractions according to the interpretation one is interested in.

Another possible use of the simplified models is to include the search results as constraints in global likelihood fits of parameters in more complete BSM models. This has been done by the GAMBIT collaboration [37] in a global fit to the electroweakino sector of the MSSM, where multiple results from both ATLAS and CMS in terms of simplified models were used as inputs to the fit.

2.7 Supersymmetric solutions

Along with the attractive feature that supersymmetry supplies a fundamental space-time symmetry between fermions and bosons, several other important features come along as consequences of this additional symmetry. In the following section, I will outline three of these consequences, which happen to provide solutions to major outstanding issues with the SM.

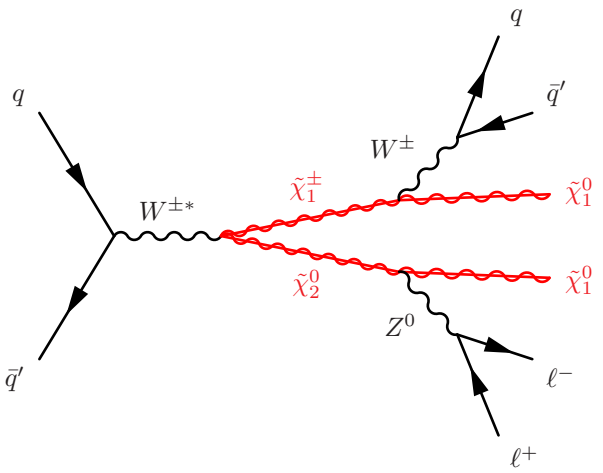


Figure 2.6: An example of a diagram representing a simplified model.

2.7.1 Natural solution to the hierarchy problem

The Higgs field potential, $V(H)$, has the form [30, 35]

$$V(H) = \mu^2|H|^2 + \lambda|H|^4, \tag{2.21}$$

where λ is some constant, and μ^2 can be recognized as the negative of the squared Higgs mass, $\mu^2 = -m_H^2$ (as the factor in front of the squared Higgs field term). In order for spontaneous symmetry breaking to occur, we must have $\lambda > 0$ and $\mu^2 < 0$, so that the Higgs field can develop a non-zero vacuum expectation value (VEV), $\langle H \rangle \equiv v$. The mass of the Higgs boson (an excitation of the Higgs field) is determined by the VEV of the Higgs field, as

$$m_H = \sqrt{2\lambda}v. \tag{2.22}$$

Via its relations to other electroweak parameters, such as the mass of the W boson and the weak isospin coupling, g, the VEV has been measured to be $v = 246\,\text{GeV}$ [19].

The SM fermions and the electroweak gauge bosons obtain their masses from their coupling strength to the Higgs field, and more specifically, to the VEV of the Higgs field. For a Dirac fermion, f, with a Yukawa coupling, λ_f, to the Higgs field, the Lagrangian term is given as $-\lambda_f H\bar{f}f$, and the mass of the fermion as

$$m_f = \frac{\lambda_f}{\sqrt{2}}v. \tag{2.23}$$

If we instead solve for the Yukawa coupling,

$$\lambda_f = \sqrt{2}\frac{m_f}{v}, \tag{2.24}$$

we find that the heavier the particle, the stronger it couples to the Higgs field.

When we calculate the interactions between quantum fields to higher orders in perturbation theory (including loops), we get additional terms that contribute to the masses of the particles, i.e., terms that are proportional to the fields squared. If we include the contribution from one fermion loop to the Higgs mass squared, we get an additional correction term of the form:

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{\mathrm{UV}}^2 + \ldots, \tag{2.25}$$

where Λ_{UV} is a regularizing term that goes to zero for momenta above some high energy (ultra-violet, UV) cut-off scale, where the SM is no longer valid, or will be modified by new physics.

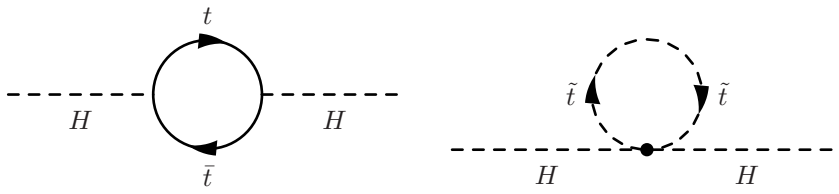
The only reference we currently have for such a new physics scale, is the Planck scale at $\sim 10^{18}\,\mathrm{GeV}$, where we expect gravity to become of similar strength as the SM interactions, and a new high energy theory of gravity will modify the physics of the SM. But this quadratic sensitivity of the Higgs mass to the cut-off scale, then leads to corrections of order 10^{36} to the mass we have measured to be $m_H^2 = (125\,\mathrm{GeV})^2$. Therefore, there needs to be a dramatic cancellation from other terms, of some 30 orders of magnitude, in order to produce the measured Higgs mass. Clearly, there needs to be some missing piece(s) in the SM that could provide an explanation for this discrepancy, but without a *natural* mechanism of the theory to provide such a cancellation, some miraculously fine-tuned counter-terms would need to be added at each order of perturbation theory to get the right mass. This is by many regarded as *unnatural*, and has become known as the hierarchy problem.

If we in addition to the fermions, would have some new scalar boson, s, the corresponding correction to the squared Higgs mass from the scalar would be of the form:

$$\Delta m_H^2 = \frac{\lambda_s}{16\pi^2}\Lambda_{\mathrm{UV}}^2 + \ldots, \tag{2.26}$$

where λ_s is the coupling strength of the scalar field, s, to the Higgs field. By comparing the corrections from fermions and scalars to the Higgs mass, we can notice that the two contributions have opposite signs, and if $2\lambda_s = |\lambda_f|^2$, the two leading terms would cancel exactly.

This is in fact precisely what happens in supersymmetry when the SM fermions and their scalar superpartners have the same masses, i.e., coupling with equal strength to the Higgs field. And there are twice as many scalars as fermions, because the fermions have two internal spin-degrees of freedom, thereby giving the same number of fermion and boson degrees of freedom in total. Since the coupling strengths to the Higgs field is proportional to the mass of the particles in question, the heaviest SM particles, with their corresponding superpartners, are expected to give the largest corrections to the Higgs mass. Figure 2.7 shows diagrams of the top quark fermion loop correction to the Higgs mass in (a), and the corresponding cancellation from the stop squark scalar loop correction in (b). The fact that supersymmetry provides this *natural* cancellation



(a) Fermion loop from the SM (b) Scalar loop from SUSY

Figure 2.7: Feynman diagrams of one-loop corrections, from (a) SM Dirac fermions and (b) SUSY scalar bosons, to the vacuum energy of the Higgs field, and consequently to the (squared) mass of the Higgs boson.

of the enormous corrections to the Higgs mass, and thereby providing a solution to the hierarchy problem, is one of the main reasons why supersymmetry has become so popular as a possible extension to the SM.

However, since we have not observed any SUSY particles so far, supersymmetry must be broken and the masses of the sparticles need to be heavier than their SM partners. Therefore, the cancellations between the fermion and boson loops will not be exact. If sparticles are found with masses up to around the TeV scale, the additional corrections needed to get the measured Higgs mass may still be regarded as acceptably natural.

Given that the ATLAS and CMS experiments have not yet seen any signs of sparticles in the LHC Run 2, many scientists have started to lose faith in SUSY being realized in nature, mostly because it does not seem to be as natural a solution to the hierarchy problem as many had hoped for. This has been referred to as the *little hierarchy problem*.

2.7.2 Dark Matter candidates

One of the main open questions in physics today, is that of the nature of dark matter (DM). Among the many new particles that arise in supersymemtry, it turns out that a few of these seem to satisfy the characteristics of DM [30].

In the early, hot universe, we believe that both SM and dark matter particles were in thermal equilibrium. This means that the energy and matter density (thereby also the temperature) of the universe was high enough for SM and DM particles to interact frequently and to pair-produce DM from SM particles, and vice versa, with equal rates, SM SM $\leftrightarrow$ DM DM. This was the case when the temperature of the universe, T, was much larger than the mass of the DM particles, $T \gg m_{\mathrm{DM}}$.

As the universe expanded and cooled, the temperature did at some point become too low for the average energy of the particle interactions to produce the

mass of the DM particles, $T < m_{\mathrm{DM}}$. When this happened, the DM production effectively stopped, $\mathrm{SM\,SM} \nrightarrow \mathrm{DM\,DM}$, but the DM would continue to annihilate into the lighter SM particles, $\mathrm{SM\,SM} \leftarrow \mathrm{DM\,DM}$.

Eventually, when the universe had diluted and cooled even more, the expansion rate of the universe exceeded the annihilation rate of the DM particles, so that the final remaining DM interactions effectively stopped as well, $\mathrm{SM\,SM} \nleftarrow \mathrm{DM\,DM}$. The consequence of this was that the comoving number density[17], or the *relic density*, was frozen to a constant value. This is called *freeze-out*.

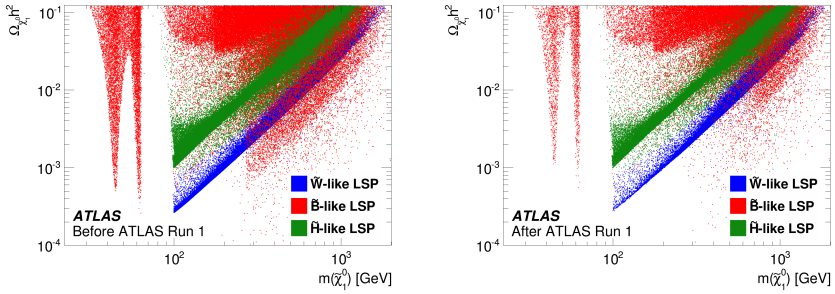
The relic DM density is dependent on the annihilation cross-section of the DM particles. The higher the annihilation cross-section, the more of the DM particles would have annihilated before the density was frozen. It turns out that a DM particle of weak scale cross-section and mass would yield almost exactly the measured relic density in the universe. Particles that fall into this category are called *weakly interacting massive particles* (WIMPs), and the fact that such particles should be in reach of current or planned experiments, has been called the *WIMP miracle*. WIMPs show up in multiple BSM theories, such as SUSY and Kaluza-Klein theory with an extra compactified dimension.

In supersymmetry, if R-parity is conserved, the lightest supersymmetric particle (LSP) will be absolutely stable, and cannot decay to other particles. In the cases where the lightest neutralino is the LSP, it is a prototypical example of a WIMP, and has been one of the most popular and well-studied DM candidates for decades. There are also other sparticles that could be candidates for DM, such as the gravitino, the sneutrino and the axino [38]. However, I will here only focus on the lightest neutralino, as it is taken to be the LSP in the simplified models presented later in the thesis. More specifically, it is taken to be a bino LSP, with the lightest chargino and the next-to-lightest neutralino as mass-degenerate wino NLSPs.

After the LHC Run 1, ATLAS performed a re-interpretation of all the exclusion limits obtained from the Run 1 data on simplified SUSY models, in terms of a scan over the 19-dimensional pMSSM [39]. 500 million samples were randomly picked from flat distributions of the 19 parameter dimensions. The ranges of the parameters were chosen to conform with experimental constraints and to values reachable at $\sqrt{s} = 8\,\mathrm{TeV}$. In addition, a pre-selection of the sampled points were applied to keep within constraints from precision measurements on electroweak and flavor physics, the measured DM relic density, and from collider experiments.

Figure 2.8 shows the distribution of the parameter points that passed the pre-selection, on the left, and the points that in addition were not excluded by any ATLAS Run 1 analyses, on the right. The distributions are shown in the plane of the lightest neutralino mass on the x-axis and the corresponding DM relic density on the y-axis. The points are also colored according to the dominant gaugino or higgsino component. The top of the plots correspond to relic density values

[17] The number density in a relatively small volume which moves along with the expanding universe.



(a) Before the ATLAS Run 1.

(b) After the ATLAS Run 1.

Figure 2.8: Phase space points sampled from the 19-dimensional pMSSM, highlighting the dominant component of the LSP [39].

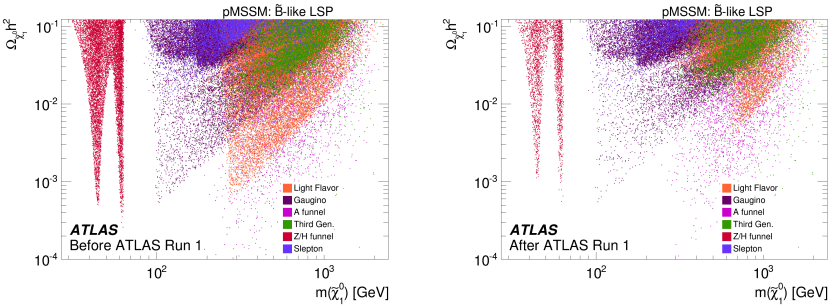
consistent with the measured DM relic density of $\Omega_{\Lambda\mathrm{CDM}}h^2 = 0.1188 \pm 0.0010$, from the Planck Collaboration. I.e., the points at the very top are able to fulfill the complete DM component, while the points lower on the y-axis will only be able to account for a part of the total DM contribution.

It is worth to note that only bino-like LSPs can satisfy the relic density constraints below an LSP mass of 100 GeV, while wino- and higgsino-like LSPs are more likely to explain the full DM density at around 1 TeV. However, after the Run 1 limits have been applied, a significant fraction of the points with a bino-like LSP below masses of 200 GeV are excluded.

In general, bino-like LSPs tend to over-produce DM, in the sense that the annihilation cross-section is too low to get below the relic density constraints. However, there are a few mechanisms that can raise the annihilation cross-section in certain parts of the parameter space, and which give rise to the distinct features shown in these plots. In Figure 2.9, only the points with a bino-like LSP are shown, and they are colored according to which mechanism that allows the points to evade the relic density constraints. Again, after pre-selection on the left, and after the Run 1 limits have been applied on the right.

The points below an LSP mass of 100 GeV are attributed to the so-called Z- and H-funnels. This happens when the mass of the bino-like LSP is close enough to either the Z- or H-mass, so that we could get resonant annihilation of the LSP along with the corresponding boson. The same could happen together with the heavier CP-odd Higgs boson, A^0. The remaining mechanism is co-annihilation of an LSP together with some other sparticle, if the two sparticles have very similar masses.

Even though a large fraction of the pMSSM points with a bino-like LSP were excluded by the ATLAS Run 1 limits, it is by no means excluded altogether as a DM candidate. Compared to the wino- and higgsino-like LSP, the bino-like LSP scenario still exhibits the richest, non-excluded phenomenology out of the three.



(a) Before the ATLAS Run 1. (b) After the ATLAS Run 1.

Figure 2.9: Phase space points sampled from the 19-dimensional pMSSM, highlighting the mechanism that lets the points evade the DM relic density constraints [39].

2.7.3 Gauge coupling unification

The gauge couplings of the SM interactions are often referred to as:

$$g' \text{ for } \mathrm{U}(1)_Y, \quad g \text{ for } \mathrm{SU}(2)_\mathrm{L}, \quad g_s \text{ for } \mathrm{SU}(3)_c, \tag{2.27}$$

where the electroweak gauge couplings are related to the electric charge of QED by

$$e = g \sin\theta_W = g' \cos\theta_W, \tag{2.28}$$

with θ_W being the Weinberg angle. Often, we instead refer to the vertex factors of the corresponding interactions as:

$$\alpha_1 = \frac{g'^2}{4\pi}, \quad \alpha_2 = \frac{g^2}{4\pi}, \quad \alpha_3 = \frac{g_s^2}{4\pi}. \tag{2.29}$$

The renormalization group (RG) evolution of the inverse of these couplings, turn out to run linearly with the RG scale at one-loop order [30, 35]:

$$\frac{d}{dt}\alpha_a^{-1} = -\frac{b_a}{4\pi}, \qquad a = 1, 2, 3, \tag{2.30}$$

where $t = \ln(Q/Q_0)$, with Q the RG scale, and the canonical normalization of $g_1 = \sqrt{5/3}g'$, $g_2 = g$ and $g_3 = g_s$ for gauge coupling unification is used. The renormalization coefficients for the SM and the MSSM at one-loop order are given in Table 2.1.

The results of running the gauge couplings of SM and MSSM up to higher energy scales at two-loop order, are shown in Figure 2.10. If we only include the SM particles in the loop-corrections, pairs of gauge couplings will meet at different energy scales, separated by orders of magnitude, as indicated by the

Table 2.1: The renormalization coefficients for the SM and the MSSM, at one-loop order [30, 35].

	b_1	b_2	b_3
SM	$41/10$	$-19/6$	-7
MSSM	$33/5$	1	-3

dashed lines. However, if we include the sparticles of the MSSM in the loops as well, all the three gauge couplings will meet at the same energy scale. The approximate point of coincidence is dependent of the mass scale at which the sparticles appear and by varying α_s within uncertainties, which is illustrated by the red and blue lines in the plot.

The calculations visualized in Figure 2.10 will not make the three gauge couplings meet at exactly the same point, but the deviations are small enough that they could be consistent within uncertainties of threshold effects by whatever particles may exist near the unification scale. The definition of the unification scale, M_{GUT}, is often taken to be the point where the U(1) and SU(2) couplings meet, i.e., where $\alpha_1(M_{\mathrm{GUT}}) = \alpha_2(M_{\mathrm{GUT}})$. This happens at $M_{\mathrm{GUT}} \sim 1.5 \times 10^{16}\,\mathrm{GeV}$.

This unification of the gauge couplings may well be a coincidence, but in light of the successful electroweak unification, this is taken by many to be a strong hint of a grand unified theory (GUT) including supersymmetry, e.g., superstring theory.

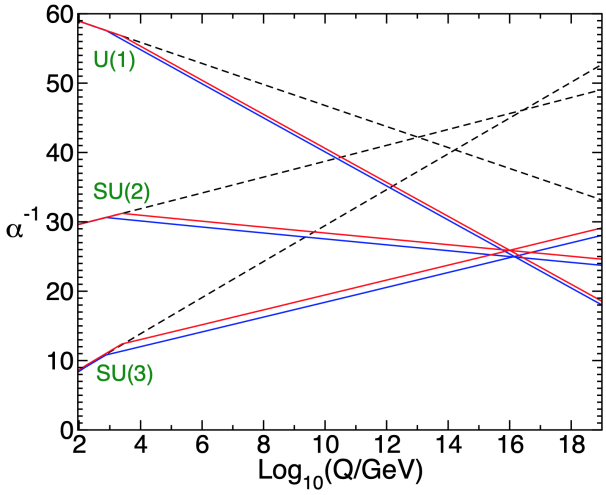


Figure 2.10: Renormalization group running of the couplings of the weak hypercharge (U(1)), the weak isospin (SU(2)) and the strong color charge (SU(3)) interactions, above the weak scale. The dashed lines show the pure SM evolution, without any contributions from SUSY particles, whereas the red and blue lines show two possible evolutions if sparticles enter the game. The blue line representing one possible realization of the sparticles, where the lightest sparticle mass enters at a lower energy (mass) scale compared to the particular realization represented by the red line. In both cases where SUSY is present, the SM couplings unify at approximately the same energy scale, which is not the case for the pure SM evolution. From Figure 6.8 in Reference [35].

Chapter 3

Proton collisions at the LHC

The *Large Hadron Collider* (LHC) is the largest and most advanced particle accelerator ever made, with its 27 km long ring of superconducting magnets, placed in a tunnel about 100 meters below the ground, on the boarder between Switzerland and France, near Geneva. It is designed to reach the unprecedented collision energy of 14 TeV in proton-proton (pp) collisions. This is about seven times more than the Tevatron ($p\bar{p}$) at Fermilab, near Chicago in the United States, which used to be the most powerful particle collider. By using protons, which are composite particles consisting of quarks and gluons, a wide range of quark and/or gluon interaction energies can be explored. This fact makes it a discovery machine for new physics phenomena at energy scales up to the pp collision energy. That makes it the ideal tool in our search for supersymmetry.

3.1 The Large Hadron Collider

The LHC [40, 41] operates with two particle beams, traveling in opposite directions, inside seperate beam pipes of ultrahigh vacuum. Because the protons are extremely tiny, about $1\,\text{fm} = 10^{-15}\,\text{m}$ in diameter, they are collected in bunches of about 10^{11} protons per bunch, in order to increase the chance for inelastic proton-proton scattering to take place in each *bunch crossing*. The proton bunches are accelerated up to nearly the speed of light, with 25 ns bunch-spacing, which corresponds to a collision rate of 40 MHz.

The collider ring consists of 1232 dipole magnets, which produce magnetic fields of 8.3 T that bend the particle trajectories, and 392 quadrupole magnets that focus the particle beams. These are superconducting electromagnets that are cooled down by liquid helium to a temperature of 1.9 K, which is colder than the temperature of outer space, at 2.7 K. A number of radiofrequency cavities are placed along the ring to accelerate the particle beams a little more each time they pass. For the Run 2 of LHC, each beam was accelerated up to 6.5 TeV, giving a *center-of-mass collision energy* of 13 TeV. The beams are made to collide at four different *interaction points* along the ring, at which the four main LHC experiments ATLAS, CMS, ALICE and LHCb are located. Special triplets of quadrupole insertion magnets are placed close to the experiments to squeeze the transverse beam areas as much as possible, from 0.2 mm to 16 µm across, for the interaction points, again to maximize the chances of inelastic scattering events.

Before the protons enter the LHC ring, they have passed through multiple stages of acceleration, as illustrated in Figure 3.1. The journey starts in a small red bottle of hydrogen, from which hydrogen atoms are extracted and the electrons separated from the proton nuclei by passing through an electric

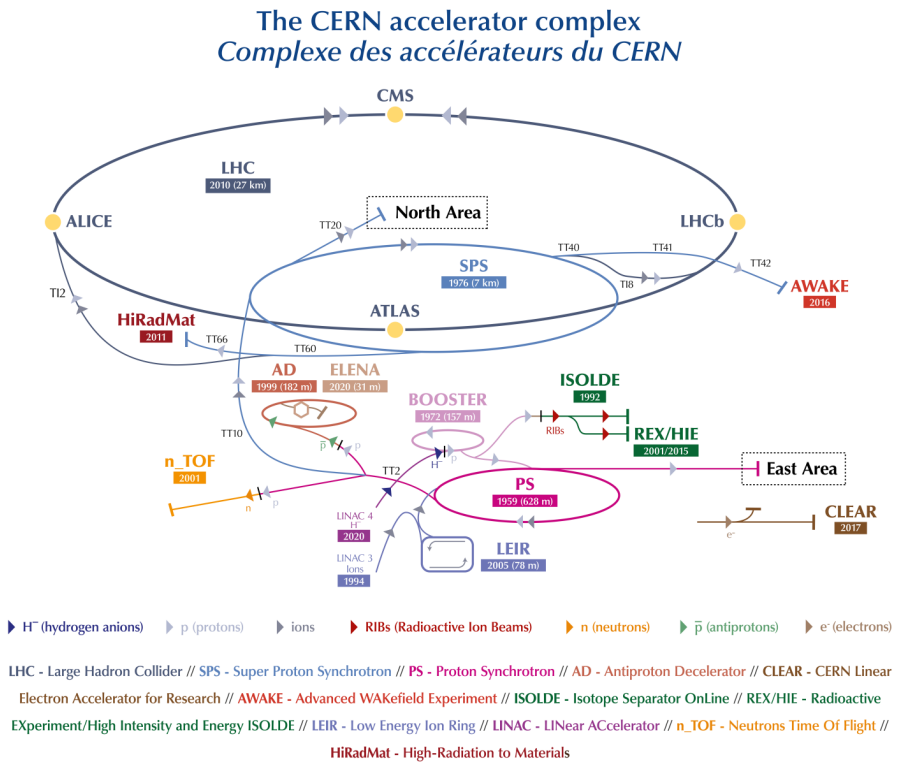


Figure 3.1: A schematic of the accelerators and associated experiments at CERN [42].

field, before being injected into LINAC2 (which in 2020 has been replaced by LINAC4, to be used for the LHC Run 3). This is a linear accelerator that uses radiofrequency cavities to accelerate the protons up to an energy of 50 MeV. The protons then enter the Proton Synchroton Booster (BOOSTER), which is made up of four superimposed synchrotron rings, and takes the protons up to 1.4 GeV. The next stage of acceleration happens in the Proton Synchrotron (PS), which sends the protons on to the Super Proton Synchrotron (SPS) with an energy of 25 GeV. The SPS is the second largest accelerator at CERN, measuring nearly seven km in circumference, and was the accelerator ring that was used to discover the W and Z bosons. It accelerates the protons up to 450 GeV, which is the energy of the protons when injected into the LHC.

In the following sections, we will discuss some of the most important parameters of particle collider experiments, namely the *particle types* that are accelerated and collided, the *center-of-mass energy* at which they collide, the *luminosity* of the particle beams, and the *cross-section*, i.e., a measure of the

probability, for the particle interactions to occur.

3.1.1 Particle type

At e^+e^- colliders, which collide the elementary electron and positron particles, we have full control of the momenta of the initial state (the colliding/interacting) particles. At a *pp* collider, however, which collide protons built up of quarks and gluons, we have full control of the momenta of the composite protons, but not the momenta of the quarks, anti-quarks and/or gluons that actually make up the initial state of the particle events we measure in the detectors. Both of these types of colliders have their pros and cons.

If we want to do precision measurements on a specific type of particle, e.g., study the detailed properties of the Higgs boson, we would be better off with an e^+e^- collider[1] tuned to operate at the energy corresponding to the Higgs mass, because of the simple and clean electroweak interactions, which are nearly free of background processes. However, we are then stuck with a fixed collision energy, which is not particularly useful if we want to search for new physics at an unknown energy scale.

In addition, the bending magnets that steer the electrons in a circular trajectory cause them to emit synchrotron radiation. The energy loss per time (power, P) due to synchrotron radiation is proportional to the Lorentz factor, γ, of the particle in question raised to the fourth power, $P \propto \gamma^4$. The Lorentz factor can be expressed in terms of the energy, E, and the rest mass, m_0, of the particle, as $\gamma = E/(m_0c^2)$. At a given energy, E, an electron (or positron) will lose 10^{13} times more energy per time to synchrotron radiation compared to a proton, due to the difference in their rest masses. Therefore, a *pp* collider can achieve higher collision energies compared to an e^+e^- collider.

If we then instead consider a *pp* collider, like the LHC, the quarks, anti-quarks and gluons inside the protons can take on almost any fraction of the proton momenta, and thereby explore a wide range of energies. This makes it a good choice for new physics searches, but comes at the cost of allowing a lot more to happen in each collision event, including chaotic QCD interactions and a huge increase in background contributions to the final states we are looking for. Therefore, accurate modeling, and finding ways to reduce the background to the event and particle selections, are among the main challenges in basically any analysis of data from the LHC.

3.1.2 Center-of-mass energy

The *center-of-mass energy* $\sqrt{s}$ of two (or more) particles, such as two colliding protons at the LHC, can be found from their energies and momenta via the

[1] Like the Large Electron Positron collider (LEP), the predecessor of the LHC.

Lorentz invariant quantity

$$s = \left(\sum_i E_i\right)^2 - \left(\sum_i \mathbf{p}_i\right)^2, \tag{3.1}$$

which has units of energy squared [43]. By taking the square root of s, we have a measure of the combined center-of-mass energy of the collision, $\sqrt{s}$, typically given in units of *electron volts*, eV.

The center-of-mass energy of the colliding protons during the LHC Run 2, was $\sqrt{s} = 1.3 \times 10^{13}\,\text{eV} = 13\,\text{TeV}$.

3.1.3 Luminosity

The *instantaneous luminosity*, $\mathcal{L}$, of two proton beams which collide head-on, with Gaussian profiles, is given by

$$\mathcal{L} = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y}, \tag{3.2}$$

where $f = 40\,\text{MHz}$ is the bunch crossing frequency at the LHC, n_1 and n_2 are the numbers of protons in the colliding bunches, and σ_x and σ_y are the root-mean-square (rms) horizontal and vertical beam sizes, respectively [43]. This quantity is a measure of the expected number of interactions per time per area, or *incident flux* of particles, $\text{s}^{-1}\text{cm}^{-2}$. It can be increased either by raising the collision frequency, packing more protons into the bunches or squeezing the beam sizes more.

The *integrated luminosity*, L, is the integral of the instantaneaous luminosities over a period of time, i.e., the sum of the instantaneous luminosity of each bunch crossing, at given points in time, $\mathcal{L}(t)$, over a period of data-taking,

$$L = \int \mathcal{L}(t)\,dt. \tag{3.3}$$

It is used as a measure of the "total amount of potential interactions" we have exposed the protons for in the experiments. The integrated luminosity has units of inverse area, and are often given in terms of the inverse unit of *barns*, where $1\,\text{barn} \equiv 10^{-28}\,\text{m}^2$. Typically the order of magnitude we are working with are picobarn (pb), where $1\,\text{pb} = 10^{-12}\,\text{barn}$, or femtobarn (fm), where $1\,\text{fb} = 10^{-15}\,\text{barn}$.

3.1.4 Cross-section

The so-called *interaction cross-section*, σ, is a quantum mechanical measure of the probability of a given particle interaction to occur. For a proper, formal definition, see any introductory book on quantum field theory, e.g., Reference [43]. Qualitatively, it can be expressed as

$$\sigma = \frac{\text{number of interactions per unit time per target particle } [\text{s}^{-1}]}{\text{incident flux } [\text{s}^{-1}\text{cm}^{-2}]} \sim [\text{cm}^2], \tag{3.4}$$

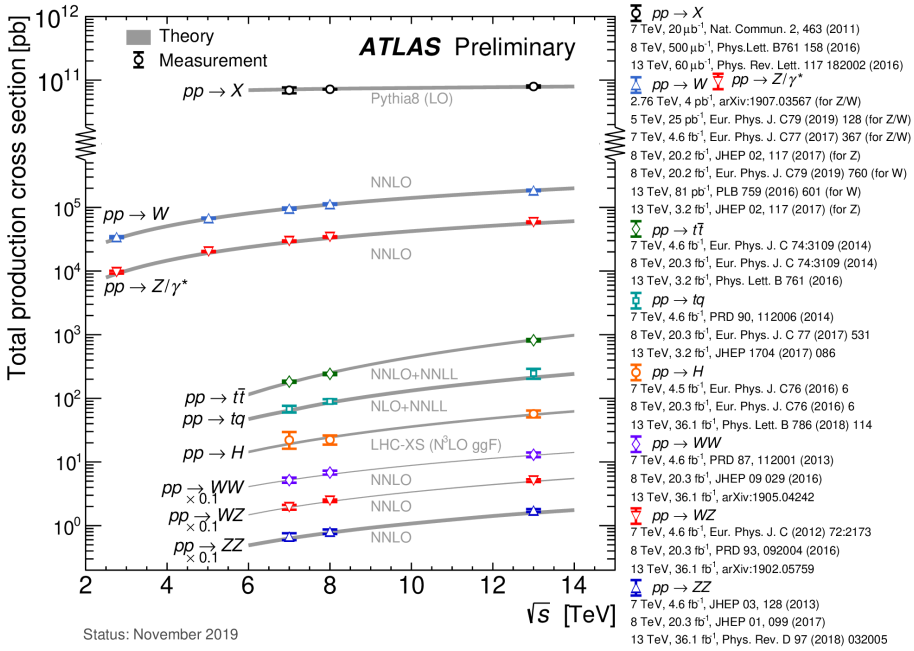


Figure 3.2: Cross-sections measured by the ATLAS detector for various SM pp interactions at the LHC, as a function of the center-of-mass energy of the pp collisions [44].

and therefore has units of area [43]. As for luminosity, this area is typically given in terms of barns, and the typical scale is pb or fb.

The cross-sections for different SM processes to take place in pp collisions, measured at $\sqrt{s} = 7$, 8 and 13 TeV at the LHC, are shown (as colored points) in Figure 3.2, and are compared to the theoretical predictions (in gray lines). The cross-section values span several orders of magnitude, from diboson production of order 1 pb, all the way up to single boson production (+jets) of order 10^5 pb. By comparison, the inclusive (total) cross-section for pp collisions at the LHC, is of order 10^{11} pb.

Figure 3.3 shows the expected cross-sections for sparticle production in pp collisions at $\sqrt{s} = 13$ TeV at the LHC. The sparticle cross-sections fall exponentially with increasing sparticle masses, where the strongly interacting sparticles, the squarks and gluinos, range from about $10^1 - 10^3$ pb at 250 GeV, down to about $10^{-2} - 1$ pb at 1 TeV; while the weakly interacting sparticles, such as the neutralinos and charginos, range from about 1 pb at 250 GeV, down to 10^{-3} pb at 1 TeV.

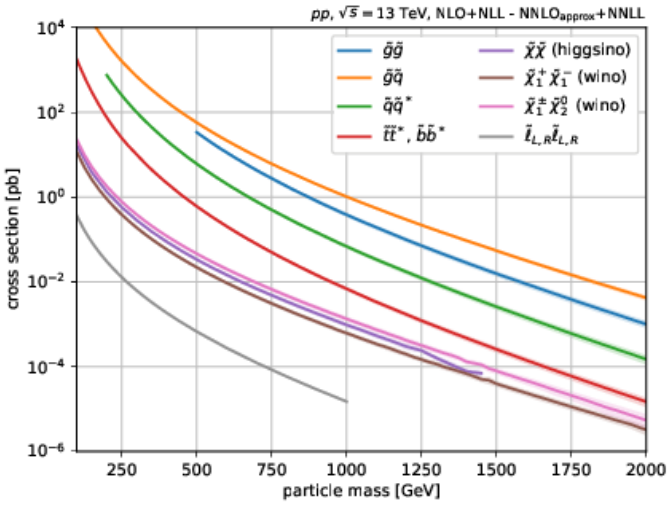


Figure 3.3: Predicted cross-sections for various SUSY production processes at the LHC, as a function of the sparticle mass [45].

3.1.5 Expected number of events

When we know the integrated luminosity, L, of a period of data-taking, and the cross-section, σ_X, for a certain process X to occur, we can calculate the expected number of *events*, N_X, for this particular process during the given period of data-taking, by

$$N_X = L \times \sigma_X. \tag{3.5}$$

3.2 Kinematics at the LHC

At the LHC, we know the full 4-momenta of the colliding protons fairly accurately, but it is typically only one parton from each proton that takes part in the inelastic scattering events we are interested in. Moreover, the individual partons may take on nearly any fraction of the longitudinal momentum of the proton, as given by the parton distribution function for the parton flavor in question, at a given proton energy/momentum scale. Therefore, we do not know the longitudinal momentum carried by the initial state partons. However, since the protons travel parallel to the beam line (the z-axis), the vector sum of the initial state partons should, because of momentum conservation, be equal to zero in the plane transverse to the beam line. This means that the vector sum of the final state particles also needs to be zero in the transverse plane. In consequence, we typically work with the transverse components of momenta, energies and masses for capturing the full event kinematics in the transverse plane.

We can broadly categorize the kinematic variables in two groups: (i) *low-level variables*, which are measurements or reconstructed quantities of a single particle or object, and (ii) *high-level variables*, which are calculated from the the low-level variables of multiple particles or objects.

3.2.1 Low-level variables

The ATLAS detector encapsulates the collision point, which is taken to be the origin of the coordinate systems, e.g., $(x, y, z)_{\text{collision point}} = (0, 0, 0)$ in Cartesian coordinates. Since the incoming protons collide at (or very close to) the defined origin of the ATLAS detector, the outgoing particles from the collisions radiate out from the origin in a spherically symmetric fashion. Therefore, the geometry and kinematics of the events can be described naturally in the 3-dimensional spherical coordinate system, (r, θ, ϕ), where r is the radius, θ the polar angle and ϕ the azimuthal angle.

In practice, we use a slightly modified version of the spherical coordinate system, inspired by the physical interpretations of the measurements. The polar angle, θ, is translated to the pseudo-rapidity, η, by the following relation, $\eta = -\ln\tan(\theta/2)$. The resulting 3-dimensional position vector can then be expressed as

$$\mathbf{x}_{\text{spherical}} = (r, \eta, \phi), \tag{3.6}$$

which can be translated to Cartesian coordinates by

$$\begin{aligned} \mathbf{x}_{\text{Cartesian}} &= (x, y, z) && (3.7)\\ &= (r\cos\phi, r\sin\phi, r\sinh\eta). && (3.8) \end{aligned}$$

If we have a measurement of the full energy deposit of a particle or an object (e.g., a jet) from a calorimeter measurement, or have inferred the identity of a particle with a known mass, we can construct the full 4-momentum of the particle/object in terms of the 4-vector

$$p^{\mu}_{\text{spherical}} = (E, p_{\text{T}}, \eta, \phi), \tag{3.9}$$

which can be translated to Cartesian coordinates by

$$\begin{aligned} p^{\mu}_{\text{Cartesian}} &= (E, p_x, p_y, p_z) && (3.10)\\ &= (E, p_{\text{T}}\cos\phi, p_{\text{T}}\sin\phi, p_{\text{T}}\sinh\eta), && (3.11) \end{aligned}$$

where $E = m$ (in natural units[2]), with m being the mass of the particle/object.

3.2.2 High-level variables

From the low-level variables of single particles/objects, we can construct higher-level quantities by combining the measurements from multiple particles/objects.

[2]In natural units, we set $c = 1$, giving the energy-momentum relation $E = mc^2 \to E = m$, and the time-component of the 4-momentum vector, $E/c \to E$.

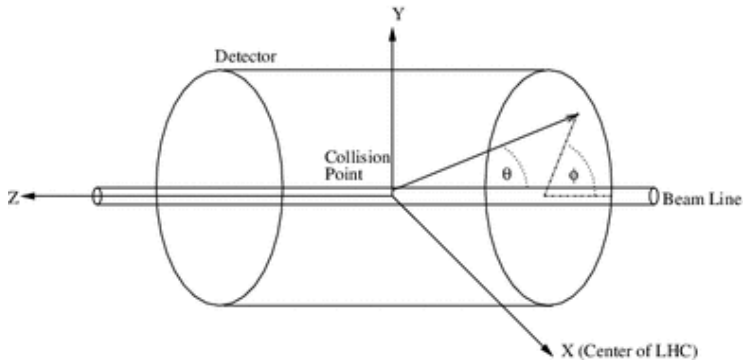


Figure 3.4: A sketch of the Cartesian (x, y, z) and the spherical (r, θ, ϕ) coordinate systems, where the radial coordinate r is given by $r = \sqrt{x^2 + y^2}$. These two coordinate systems are used to describe the geometry of the ATLAS detector, and the characteristics of the particles involved in the collisions taking place inside of the detector volume. From Reference [46].

This can be simple relations, such as the the angular separation between a pair of particles/objects, or more involved relations, such as the "stransverse" mass of a system of particles.

Below is a list of the high-level variables that will be considered in this thesis.

3.2.2.1 Momentum/energy scales

Variables that quantify the total momentum or energy of all, or part of, the particles of an event, can be useful for characterizing the energy scale at which an event, or part of an event, is taking place at. The energy scale typically correlates strongly with the topology of the event (angular distributions), and with the amount and composition of the expected contributions to the background.

Scalar summed p_{T} The H_{T}-variable is calculated as the scalar sum of the p_{T} of all the hadronic jets in the event:

$$H_{\mathrm{T}} = \sum_{i \in \{\mathrm{jets}\}} ||\mathbf{p}_{\mathrm{T},i}||, \tag{3.12}$$

and gives a measure of the hadronic energy scale of the event.

Vector summed p_{T} The absolute value of the vector summed p_{T} of two particles/objects, 1 and 2, which are decay products of a vector boson, V, are calculated as

$$p_{\mathrm{T},12} = ||\mathbf{p}_{\mathrm{T},1} + \mathbf{p}_{\mathrm{T},2}|| \tag{3.13}$$

$$= p_{\mathrm{T},V}. \tag{3.14}$$

Missing transverse momentum/energy By summing the transverse momentum vectors of all the reconstructed objects (the first term in the equation below) and the so-called soft terms from low-p_{T} tracks that have not been associated with a reconstructed object (the second term in the equation), we can negate this sum to get an estimate of the missing transverse (2-vector) momentum that is needed for momentum conservation in the transverse plane,

$$\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}} = -\left(\sum_{i \in \{\mathrm{obj.}\}} \mathbf{p}_{\mathrm{T},i} + \sum_{j \in \{\mathrm{soft}\}} \mathbf{p}_{\mathrm{T},j} \right). \tag{3.15}$$

The absolute value of this 2-vector is what we call the missing transverse energy,

$$E_{\mathrm{T}}^{\mathrm{miss}} = ||\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}||. \tag{3.16}$$

Relative $E_{\mathrm{T}}^{\mathrm{miss}}$ In order to deal with sizeable mis-measurements of jets or leptons, which lead to so-called fake $E_{\mathrm{T}}^{\mathrm{miss}}$ (becomes important in chapter 7 and 8), we can construct a new variable which only counts the component of $\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}$ that is perpendicular to the ϕ-direction of the nearest lepton or jet, in events where the $\Delta\phi_{\ell,j}$ is less than 90 degrees:

$$E_{\mathrm{T}}^{\mathrm{miss,rel}} = \begin{cases} E_{\mathrm{T}}^{\mathrm{miss}} & \text{if } \Delta\phi_{\ell,j} \geq \pi/2 \\ E_{\mathrm{T}}^{\mathrm{miss}} \times \sin\Delta\phi_{\ell,j} & \text{if } \Delta\phi_{\ell,j} < \pi/2 \end{cases}, \tag{3.17}$$

where $\Delta\phi_{\ell,j} = \min_{i \in \{\ell,j\}} \Delta\phi(\mathbf{p}_{\mathrm{T},i}, \mathbf{p}_{\mathrm{T}}^{\mathrm{miss}})$, and the minimization is performed over the set of all reconstructed leptons and jets in the event, $\{\ell, j\}$.

Mis-measurements of visible objects lead to an artificial momentum imbalance in the event, which further cause artificial or fake $E_{\mathrm{T}}^{\mathrm{miss}}$. Significant visible momenta traveling in the same direction as the $E_{\mathrm{T}}^{\mathrm{miss}}$, may indicate such a mismodeling. By disregarding any $E_{\mathrm{T}}^{\mathrm{miss}}$ component parallel (in the same direction) to the nearest lepton or jet, which is more likely to occur for background processes, such as Z + jets, we can effectively obtain better separation between the distribution of background and signal events, and thereby increase our sensitivity to the signal.

Object-based $E_{\mathrm{T}}^{\mathrm{miss}}$ significance Another way to deal with artificial or fake $E_{\mathrm{T}}^{\mathrm{miss}}$, is to weigh the value of the $E_{\mathrm{T}}^{\mathrm{miss}}$ by the precision of its reconstruction. The $E_{\mathrm{T}}^{\mathrm{miss}}$ significance can, in a simplified manner, be expressed as

$$E_{\mathrm{T}}^{\mathrm{miss,sig}} = \frac{E_{\mathrm{T}}^{\mathrm{miss}}}{\sigma(E_{\mathrm{T}}^{\mathrm{miss}})}, \tag{3.18}$$

where $\sigma(E_{\mathrm{T}}^{\mathrm{miss}})$ is the uncertainty in the reconstruction of the $E_{\mathrm{T}}^{\mathrm{miss}}$, which considers the individual uncertainties of the objects that enter into the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation. See Section 8.3 for a more detailed description.

3.2.2.2 Masses

By calculating the combined mass of multiple particles or objects, we can statistically target particle combinations which are likely to be decay products of a given mother particle with a certain mass. We can use this information to either keep events, if the mother particle is part of the signal process, or throw away events, if the mother particle is part of a background process, or is otherwise not of interest for the study at hand.

Invariant mass The invariant mass of two particles/objects, labeled 1 and 2, can be calculated in the following way:

$$m_{12}^2 = (E_1 + E_2)^2 - ||\mathbf{p}_1 + \mathbf{p}_2||^2 \tag{3.19}$$

$$= m_1^2 + m_2^2 + 2(E_1 E_2 - \mathbf{p}_1 \cdot \mathbf{p}_2). \tag{3.20}$$

If these two particles are the only decay products of a mother particle of mass m_{mother}, then the invariant mass of the daughter particles should be equal to the mass of the mother particle, $m_{12} = m_{\text{mother}}(\pm\Gamma/2)$, where m_{mother} is the pole mass and Γ is the width of the resonance in the invariant mass spectrum.

Transverse mass The transverse equivalent of the invariant mass of a lepton, ℓ, and $E_{\text{T}}^{\text{miss}}$, stemming from a decay of a heavy particle into a visible and an invisible particle (e.g. a neutrino), is called the transverse mass,

$$m_{\text{T}}(\mathbf{p}_{\text{T}}^{\ell}, \mathbf{p}_{\text{T}}^{\text{miss}}) = \sqrt{2(p_{\text{T}}^{\ell} p_{\text{T}}^{\text{miss}} - \mathbf{p}_{\text{T}}^{\ell} \cdot \mathbf{p}_{\text{T}}^{\text{miss}})}, \tag{3.21}$$

where the invisible particle is taken to be massless, which is a good approximation for the neutrinos. The transverse mass distribution forms a kinematic endpoint at the mass of the mother particle. The events reaching the endpoint are characterized by fully taking place in the transverse plane, i.e., the longitudinal components of the momenta of the lepton and the invisible particle are equal to zero.

Stransverse mass In R-parity-conserving SUSY models, the sparticles are always produced in pairs, and if both of these sparticles (e.g. $\tilde{\chi}_1^{\pm}\ \tilde{\chi}_1^{\mp}$) decay to leptons, we have (at least) two leptons and (at least) two particles escaping detection (e.g. the $\tilde{\chi}_1^0$ LSPs). In that case, there is a supersymmetric version of the transverse mass, which calculates a transverse mass for each of the two leptons by distributing the total $\mathbf{p}_{\text{T}}^{\text{miss}}$ among the two systems, and minimizing the maximum of the two transverse masses by varying the distribution of the $\mathbf{p}_{\text{T}}^{\text{miss}}$-vector in terms of the size of $\mathbf{q}_{\text{T}}$:

$$m_{\text{T2}} = \min_{\mathbf{q}_{\text{T}}} \left[\max \left(m_{\text{T}}(\mathbf{p}_{\text{T}}^{\ell_1}, \mathbf{q}_{\text{T}}), m_{\text{T}}(\mathbf{p}_{\text{T}}^{\ell_2}, \mathbf{p}_{\text{T}}^{\text{miss}} - \mathbf{q}_{\text{T}}) \right) \right], \tag{3.22}$$

where the transverse masses are given by

$$m_{\text{T}}(\mathbf{p}_{\text{T}}, \mathbf{q}_{\text{T}}) = \sqrt{2(p_{\text{T}} q_{\text{T}} - \mathbf{p}_{\text{T}} \cdot \mathbf{q}_{\text{T}})}. \tag{3.23}$$

As SM processes also can produce a similar kinematic endpoint, such as the diboson process WW, where each of the bosons decay leptonically as $W \to l + \nu_l$, and produce an endpoint at the W mass, we can use m_{T2} to efficiently reduce WW background by throwing away events with $m_{\mathrm{T2}} < m_W$.

3.2.2.3 Angular separation

The combined angular separation between the direction of two objects in both the η- and ϕ-angles, are calculated by

$$\Delta R_{12} = \sqrt{(\Delta\eta_{12})^2 + (\Delta\phi_{12})^2} \tag{3.24}$$

$$= \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2}. \tag{3.25}$$

This variable can also be used to define a cone of radius ΔR around an object, and be used for isolation requirements, e.g., by restricting the presence of tracks from other object within this cone.

3.2.2.4 Ratios

We will in some cases also be interested in the balance between different high- and/or low-level variables in the event. E.g., the amount of $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event compared to the p_{T} of one or (vector sum of) multiple visible object(s), X:

$$E_{\mathrm{T}}^{\mathrm{miss}}/p_{\mathrm{T}}(X). \tag{3.26}$$

Chapter 4

13 TeV pp data collected by the ATLAS detector

4.1 The ATLAS detector

The ATLAS experiment is located at interaction point 1 along the LHC ring, right next to the main CERN site in Meyrin. The ATLAS detector [47] is designed to encapsulate the interaction point as much as possible, to reach almost 4π of angular coverage. It has a cylindrical geometry, consisting of multiple layers of sub-detector systems, which target measurements of the momentum or energy of different types of particles. Figure 4.1 shows a sketch of the ATLAS detector and its sub-detector systems.

A simplified, transverse cross-section of the ATLAS detector components is sketched in Figure 4.2, along with how the different particle types interact with each layer. The trajectories of the particles are represented by solid lines inside the detector components they interact with (are visible to), and by dashed lines in the detector components they do not interact with (are invisible to). Charged particles have curved trajectories, because of the Lorentz force exerted on them

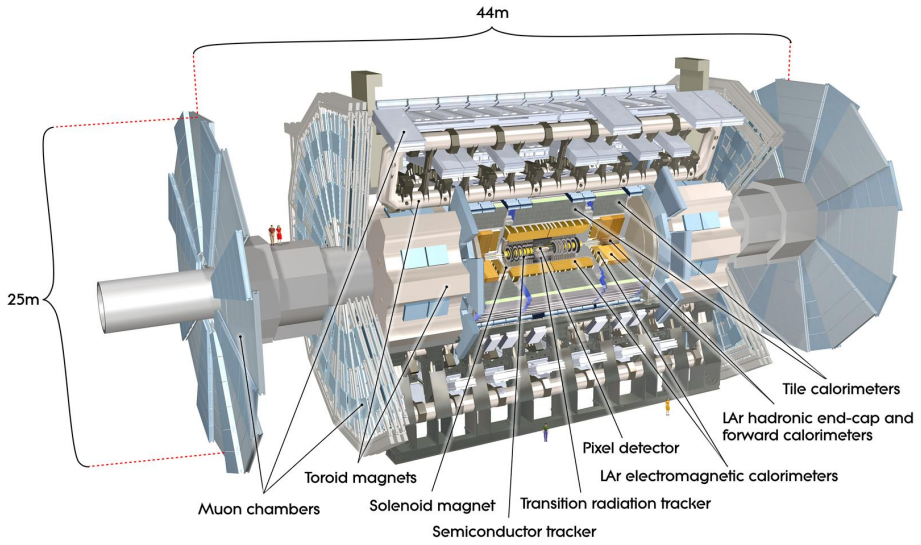


Figure 4.1: A schematic of the ATLAS detector [47].

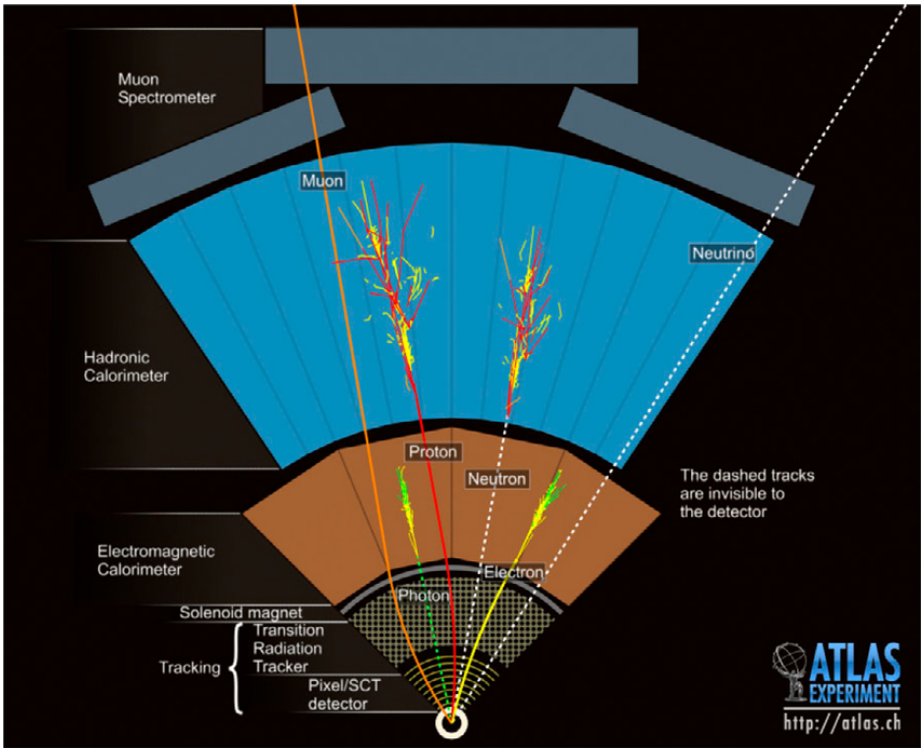


Figure 4.2: An illustration showing how different types of particles interact with the different layers of the ATLAS detector [47]. Note that the relative sizes of the detector components, and distances between them, are not representative of the real detector, but are modified for the purpose of displaying the particle interactions.

by the magnetic field. The sign of the electric charge (positive or negative) of a particle determines its direction of deflection.

The *inner detector trackers* only interact with electrically charged particles, like the electron, muon and proton, which means that the electrically neutral photon, neutron and neutrino do not leave hits in the tracking detectors. The momentum and electric charge of the electrically charged particles can be determined from the curvature of their reconstructed tracks.

The *electromagnetic (EM) calorimeter* interacts with all particles that feel the EM interaction, and stops the photon and the electron entirely through EM showers of particles, thereby absorbs and measures their total energy. The proton and the muon interact slightly with the EM calorimeter, but do only lose a small portion of their energy there.

The *hadronic calorimeter* is designed to interact strongly with hadrons, such as the proton and the neutron, to completely absorb and stop these particles, by producing hadronic showers in the calorimeter, and thereby measures the total

energy of the incident particles.

The *muon spectrometer* makes up the outermost layers of the ATLAS detector. Similarly to the inner detector, it aims to measure hits from the electrically charged muons that pass through the muon tracking layers. It does, however, not aim to stop the muons, but rather determine their momenta from the curvature of their reconstructed tracks. Therefore, the only detectable particles which are expected to escape from the ATLAS detector, are the muons.

There are a few other elementary particles we have not yet mentioned. The tau lepton, the EW gauge bosons and the Higgs boson are heavy enough that they decay to lighter particles before they even reach the innermost pixel layer. The final, unmentioned elementary particles, with large enough life time to traverse the detector volume, are the neutrinos. They only interact via the weak interaction, and are therefore not detectable by any of the ATLAS sub-systems. The energy/momentum carried away by the undetected neutrinos leads to *missing energy*. However, it is not only the neutrinos that can produce missing energy, also exotic particles that only interact weakly, could be responsible for part of the missing energy in collision events. This fact is heavily exploited in searches for dark matter, supersymmetric particles and other BSM searches.

4.1.1 The inner detector

The detector layers closest to the interaction point are referred to as the *inner detector* (ID) or *inner trackers*. Their role is to measure the momenta of the electrically charged particles. This is done by exposing the charged particles to a 2 T magnetic field, produced by a solenoid surrounding the ID, which bends the particle trajectories in the plane transverse to the beam axis. By reconstructing the particle tracks from the hits they leave in the detector layers, the momentum and the sign of the electric charge can be determined, from the straightness of the track and the direction of deflection, respectively.

The ID consists of three sub-systems: (i) the *pixel detector*, (ii) the *semiconductor tracker* (SCT) and (iii) the *transition radiation tracker* (TRT). A sketch of the inner detector sub-systems can be found in Figure 4.3.

The pixel detector is made up of 80 million silicon pixel sensors, each of size $50 \times 400\,\mu\text{m}^2$, with a resolution of $14 \times 115\,\mu\text{m}^2$, and spread over multiple layers, both in the barrel and the end-cap regions. The innermost layer is called the *insertable b-layer* (IBL), and was inserted before the start of Run 2 in order to allow for a better determination of secondary vertices from b-hadron decays.

Situated outside of the pixel layers, are the SCTs, which consist of silicon microstrip trackers, also placed on multiple layers in the barrel and disks in the end-caps. Readout strips every 80 µm on the silicon, enables a spatial resolution of 17 µm, in the direction transverse to the strips. The pixel and SCT (precision tracking) detectors cover the pseudo-rapidity range $|\eta| < 2.5$.

Furthest away from the interaction point, we find the TRTs. They consist of 4 mm in diameter straw tubes, 144 cm long, with a 0.03 mm in diameter gold-plated tungsten wire in the center. They provide spatial measurements with a precision of 0.17 mm. The TRTs extend out to $|\eta| = 2.0$.

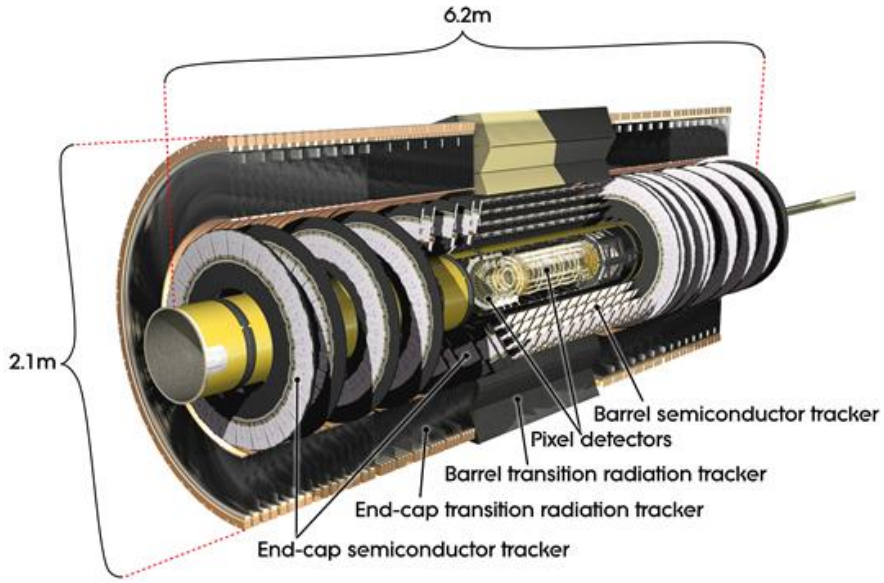


Figure 4.3: A schematic of the ATLAS inner trackers [47].

4.1.2 The calorimeters

ATLAS has, broadly speaking, two types of calorimeters: (i) *electromagnetic calorimeter* (ECAL) and (ii) *hadronic calorimeter* (HCAL) systems, which are both designed to fully stop certain types of incident particles. This is done by installing high-density, *passive* or *absorbing* detector plates that cause lots of interactions with the incoming particles, which then lose their energy by emitting showers of new particles. So-called *active* detector volumes are placed in between the passive material, in regular intervals, in order to measure all the energy emitted by the shower particles, and thereby measure the total amount of energy of the incoming particle(s) that hit the detector in the first place. A sketch of the ATLAS calorimeters is shown in Figure 4.4.

The ECAL is immediately surrounding the inner detector, both in the barrel ($|\eta| < 1.475$) and the end-cap regions ($1.375 < |\eta| < 3.2$). It consists of *absorbing* lead plates, with liquid Argon (LAr) volumes in between, working as the active part of the calorimeter. The ECAL is split in a barrel and two end-cap systems (one on each end of the cylinder), all three being lead-LAr detectors. The thickness of the calorimeter is made large enough to fully measure the showers of incident photons and electrons/positrons. The muons naturally have longer interaction lengths (average distance between interactions) with lead, and therefore only lose a small fraction of their energy in the ECAL.

The HCAL is immediately surrounding the ECAL on all sides. It consists of two different types of detectors. In the barrel ($|\eta| < 1.0$) and extended barrel

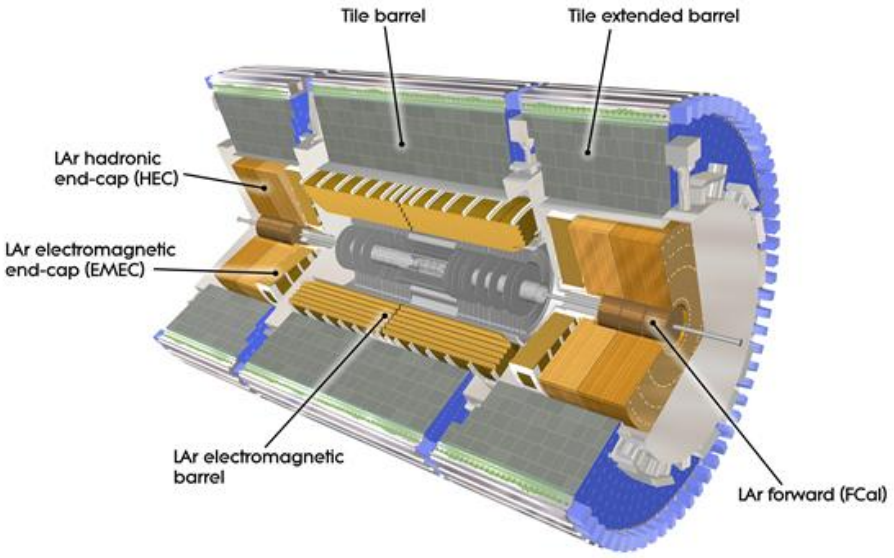


Figure 4.4: ATLAS electromagnetic and hadronic calorimeters
A schematic of the electromagnetic and hadronic calorimeters of the ATLAS detector [47].

regions ($0.8 < |\eta| < 1.7$), the HCAL is made of absorbing steal plates, and plastic scintillator tiles are used as the active material. In the end-cap regions ($1.5 < |\eta| < 3.2$), however, there are hadronic LAr detectors, with absorbing copper plates used as the active material; whereas in the very forward regions ($3.1 < |\eta| < 4.9$), a combination of copper and tungsten plates are used instead. The active materials are chosen to maximize the interaction cross-section with hadrons, such as neutrons, protons and pions. The depth of the HCALs are also designed to fully stop the particle showers which arise from the incident hadrons, in order to measure their total energies and avoid *punch-through* to the muon system. Hadrons are efficiently stopped by the HCAL, which in principle means that the muons and the weakly interacting neutrinos are the only particles that leave the HCAL.

4.1.3 The muon spectrometer

The outermost layers of the ATLAS detector are dedicated to measurements of muon momenta. Similar to the ID, the *muon system* or *muon spectrometer* (MS) is also a tracking detector, consisting of multiple layers of detector material, immersed in a strong magnetic field in order to bend the trajectories of the electrically charged muons.

In contrast to the ID, which uses a solenoid magnet, the MS has eight

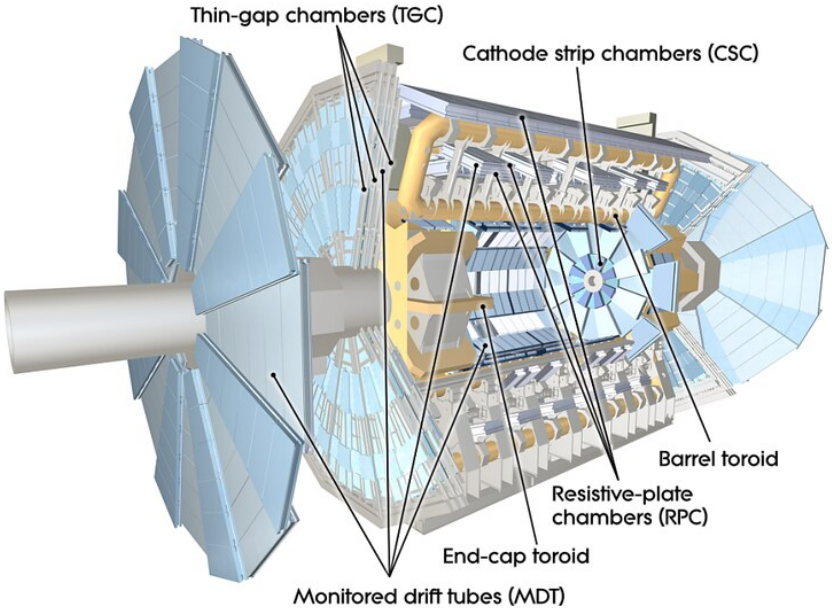


Figure 4.5: A schematic of the muon spectrometer of the ATLAS detector [47].

large air-core toroid magnets in the barrel region (which provide up to 2.5 T of magnetic field strength), and two smaller toroid systems in the end-cap regions, each with eight smaller toroid coils (which provide field strengths up to 3.5 T). The muon detectors, along with the toroid magnets, are illustrated in Figure 4.5.

The MS consists of four different types of detector components, which fulfill two different purposes: (i) precision tracking is predominantly done by the *Monitored Drift Tubes* (MDTs) in most of the pseudo-rapidity range, with the *Cathode Strip Chambers* (CSCs) (multi-wire proportional chambers) of higher granularity, dealing with the harsher conditions closer to the beam line in the forward directions; while (ii) triggering and second coordinate tracking is done by *Resistive Plate Chambers* (RPCs) in the barrel and by *Thin Gap Chambers* (TGCs) in the end-caps. Tracking is provided for pseudorapidities up to $|\eta| < 2.7$, whereas the trigger system only extends to $|\eta| < 2.4$.

Alignment of the muon chambers is essential to obtain precise track reconstruction, and thereby precise momentum estimates. Internal deformation and relative positions of the MDTs are monitored by 12,000 precision-mounted, optical sensors, in order to obtain the required 30 µm precision on the relative alignment of the chambers.

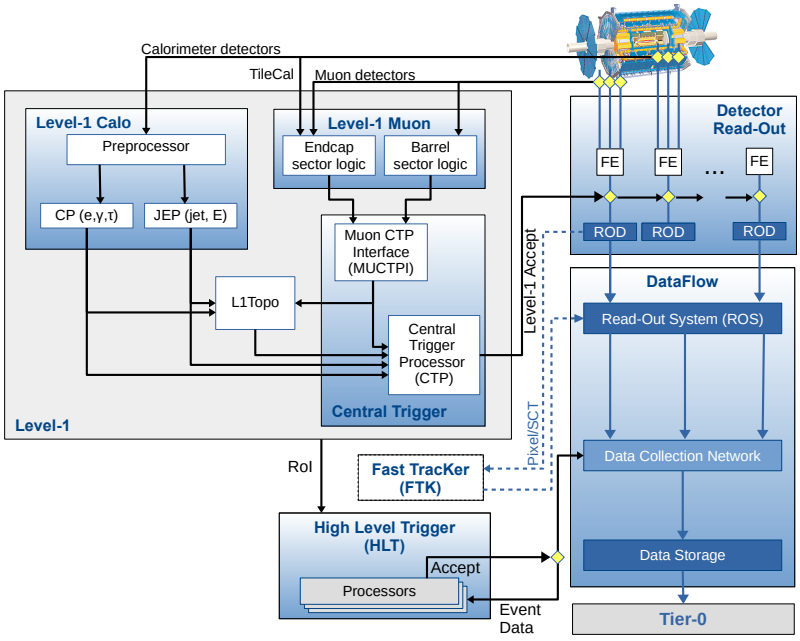


Figure 4.6: The components of the ATLAS TDAQ system [48] for Run 2.

4.2 Experimental data

The experimental dataset studied in this thesis is taken from measurements made by the ATLAS detector, of *pp* collisions provided by the LHC, from mid-2015 until the end of 2018. The *pp* collisions were conducted at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, corresponding to 6.5 TeV per proton. The protons were packed in bunches of 1.1×10^{11} protons per bunch, and accelerated up to nearly the speed of light, with a final bunch spacing of 25 ns. This translates to a collision rate of 40 MHz, i.e., 40 million collisions per second.

4.2.1 Triggering and data acquisition

The ATLAS *trigger and data acquisition* (TDAQ) system [48] performs *online* (real-time) processing of the detector measurements, and is responsible for selecting and storing the most interesting events for *offline* analysis (of recorded events). The TDAQ system is illustrated in Figure 4.6.

The *Level 1 trigger* (L1) is hardware-based, and uses custom electronics to trigger on reduced-granularity information from the calorimeters (*L1Calo*) and the muon system (*L1Muon*). In the L1Calo, the analogue signals are first digitized and calibrated by the preprocessor, before, in parallel, being handed over to the *Cluster Processor* (CP) for identification of electron, photon or

tau-candidates, and the *Jet/Energy-sum Processor* (JEP) for identification of jet candidates, and for performing global sums of total and missing transverse energy. The L1Muon trigger uses hits from the RPCs in the barrel and the TGCs in the end-caps to measure the momentum of the muons. In order to reject contributions from muon measurements in the end-caps that do not come from the interaction point, L1Muon puts coincidence requirements on the hits in the inner and outer TGC stations, and between the TGCs and the tile calorimeter.

The L1 trigger information is collected in the so-called *Central Trigger*, which outputs the final L1 trigger decisions. The decision can be based on event-level quantities, object multiplicities (above pre-defined threshold) or topological quantities, such as invariant masses or angular separations. Geometric and kinematic combinations of trigger objects from both the L1Calo and the L1Muon systems are collected and processed by the *L1Topo* trigger, which applies the topological requirements on the events. The L1 trigger reduces the event rates from the collision frequency of 40 MHz down to a maximum L1 trigger rate of 100 kHz.

For each event that has been accepted by the L1 trigger, the *Front-End* (FE) electronics read out the event data from all detectors. The data are then passed on to the *Read-Out Drivers* (RODs) for initial processing and formatting, before reaching the *Read-Out System* (ROS), which buffers the event data, awaiting a decision from the second trigger level.

The *High Level Trigger* (HLT) is software-based, and makes use of *Regions of Interest* (RoIs) in η and ϕ, which have been identified by the L1 trigger. The HLT typically runs dedicated fast trigger algorithms to provide early rejection, before running more CPU-intensive algorithms that are similar to full offline object and event reconstruction. These algorithms are run on a dedicated computing farm, making use of approximately 40,000 selection applications called *Processing Units* (PUs). The PUs make decisions within the order of a few hundred milliseconds. Multiple steps of feature-extraction algorithms are executed, using data fragments from within the RoI, until, finally, a hypothesis algorithm makes a decision on whether the trigger condition has been met or not. However, in some cases, such as triggering on $E_{\mathrm{T}}^{\mathrm{miss}}$, the full detector information of the event is needed. The HLT reduces the event rate even further, from the L1 rate of 100 kHz, down to about 1.2 kHz. This corresponds to about 1.2 GB/s of physics throughput.

The events that have been accepted by the HLT, are forwarded to the Tier-0 facility at the CERN Data Center, for offline reconstruction and permanent storage.

4.2.2 The Worldwide LHC Computing Grid

The data of the events that have passed the HLT, is transferred to the *CERN Data Center*, also referred to as *Tier 0*, for further processing and for permanent storage of the raw data. The first pass of the reconstruction of particle tracks and objects from the raw data, is done at Tier 0, before the reconstructed data

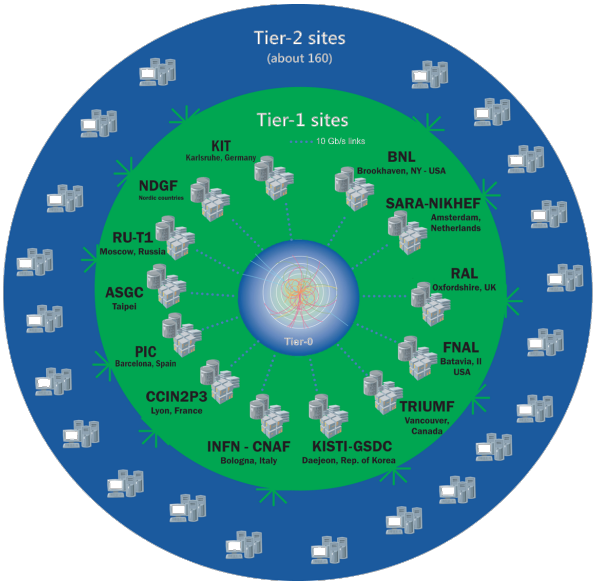


Figure 4.7: A schematic of the data centers which are part of the Worldwide LHC Computing Grid (WLCG).

are distributed to 13 large data centers, also known as *Tier 1s*, spread across the world.

These data centers are part of a global computing infrastructure called the *Worldwide LHC Computing Grid* (WLCG). The purpose of this computing infrastructure is to provide resources to store, distribute and analyze the data produced by the LHC experiments, and to make the data available to all partners regardless of their location on the globe.

The Tier 1s are responsible for storing a proportional amount of the raw and reconstructed data, running large re-processing campaigns for calibration of the data, and to distribute these data to smaller computing centers, called *Tier 2s*, which are located at around 160 universities and other scientific institutes, in 42 countries.

The Tier 2s have enough computing resources and disk space to handle user analysis tasks, and to produce and store the outputs of Monte Carlo simulations.

Figure 4.7 shows an illustration of the hierarchy of the computing centers which are part of the WLCG.

4.2.3 The dataset

The dataset [49] is divided into multiple levels of granularity, concerning different aspects of the data-taking. Given that the LHC has had an extended shut-down at the end of each year, for maintenance or upgrades, it is natural to divide the data according to which *year* it has been taken in. Furthermore, the data-taking

conditions often change throughout a given year, and so we divide a year into *periods* of uniform data-taking conditions (often separated by short technical stops). On a day-to-day basis, the LHC is normally able to provide collisions with so-called stable beams for a few hours at a time, before having to dump the remaining parts of the particle beams and do a restart, by re-filling the beams with new protons. This is referred to as a fill, and every time interval that ATLAS is able to uninterruptedly take data from an LHC fill is called a *run*. And finally, the smallest data collection unit (apart from single events), is what we call a *luminosity block* (LB). A LB typically corresponds to 60 seconds of data-taking.

4.2.4 Luminosity and pile-up measurements

The primary luminosity detectors of ATLAS are the *LUCID2* Cherenkov detectors, which consist of 16 photomultiplier tubes, and are placed about 17 m from the interaction point, in either direction along the beam line [49]. They measure the hit counts for every bunch crossing, and integrate these numbers for every luminosity block. Based on absolute luminosity calibrations, which happened once in each data-taking year, using beam separation scans in the x- and y-directions (so-called van der Meer scans), dedicated algorithms are used to estimate the visible interaction rate per bunch crossing, μ_{vis}, which in turn is proportional to the instantaneous luminosity. The integrated luminosity measurements are summarized per year of Run 2 in Table 4.1.

Figure 4.8 shows the integrated luminosity accumulated as a function of time. The green histogram represents the full luminosity delivered by the LHC, the yellow histogram the part that ATLAS was able to record, whereas the blue histogram shows the amount of this data that passed all the data quality (DQ) requirements needed to be used for physics analysis. Details on the DQ requirements are given in section 4.2.5. Figure 4.9 shows the distributions of the *mean number of interactions per bunch crossing*, often referred to as *pile-up*, split by year. We can see a large spread in the number of pile-up events per bunch crossing from year to year, but also within a given year. From the average of 13.4 interactions per crossing in 2015 to 37.8 in 2017, with the maximum number achieved during Run 2 extending up 70 interactions per bunch crossing. The average of the whole of Run 2 was 33.7 mean interactions per bunch crossing.

4.2.5 Data quality

In order to make sure that the quality of the experimental data, recorded by the ATLAS detector and subsequently reconstructed by dedicated software, is sufficiently high to be used for physics analysis, the performance of every detector component, the triggers and the data acquisition systems are constantly being monitored and scrutinized during periods of data-taking [50]. This monitoring consists of both automatic surveillance and flagging of potentially compromized data by software, but also by human shifters on-site in the ATLAS control room and remote shifters monitoring the performance via webpages. Some

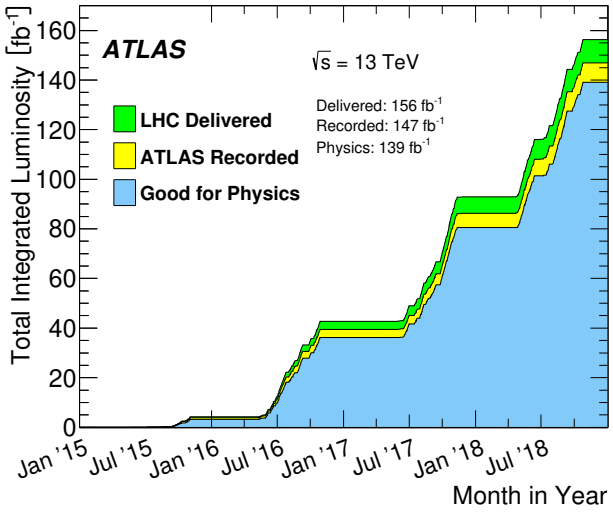


Figure 4.8: Integrated luminosity in Run 2, separated into the amount delivered by LHC (green), how much was recorded by ATLAS (yellow) and how much passed the data quality requirements in order to be labeled *good for physics* (blue) [50].

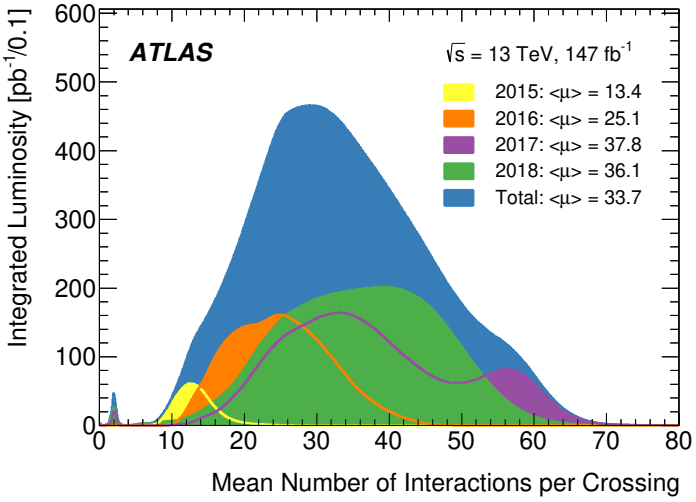


Figure 4.9: Distributions of the mean number of interactions per bunch crossing in Run 2, both by year and the combination of all the years in the run [50].

Table 4.1: Summary of the integrated luminosity taken by ATLAS during Run 2, which has been labeled *good for physics* [49]. The columns represent the year of data-taking, the measured integrated luminosity (int. lumi.), and the total (tot. unc.) and relative uncertainties (rel. unc.) on the integrated luminosity measurements.

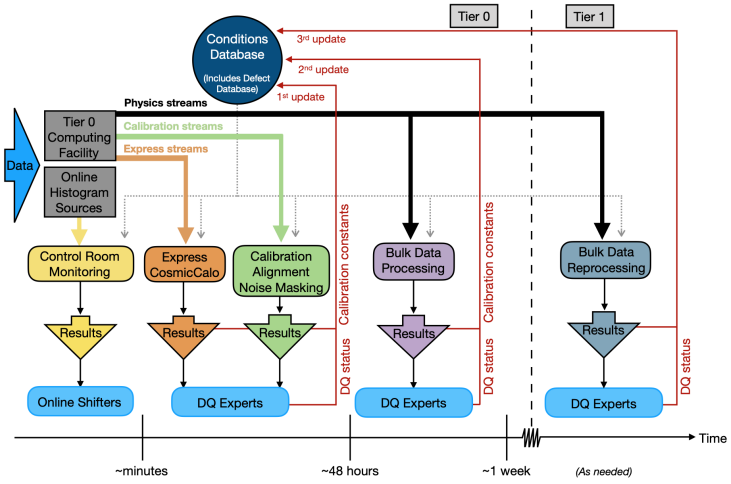
Year	**Int. lumi.** [$\mathbf{fb^{-1}}$]	**Tot. unc.** [$\mathbf{fb^{-1}}$]	**Rel. unc.** [%]
2015+16	36.2	0.8	2.1
2017	44.3	1.0	2.4
2018	58.5	1.2	2.0
Combined	139.0	2.4	1.7

of the shifters are responsible for specific detector systems, such as the inner tracker, the calorimeters or the muon system, separately, while other shifters are responsible for monitoring the combined performance of several sub-systems or reconstruction steps. In case of a change in performace, the software or the shifters can either mark the affected data as having a certain defect or supply calibration information to the conditions database, which stores the status of all parts affecting the measurements, acquisition and reconstruction of the data.

Figure 4.10 shows the flow of data processing, including feedback on the conditions of the data from the *data quality* (DQ) monitoring. Subsequent data (re-)processing then make use of the updated conditions to calibrate the reconstructed data. There is also DQ monitoring taking place during and after a data (re-)processing campaign to make sure that the calibrated data looks reasonable. In case something does not look right, the whole luminosity block (LB) containing the compromized data-taking interval is flagged as bad. Conversely, all the LBs that are flagged as *good for physics*, i.e., passing all the DQ requirements, are stored in the so-called *Good Run List* (GRL) for the corresponding year. Physics analyses only use data events from LBs that are listed in the GRL.

During a (re-)processing of the data, it is possible to drop data-taking intervals down to the order of milliseconds, while marking the rest of the corresponding LB as *good for physics*. In that case, the calculation of the integrated luminosity subtracts off the integrated luminosity of the sub-intervals within an LB which have been flagged as bad. In this way, one can retrieve some of the data that were actually *good for physics*, but initially fell into an LB where a fraction of the events were flagged as bad, and therefore the whole LB was flagged as bad.

Figure 4.11 shows the DQ efficiency for the 2015 through 2018 data-taking period. The DQ efficiency is calculated w.r.t. the amount of data that was recorded by ATLAS and intended for physics, and not w.r.t. the luminosity delivered by the LHC. The figure shows the cumulative DQ efficiency as function of the integrated luminosity, split by year. Because the DQ and ATLAS operation procedures have continuously been reviewed and developed throughout the Run 2 data-taking campaign, we can see a tendency of improvement in the DQ efficiency over time.



(a)

Figure 4.10: A schematic view of the data quality monitoring and (re-)processing procedures in ATLAS [50].

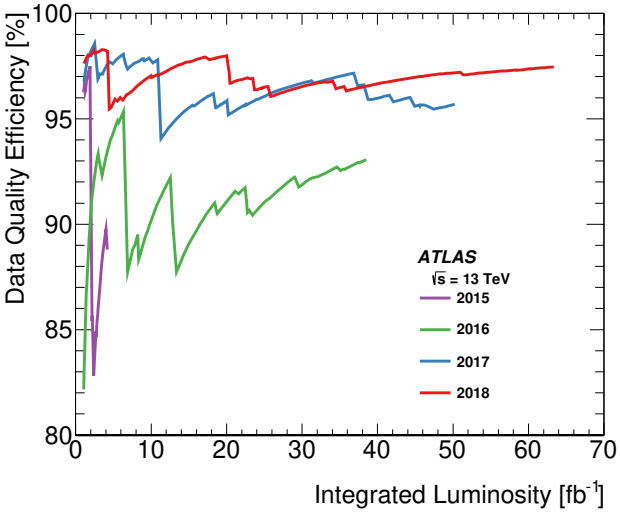


Figure 4.11: The fractions of the recorded data that were found to pass all the DQ requirements, as a function of integrated luminosity [50].

4.3 Simulated events

In order to make meaningful analyses of the experimental data, we need to understand which processes we expect to contribute to a given event/object selection, and at which rates. To that end, extensive resources are applied to perform *Monte Carlo (MC) simulations* [51] of the particle interactions that we expect to take place in the proton-proton collisions at the LHC, and eventually be measured by the ATLAS detector.

The simulations go through three main steps to get from theory to experimental signatures: (i) event generation, (ii) parton showering, hadronization and modeling of the underlying event, and (iii) propagation through the detector media.

4.3.1 Event generation

The first step in the simulation of a proton-proton interaction is to decide which, out of all possible interactions, is going to take place between the partons of the incoming protons in a given event. We make use of theoretical knowledge about which particles are allowed to take part in the given interaction, and at which rates, in terms of the S-matrix[1], or scattering matrix, formalism of perturbation theory (see any introductory book on quantum field theory, e.g., [43]). The number and type of matrix elements we decide to include in the event generation determine the accuracy at which we simulate the processes. Each set of simulations (*MC sample or dataset*) typically only target a specific type of interaction, or a set of interactions. In order to capture the full distribution of event types and kinematics for a physics analysis sufficiently well, we often need to include the contributions from many different types of independent simulations, which focus on different processes.

4.3.2 Parton showering, hadronization and pile-up

Due to the compositeness of the protons and the complexity of the QCD interactions, the complete dynamics of a proton-proton collision at the LHC is extremely complicated [52]. A pp collision typically produces hundreds of particles in a large number of interactions. Still, a few distinct stages in the evolution of such events have been identified, which are used in MC simulations to mimic the real events, before they reach the detectors.

[1] Each element of an S-matrix, or scattering matrix, holds (an expression for) the transition amplitude from a specific initial state to a specific final state in a scattering process. Each matrix element corresponds to a unique Feynman diagram. The transition from a given initial state to a given final state can happen via multiple different sub-processes, where the more processes we include in the S-matrix, the more accurate the calculation becomes. Typically, we include all possible processes up to a chosen order of perturbation theory, such as leading order (LO) or next-to-leading order (NLO) in a given coupling constant.

4.3.2.1 Parton showering and hadronization

Similarly to how electrically charged particles undergo bremsstrahlung (braking radiation) in the presence of an external electric field, color charged particles produce gluon radiation, called parton showering (PS), in the presence of an external color field. Developers of simulation software apply clever techniques to efficiently simulate realistic radiation processes and match them with the matrix elements of the event generation [52].

However, the bremsstrahlung of QED and the PS of QCD are not completely equivalent, as the photons do not have electric charge and thereby do not interact among themselves. The opposite is true for the gluons, as they do carry two color charges each (one color and one anti-color, e.g., $r\bar{b}$). This leads to the QCD phenomenon of confinement, where the attractive force between partons increase with separation. The products of the PS eventually experience the process of hadronization, where the color charged partons are confined to color neutral hadron states. A schematic showing the evolution of a pp collision simulation is depicted in Figure 4.12.

4.3.2.2 Pile-up conditions

The event generation step only produces one hard-scatter event between a pair of protons. In the experimental data, however, there are multiple proton-proton interactions/collisions happening in each bunch crossing. To simulate this, additional *minimum-bias events*[2] have been simulated separately, and are overlaid on top of the primary simulated hard-scatter event. The number of minimum-bias events to add to each event is drawn randomly from a distribution that attempts to replicate that of the mean number of interactions per bunch crossing in data.

As the distribution of the mean number of interactions per bunch crossing changed significantly with the year of data-taking in Run 2, it was decided to produce one set of MC samples for each data-taking year. Each set of MC samples reflects either a guess of what the pile-up distribution in data would look like, for early production of MC samples before or during the data-taking period, or the actual pile-up distribution, in the cases where the MC samples needed to be reproduced with a more accurate pile-up distribution after the data-taking period was over. The pile-up distributions that were used for the three MC production sub-campaigns are shown in Figure 4.13.

Additionally, to correct for any discrepancies in the simulated and measured pile-up distributions for a given year, a dedicated re-weighting procedure is applied to the simulated events. This correction up-weights the events with a pile-up number which was underrepresented compared to the experimental data, and vice-versa.

[2]Minimum bias in the sense that these events mimic events recorded by minimum-bias triggers in real data, which are minimally biased in their event selection, i.e., do not select events of particular characteristics (close to random selection).

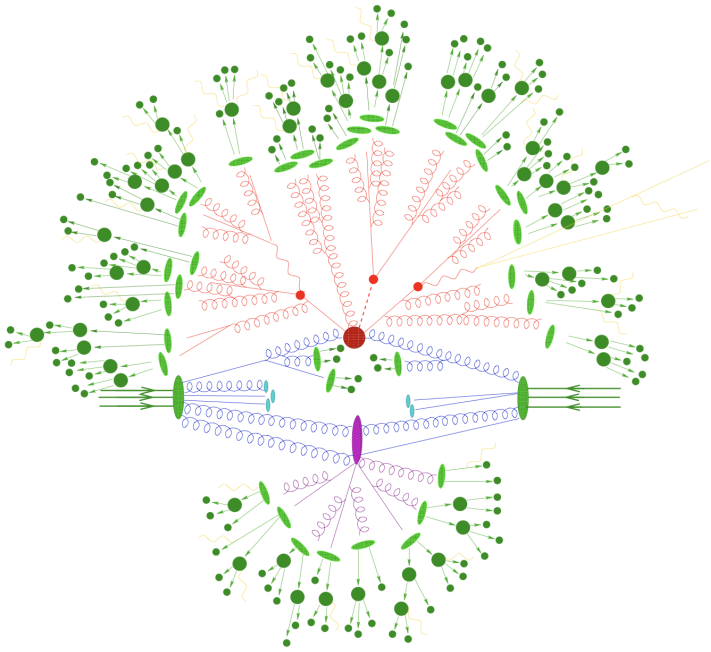


Figure 4.12: An example schematic of a simulated proton-proton collision event [53]. Upon incidence, the incoming protons (straight green lines, from left and right) break up into individual quarks (straight blue lines) and additional gluon radiation (blue curly lines). A quark from the left proton and a gluon from the right proton go through *initial state radiation* (ISR), producing, among other particles, a gluon each, which make up the intial state of the hard scatter event (big red blob). The hard scatter produces a $t\bar{t}H$ final state, where the unstable top quarks (red solid lines) and the Higgs boson (red dashed line) decay further to lighter quarks (red solid lines) and/or leptons (yellow solid lines), involving W bosons (red wavy lines) and photons (yellow wavy lines). In the presence of fields from other color charges, the hard final state quarks undergo parton showering (PS), whereby they lose momentum by gluon radiation (red curly lines). As the quarks/anti-quarks move away from each other, their mutual attraction continues to grow, until the energy density of their common color field is high enough to produce new quark-anti-quark-pairs. These newly created quarks/anti-quarks are then confined to the already existing quarks/anti-quarks, to form color neutral hadron states (light green blobs), in a process called hadronization. The hadrons can decay further to lighter and more stable hadrons (dark green blobs), which eventually reach the detectors. At the same time, the remnants of the incident protons, that do not take part in the hard(est) scatter interactions, may also interact with each other (purple blob). These additional particle interactions are referred to as the underlying event. The underlying event contribute so-called pile-up interactions or collisions, that is, additional interactions and particles that are not part of the main hard interaction. Photon radiation also happens at all stages of the evolution (yellow wavy lines).

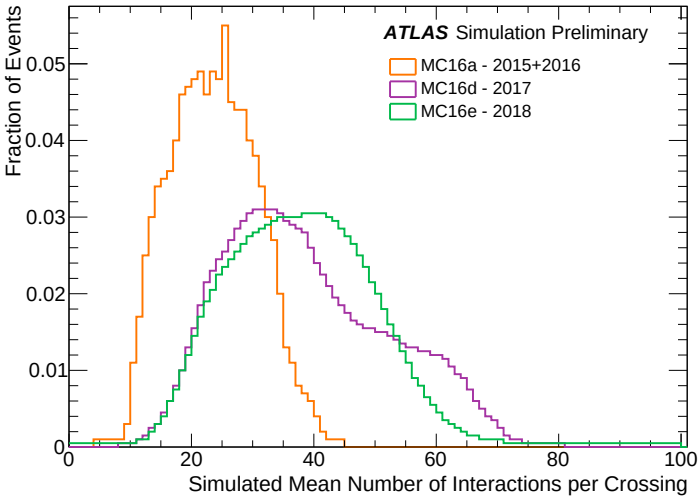


Figure 4.13: The plot shows the pile-up distributions of the three MC simulation sub-campaigns produced to match the actual distributions of the experimental datasets taken in 2015+2016 (orange), in 2017 (purple) and in 2018 (green) [54].

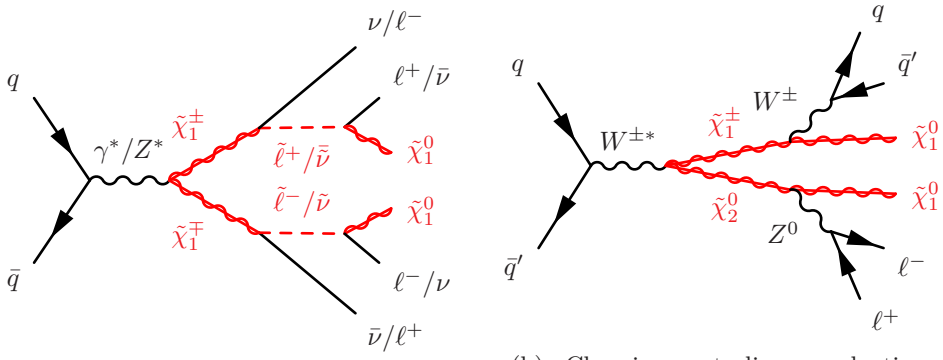
4.3.3 Detector simulation

Finally, when the particles that leave the collision point have been evolved in space and time, they eventually reach the different ATLAS sub-detectors. The simulation of how these particles interact with the detector materials are carried out using the GEANT4 software toolkit [55].

The full detector simulations consume a lot of computing resources, and this is in particular the case for the particle showers that take place in the electromagnetic and hadronic calorimeters. As a means to reduce the time and resources spent on this task, a fast calorimeter simulation procedure (Atlfast-II) [56] has been developed that uses a parametrization of the energy deposits of these electromagnetic and hadronic showers. The reduction in computational resources comes at the cost of slightly less accurate modeling of the energy deposits, and is therefore only used for cases where the analyses are not sensitive to this effect, or for BSM signal models where a high number of signal grid points is more important than the slight degradation in the simulated calorimeter response.

4.3.4 SUSY signal

In order to get a handle on what the experimental signatures of the SUSY particles would look like, we make use of MC simulated SUSY events. The two SUSY signal models considered in this thesis are: (i) chargino-pair ($\tilde{\chi}_1^{\pm}$ $\tilde{\chi}_1^{\mp}$) production with decays via sleptons into two leptons and two neutralino ($\tilde{\chi}_1^0$)



(a) Chargino-pair production with decays via sleptons or sneutrinoes into two leptons and two lightest neutralinos.

(b) Chargino-neutralino production with decays via the W and Z bosons into two leptons, two hadronic jets and two lightest neutralinos.

Figure 4.14: Diagrams of MC simulated SUSY processes.

LSPs in the final state (Section 6), and (ii) $\tilde{\chi}_1^{\pm}\,\tilde{\chi}_2^0$-production with decays via EW gauge bosons into two leptons, two light-flavor jets and two $\tilde{\chi}_1^0$ LSPs (Section 7 and 8). Diagrams illustrating these two processes are shown in Figure 4.14a and 4.14b, respectively.

Both of the SUSY signal models mentioned above have been generated with the MadGraph v2.2.3 generator [57], where the parton showering, hadronization and the underlying event have been handled by Pythia 8.186 [58] with the A14 tune [59]. Parton luminosities are given by the NNPDF2.3LO PDF set [60], and jet-parton matching was done using the CKKW-L prescription [61], at the matching scale equal to one quarter of the pair-produced sparticle mass.

The signal cross-sections have been calculated to NLO in the strong coupling constant, with soft gluon emission effects added to NNLL accuracy [62, 63, 64, 65, 66]. The nominal cross-sections and their uncertainties are taken from an envelope of predictions obtained by different PDF sets and renormalization and factorization scales, as described in Reference [67].

4.3.5 SM background processes

The SUSY signals described in the previous section both result in events with two leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$, plus two jets from a W boson in the $\tilde{\chi}_1^{\pm}\,\tilde{\chi}_2^0$ case. There also exist SM processes that result in the same final states as the SUSY signal, which we refer to as background.

In order to estimate the expected contributions from SM background processes to our event selections, we often make use of MC simulations to model these processes as well. Below, I will give a short overview of the MC generators and simulation setups that have been used to produce the background estimates for the analyses presented in the following chapters.

The simulated datasets come from two different ATLAS MC production campaigns: (i) *early Run 2*, samples generated prior to the start of Run 2 data-taking in 2015, which were basically in use for analyses of 2015+2016 data only, and (ii) *full Run 2*, samples generated for analyses of the full Run 2 dataset, which were basically in use from the summer of 2017 and onwards.

The first two analyses of this thesis, which I will refer to as *conventional analysis* (CA) 1 & 2, were conducted in the first half of Run 2, and therefore make use of simulated samples from the early Run 2 MC campaign. The two final analyses of the thesis, which I will refer to as CA 3 & *multivariate analysis* (MVA), were conducted in the second half and after the end of Run 2, and therefore make use of the simulated samples from the full Run 2 MC campaign.

Below, the MC samples from the two MC production campaigns are described separately, where the early Run 2 datasets are given under the label CA 1 & 2, and the full Run 2 datasets are given under the label CA 3 & MVA. The generator versions that were used for each of these two cases, are also summarized in Table 4.2.

4.3.5.1 Multiboson

Diagrams showing examples of SM diboson and triboson production processes are given in Figure 4.15 and 4.16, respectively.

CA 1 & 2 The SM diboson and triboson processes, involving W and Z bosons with leptons in the final state, were simulated using the Sherpa 2.2.1 generator [53, 68].

The fully-leptonic diboson processes, with four electroweak vertices, were generated with the NNPDF3.0nnlo PDF [69] at next-to-leading order (NLO) in the strong coupling constant with up to one additional parton for 4ℓ and $2\ell + 2\nu$, at NLO with no additional partons for $3\ell + \nu$, and at leading order (LO) with up to three additional partons for all the aforementioned processes.

The cases where one of the bosons decays hadronically, have similarly been simulated with up to one additional parton at NLO, and up to three additional partons at LO.

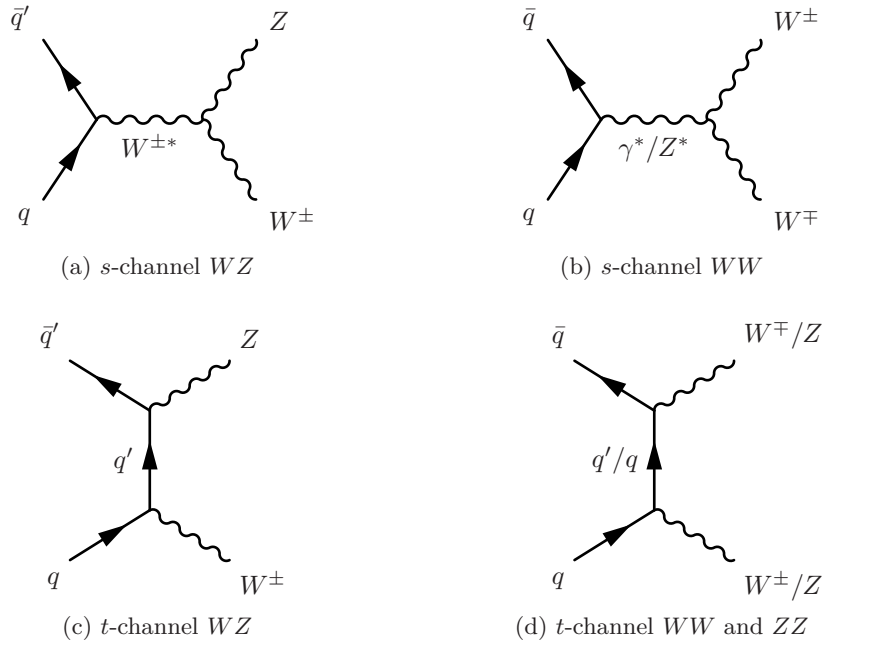
The di- and triboson processes with up to six electroweak vertices, including the same-sign WW process with associated jets, $W^{\pm}W^{\pm}jj$, were generated with the CT10 [70] PDF at NLO with no additional partons, and at LO for up to one and two additional partons for the diboson and triboson processes, respectively.

In all cases, the Comix [71] and OpenLoops [72, 73, 74] matrix element generators have been used to calculate additional hard parton emissions and virtual QCD corrections, respectively, and the resulting matrix elements merged with the Sherpa parton shower using the MEPS@NLO prescription [75, 76, 77, 78].

CA 3 & MVA Samples of diboson final states (VV) were simulated with the Sherpa 2.2.1 or 2.2.2 [68] generator depending on the process, including off-shell effects and Higgs boson contributions, where appropriate. Fully leptonic final

Table 4.2: An overview of the generator versions used to produce the MC simulated events for SM background estimation, for the three conventional analyses (CA 1, 2 & 3) and the multivariate analysis (MVA) presented in this thesis.

	CA 1 & 2	**CA 3 & MVA**
Diboson (VV)	Sherpa 2.2.1	Sherpa 2.2.2
Triboson (VVV)	Sherpa 2.1.1	Sherpa 2.2.2
V+jets	Sherpa 2.2.0/1	Sherpa 2.2.1
$t\bar{t}$, tW	Powheg + Pythia 8.186	Powheg + Pythia 8.230
tZ, $t\bar{t}V$, $t\bar{t}WW$, 3t, 4t	MadGraph5_aMC@NLO 2.2.2 + Pythia 8.186	MadGraph5_aMC@NLO 2.3.3 + Pythia 8.210
H (ggF, VH, VBF)	Powheg + Pythia 8.186	Powheg + Pythia 8.212
$t\bar{t}H$	MadGraph5_aMC@NLO 2.3.2 + Pythia 8.186	Powheg + Pythia 8.230



(a) s-channel WZ (b) s-channel WW

(c) t-channel WZ (d) t-channel WW and ZZ

Figure 4.15: Diagrams of SM diboson production. A superscript star indicates that the vector boson needs to be virtual and off mass shell, in order to produce its subsequent decay products. A quark marked with a prime, indicates that the quark is of a different quark flavor compared with a non-primed quark, i.e., by emitting or absorbing a W boson, an up-type quark becomes a down-type quark, and vice versa.

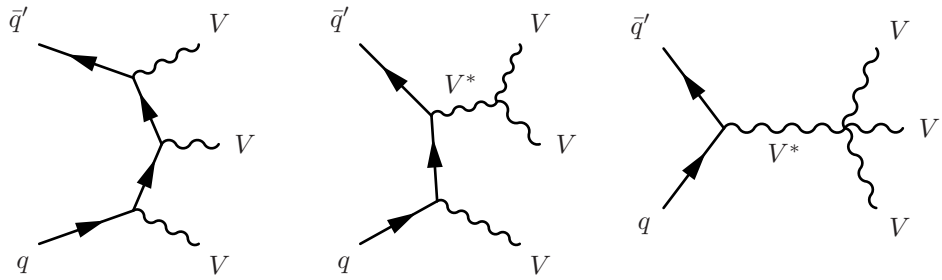


Figure 4.16: Diagrams of SM triboson production. V denotes an EW vector boson (W or Z), where a superscript star indicates that the vector boson needs to be virtual and off mass shell, in order to produce its subsequent decay products. A quark marked with a prime, indicates that the quark may be of a different quark flavor compared with a non-primed quark, depending on the actual vector bosons involved in the process.

states and semileptonic final states, where one boson decays leptonically and the other hadronically, were generated using matrix elements at NLO accuracy in QCD for up to one additional parton and at LO accuracy for up to three additional parton emissions. Samples for the loop-induced processes $gg \to VV$ were generated using LO-accurate matrix elements for up to one additional parton emission for both the cases of fully leptonic and semileptonic final states. The matrix element calculations were matched and merged with the Sherpa parton shower based on Catani–Seymour dipole factorisation [71, 79] using the MEPS@NLO prescription [75, 76, 77, 78]. The virtual QCD corrections were provided by the OpenLoops library [72, 73, 74]. The NNPDF3.0nnlo set of PDFs was used [69], along with the dedicated set of tuned parton-shower parameters developed by the Sherpa authors.

Electroweak production of a diboson in association with two jets ($VVjj$) was simulated with the Sherpa 2.2.2 [68] generator. The LO-accurate matrix elements were matched to a parton shower based on Catani–Seymour dipole factorisation [71, 79] using the MEPS@LO prescription [75, 76, 77, 78]. Samples were generated using the NNPDF3.0nnlo PDF set [69], along with the dedicated set of tuned parton-shower parameters developed by the Sherpa authors.

The production of triboson (VVV) events was simulated with the Sherpa 2.2.1 [68] generator. Matrix elements accurate to LO in QCD for up to one additional parton emission were matched and merged with the Sherpa parton shower based on Catani–Seymour dipole factorisation [71, 79] using the MEPS@LO prescription [75, 76, 77, 78]. Samples were generated using the NNPDF3.0nnlo PDF set [69], along with the dedicated set of tuned parton-shower parameters developed by the Sherpa authors.

4.3.5.2 V+jets

Diagrams showing examples of SM V+jets production processes are given in Figure 4.17.

CA 1 & 2 The production of a vector boson (W or Z) in association with hadronic jets have also been simulated with the Sherpa 2.2.0 and 2.2.1 generators [68], for CA 1 and CA 2, respectively..

Massive b- and c-quarks have been used in the simulations to improve the treatment of vector boson production with associated jets containing b- or c-hadrons.

The matrix elements have been calculated with the Comix [71] and OpenLoops [72, 73, 74] generators, and merged with the Sherpa parton shower using the ME+PS@NLO prescription [75, 76, 77, 78]. Matrix elements with up to two additional partons were calculated at NLO, whereas up to four additional partons were calculated at LO. Global k-factors[3] were applied to normalize the

[3] k-factors are scale factors that correct the cross-section used in the event generation to a higher order of perturbation theory. They are ratios that typically take an LO cross-section to NLO-precision, $k \equiv \sigma_{\mathrm{NLO}}/\sigma_{\mathrm{LO}}$, or from NLO to NNLO precision, $k \equiv \sigma_{\mathrm{NNLO}}/\sigma_{\mathrm{NLO}}$. The latter is the case for the V+jets samples considered here.

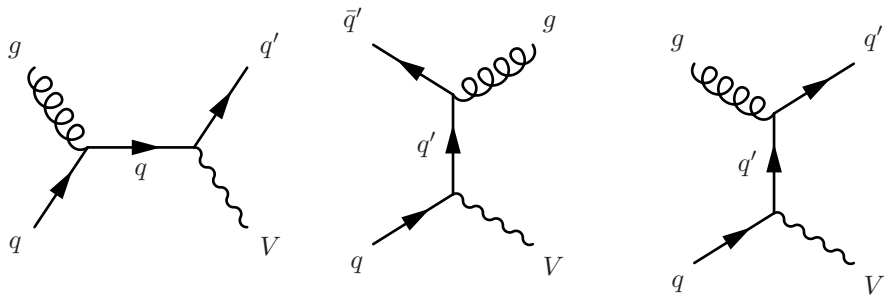


Figure 4.17: Diagrams of SM V+jets production. V denotes an EW vector boson (W or Z). A quark marked with a prime, indicates that the quark may be of a different quark flavor compared with a non-primed quark, depending on the actual vector bosons involved in the process.

QCD cross-sections to next-to-next-to-leading order (NNLO).

CA 3 & MVA The production of V+jets was simulated with the SHERPA 2.2.1 [68] generator using next-to-leading-order (NLO) matrix elements (ME) for up to two partons, and leading-order (LO) matrix elements for up to four partons calculated with the Comix [71] and OPENLOOPS [72, 73, 74] libraries. They were matched with the SHERPA parton shower [79] using the MEPS@NLO prescription [75, 76, 77, 78] using the set of tuned parameters developed by the SHERPA authors. The NNPDF3.0NNLO set of PDFs [69] was used and the samples were normalised to NNLO prediction [80].

Electroweak production of $\ell\ell jj$, $\ell\nu jj$ and $\nu\nu jj$ final states was simulated with SHERPA 2.2.1 [68] using leading-order (LO) matrix elements with up to two additional parton emissions. The matrix elements were merged with the SHERPA parton shower [79] following the MEPS@LO prescription [78] and using the set of tuned parameters developed by the SHERPA authors. The NNPDF3.0NNLO set of PDFs [69] was employed. The samples were produced using the VBF approximation, which avoids overlap with semileptonic diboson topologies by requiring a t-channel colour-singlet exchange.

4.3.5.3 Top-quark processes

Diagrams showing examples of SM $t\bar{t}$ and single top quark production processes are given in Figure 4.18 and 4.19, respectively.

CA 1 & 2 Production of $t\bar{t}$ and single-top in the tW-channel have been done with the POWHEG BOX v2 [81, 82, 83, 84] generator, using the NNPDF3.0NLO [69] PDF set, with parton showering done by Pythia 8.186. The mass of the top quark was taken to be 172.5 GeV. The $t\bar{t}$ events were normalized to NNLO+next-to-

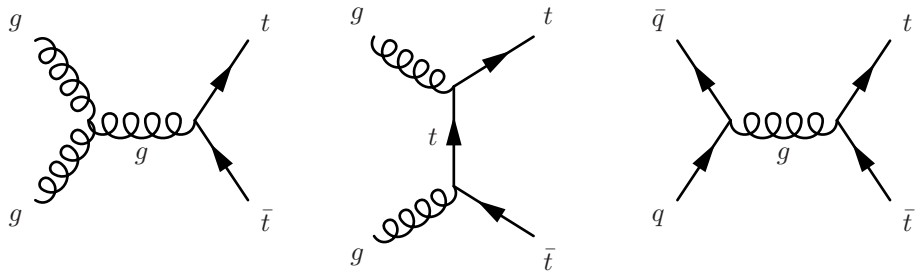
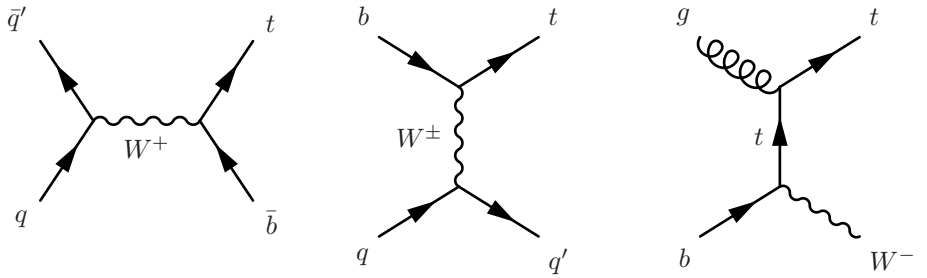


Figure 4.18: Diagrams of SM $t\bar{t}$ production.



(a) s-channel (b) t-channel (c) tW

Figure 4.19: Diagrams of SM single top production. A quark marked with a prime, indicates that the quark is of a different quark flavor compared with a non-primed quark.

next-to-leading-logarithm (NNLL) QCD cross-sections, whereas the tW events were normalized to NLO+NNLL cross-section.

The $t\bar{t}V$ and $t\bar{t}WW$ processes were generated at LO using the MadGraph5_aMC@NLO v2.2.2 [57] generator, which was interfaced to Pythia 8.186 [58] for parton showering, hadronization and description of the underlying event, with up to two ($t\bar{t}W$), one ($t\bar{t}Z$) or no ($t\bar{t}WW$) additional partons included in the matrix elements. MadGraph was also used to simulate the tZ, $t\bar{t}t$ and $t\bar{t}t\bar{t}$ processes. The NNPDF2.3LO [60] PDF set was used along with the A14 set of tuned parameters [59]. Events containing these top quark processes were normalized to their corresponding NLO cross-sections, apart from tZ and $t\bar{t}$ which were normalized to their LO generator cross-sections.

CA 3 & MVA The production of $t\bar{t}$ events was modelled using the POWHEG BOX v2 [81, 83, 84, 85] generator at NLO with the NNPDF3.0NLO [69]

PDF set and the h_{damp} parameter[4] set to $1.5\,m_{\text{top}}$ [86]. The events were interfaced to Pythia 8.230 [87] to model the parton shower, hadronisation, and underlying event, with parameters set according to the A14 tune [88] and using the NNPDF2.3lo set of PDFs [60]. The decays of bottom and charm hadrons were performed by EvtGen 1.6.0 [89].

The associated production of top quarks with W bosons (tW) was modelled by the Powheg Box v2 [81, 83, 85, 90] generator at NLO in QCD using the five-flavour scheme and the NNPDF3.0nlo set of PDFs [69]. The diagram removal scheme [91] was used to remove interference and overlap with $t\bar{t}$ production. The related uncertainty was estimated by comparison with an alternative sample generated using the diagram subtraction scheme [86, 91].The events were interfaced to Pythia 8.230 [87] using the A14 tune [88] and the NNPDF2.3lo set of PDFs [60].

Single-top t-channel production was modelled using the Powheg Box v2 [81, 83, 85, 92] generator at NLO in QCD using the four-flavour scheme and the corresponding NNPDF3.0nlo set of PDFs [69]. The events were interfaced with Pythia 8.230 [87] using the A14 tune [88] and the NNPDF2.3lo set of PDFs [60].

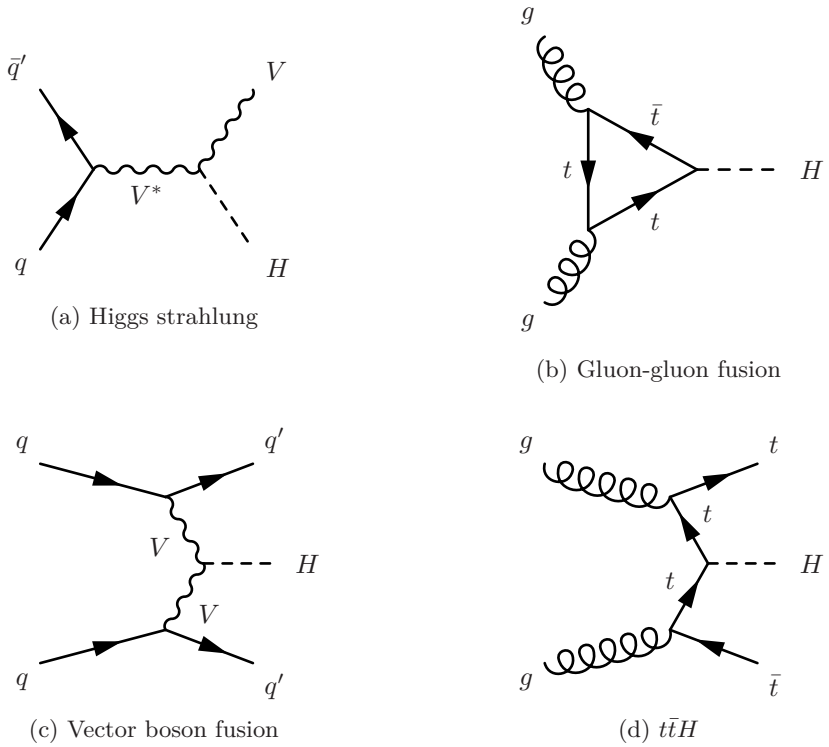
Single-top s-channel production was modelled using the Powheg Box v2 [81, 83, 85, 93] generator at NLO in QCD in the five-flavour scheme with the NNPDF3.0nlo [69] parton distribution function (PDF) set. The events were interfaced with Pythia 8.230 [87] using the A14 tune [88] and the NNPDF2.3lo PDF set.

The production of $t\bar{t}V$ events was modelled using the MadGraph5_aMC@NLO 2.3.3 [57] generator at NLO with the NNPDF3.0nlo [69] parton distribution function (PDF). The events were interfaced to Pythia 8.210 [87] using the A14 tune [88] and the NNPDF2.3lo [69] PDF set. The decays of bottom and charm hadrons were simulated using the EvtGen 1.2.0 program [89].

The production of tZq events was modelled using the MadGraph5_aMC@NLO 2.3.3 [57] generator at NLO with the NNPDF3.0nlo [69] parton distribution function (PDF). The events were interfaced with Pythia 8.230 [87] using the A14 tune [88] and the NNPDF2.3lo [69] PDF set.

The production of tWZ events was modelled using the MadGraph5_aMC@NLO 2.3.3 [57] generator at NLO with the NNPDF3.0nlo [69] parton distribution function (PDF). The events were interfaced with Pythia 8.212 [87] using the A14 tune [88] and the NNPDF2.3lo [69] PDF set. The decays of bottom and charm hadrons were simulated using the EvtGen 1.2.0 program [89].

[4]The h_{damp} parameter is a resummation damping factor and one of the parameters that controls the matching of Powheg matrix elements to the parton shower and thus effectively regulates the high-p_{T} radiation against which the $t\bar{t}$ system recoils.



(a) Higgs strahlung

(b) Gluon-gluon fusion

(c) Vector boson fusion

(d) $t\bar{t}H$

Figure 4.20: Diagrams of SM Higgs production. V denotes an EW vector boson (W or Z), where a superscript star indicates that the vector boson needs to be virtual and off mass shell, in order to produce its subsequent decay products. A quark marked with a prime, indicates that the quark may be of a different quark flavor compared with a non-primed quark, depending on whether the vector boson involved in the process is a Z or a W.

4.3.5.4 Higgs

Diagrams showing examples of SM Higgs production processes are given in Figure 4.20.

CA 1 & 2 Production of processes involving the Higgs boson, including gluon-gluon fusion, associated VH production and vector boson fusion, have been generated with POWHEG BOX v2 [81, 82, 83, 94] and Pythia 8.186 [58]. The resulting events were normalized to cross-sections calculated at NNLO with soft gluon emission effects added at NNLL accuracy.

$t\bar{t}H$ production was simulated with MadGraph5_aMC@NLO v2.3.2 [57] and Pythia 8.186, and the corresponding events normalized to their NLO cross-sections.

All of the aforementioned processes have been produced with the Higgs boson mass at 125 GeV.

CA 3 & MVA Higgs boson production via gluon–gluon fusion was simulated at next-to-next-to-leading-order (NNLO) accuracy in QCD using Powheg Box v2 [81, 83, 85, 95, 96]. The simulation achieved NNLO accuracy for arbitrary inclusive $gg \to H$ observables by reweighting the Higgs boson rapidity spectrum in Hj-MiNLO [97, 98, 99] to that of HNNLO [100]. The PDF4LHC15nnlo PDF set [101] and the AZNLO tune [102] of Pythia 8 [87] were used.

The gluon–gluon fusion prediction from the Monte Carlo samples was normalised to the next-to-next-to-next-to-leading-order cross-section in QCD plus electroweak corrections at next-to-leading order (NLO) [103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113]. The decays of bottom and charm hadrons were performed by EvtGen [89]. The normalisation of all Higgs boson samples accounts for the decay branching ratio calculated with HDECAY [114, 115, 116] and Prophecy4f [117, 118, 119].

Higgs boson production via vector-boson fusion was simulated with Powheg Box v2 [81, 83, 85, 94] and interfaced with Pythia 8 [87] for parton shower and non-perturbative effects, with parameters set according to the AZNLO tune [102]. The Powheg Box prediction is accurate to next-to-leading order (NLO) and uses the PDF4LHC15nlo PDF set [101]. It was normalised to an approximate-NNLO QCD cross-section with NLO electroweak corrections [120, 121, 122]. The decays of bottom and charm hadrons were performed by EvtGen [89]. The normalisation of all Higgs boson samples accounts for the decay branching ratio calculated with HDECAY [114, 115, 116] and Prophecy4f [117, 118, 119].

Higgs boson production in association with a vector boson was simulated using Powheg Box v2 [81, 83, 85, 94] and interfaced with Pythia 8 [87] for parton shower and non-perturbative effects. The Powheg Box prediction is accurate to next-to-leading order for VH boson plus one-jet production. The loop-induced $gg \to ZH$ process was generated separately at leading order. The PDF4LHC15nlo PDF set [101] and the AZNLO tune [102] of Pythia 8 [87] were used. The decays of bottom and charm hadrons were performed by EvtGen [89]. The Monte Carlo prediction was normalised to cross-sections calculated at NNLO in QCD with NLO electroweak corrections for $q\bar{q}/qg \to VH$ and at NLO and next-to-leading-logarithm accuracy in QCD for $gg \to ZH$ [123, 124, 125, 126, 127, 128, 129]. The normalisation of all Higgs boson samples accounts for the decay branching ratio calculated with HDECAY [114, 115, 116] and Prophecy4f [117, 118, 119].

The production of $t\bar{t}H$ events was modelled using the Powheg Box v2 [81, 83, 84, 85, 130] generator at NLO with the NNPDF3.0nlo [69] PDF set. The events were interfaced to Pythia 8.230 [87] using the A14 tune [88] and the NNPDF2.3lo [69] PDF set. The decays of bottom and charm hadrons were performed by EvtGen 1.6.0 [89].

4.3.6 b/c-hadrons and pile-up

CA 1 & 2 For all MC events, except those produced by Sherpa, the EvtGen 1.2.0 [89] program was used to simulate the properties of the b- and c-hadrons.

The effects of additional pp collisions in the same bunch crossing (in-time pile-up) and in neighboring bunch crossings (out-of-time pile-up) were simulated using the soft QCD processes of Pythia 8.186 [58] with the A2 [131] tune and the MSTW2008LO [132, 133].

CA 3 & MVA The effect of multiple interactions in the same and neighbouring bunch crossings (pileup) was modelled by overlaying the simulated hard-scattering event with inelastic proton–proton (pp) events generated with Pythia 8.186 [58] using the NNPDF2.3lo set of parton distribution functions (PDF) [60] and the A3 set of tuned parameters [134].

4.4 Fake and non-prompt lepton backgrounds

The background sources described so far, where the final-states consist of exactly the same particle content as the signal of interest, we refer to as *irreducible backgrounds*. There are, however, other SM processes which do not truly produce the same final-state as the signal, but still end up being measured or reconstructed that way. For final-states with leptons, which is the focus of this thesis, this can be caused either by (i) hadronic jets being interpreted as leptons by the object reconstruction, so-called *fake leptons*, or (ii) additional leptons arising from so-called *non-prompt* processes, e.g., final-states with semi-leptonic decays of heavy flavor hadrons or photon conversion to an e^+e^--pair. These fake and non-prompt (FNP) lepton backgrounds are referred to as *reducible backgrounds*, as we can reduce their contribution by improving the quality of the measurements or the reconstruction algorithms.

The sources of FNP leptons have been known to be sub-optimally modeled by MC in the past, and we therefore use data-driven estimates for these sources of background instead. The so-called *Matrix Method* (MM) [135] is used to calculate FNP weights to reflect the expected contributions from three different FNP sources: (i) *photon conversion* to an e^+e^--pair, (ii) semi-leptonic decays of *heavy-flavor* hadrons and (iii) mis-identification of *light-flavor* jets as electrons, to each of the signal regions.

The FNP weights are applied to the observed data, event by event, according to the event characteristics, and the resulting event yields are added to the background estimates.

4.5 Reconstruction and selection of events and objects

The events and the particles/objects are reconstructed in exactly the same way for both data and MC. Given the detector signals, the ATLAS reconstruction

software performs various forms of pattern recognition in order to link measurements from different sub-detector layers together, and form particle tracks and energy clusters arising from single particles/objects. Based on the reconstructed particle tracks and energy deposits, the software is then able to infer the particle identity and kinematic features of each object.

4.5.1 Events

The term *event*, in the context of particle collisions, refers to measurements taken in a time interval that contains only one bunch crossing. After particle tracks have been reconstructed, *interaction points/vertices* are also reconstructed based on where collections of tracks point back to the beam line. We have seen that during the LHC Run 2, the average number of interactions per bunch crossing was 33.7. This means that about 34 such interaction vertices were on average reconstructed per event.

For physics analysis, we only consider particles coming from a single interaction vertex per event. This vertex is chosen to be the one with the highest sum of associated track p_{T} squared, $\sum p_{\mathrm{T}}^2$, and is called the *primary vertex*. In order for an event to make it into the event selection, we require that it has a primary vertex with at least two associated tracks of $p_{\mathrm{T}} > 400\,\mathrm{MeV}$ each.

4.5.2 Baseline objects

We define two categories of quality criteria for the physics objects, such as electrons, muons and hadronic jets. These are referred to as (i) *baseline* objects, which are primarily used to estimate the so-called fake and non-prompt lepton contributions, and (ii) *signal* objects, which is a subset of the baseline objects and satisfy stronger quality requirements. The objects need to satisfy the signal requirements in order to make it into the final analysis selections.

4.5.2.1 Baseline electrons

Baseline electrons are reconstructed from isolated calorimeter deposits, which are matched to inner detector (ID) tracks. They need to satisfy $p_{\mathrm{T}} > 10\,\mathrm{GeV}$, have been measured with pseudorapidity $|\eta| < 2.47$, and satisfy the *Loose* likelihood-based identification *working point* (WP), described in [136, 137].

4.5.2.2 Baseline muons

Baseline muons are reconstructed from muon spectrometer (MS) tracks that satisfy $|\eta| < 2.7$ and are matched to ID tracks. They must have $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and satisfy the *Medium* muon identification WP, described in [138]. The latter is based on the number of hits and curvature measurements in the ID and MS systems.

4.5.2.3 Baseline jets

For reconstruction of hadronic jets, the anti-k_t algorithm [139], as implemented in the FastJet package [140], is used. The jets considered in this thesis are reconstructed from three-dimensional energy clusters [141] in the calorimeter within a cone of radius $R = 0.4$. In order to pass the baseline jet criteria, the jets must have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and have been measured within $|\eta| < 4.5$.

In order to reduce effects from pile-up, jets with $p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and $|\eta| < 2.4$ must have a significant fraction of their associated tracks compatible with coming from the primary vertex. We make this decision based on the score given to each jet by the multivariate Jet Vertex Tagger [142]. Additionally, the expected average energy contribution from pile-up jets is subtracted off the jet energies, according to the jet areas [143]. Events are also dropped if they contain any jet which is failing basic quality criteria or are identified to come from noise or non-collision background.

4.5.2.4 Baseline b-jets

A multivariate discriminant, called MV2c10 [144, 145], identifies jets that contain b-hadrons, making use of information about track impact parameters and reconstructed secondary vertices. This procedure is referred to as *b-tagging*, and we here use an identification requirement that correspond to 77% average efficiency for b-jet identification in simulated $t\bar{t}$ events. Apart from being b-tagged, the baseline b-jets also need to satisfy the regular baseline jet requirements.

4.5.2.5 Baseline photons

Photons do not leave tracks, as they are not electrically charged, but they are instead reconstructed from the energy they deposit in the electromagnetic calorimeters. In order to satisfy the baseline photon requirements, they must pass criteria of the *Tight* identification WP, described in [146, 147], have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.37$, but excluding the transition/crack region $1.37 < |\eta| < 1.52$, where the calorimeter performance is degraded.

4.5.2.6 Overlap removal

After all the baseline objects have been identified, we need to check for possible double-counting of objects. The overlap removal procedure is explained step-by step, in the order it is applied, below:

1. An electron sharing an ID track with a muon is removed.

2. If a b-tagged[5] jet is within $\Delta R = 0.2$ of an electron candidate, the electron is rejected, as it is likely coming from a semi-leptonic b-hadron decay. Otherwise, if the jet which is within $\Delta R = 0.2$ of the electron candidate

[5]The overlap removal procedure uses the 85% efficiency working point of the MV2c10 algorithm.

is not b-tagged, the jet is rejected, as it is likely coming from an electron-induced shower.

3. Electrons within $\Delta R = 0.4$ of a remaining jet candidate are discarded, in order to further suppress semi-leptonic decays of b- and c-hadrons.

4. Jets with a nearby muon that carries a significant fraction of the transverse momentum of the jet[6] are discarded either if the muon candidate is within $\Delta R = 0.2$ of the jet or if the muon is matched to a track associated with the jet. Only jets with less than three associated tracks can be discarded in this step.

5. Muons within $\Delta R = 0.4$ of a remaining jet candidate are discarded, to suppress muons from semi-leptonic decays of b- or c-hadrons.

4.5.3 Signal objects

In addition to the baseline object requirements, the signal objects also need to satisfy the criteria described below.

4.5.3.1 Signal electrons

Signal electrons must satisfy the *Medium* likelihood-based identification WP, as well as the *Tight* isolation WP [136, 137] in order to reduce contributions from fake and non-prompt electrons. The track associated with the electron must have a transverse impact parameter (with respect to the primary vertex), d_0, of $d_0/\sigma(d_0) < 5$, with $\sigma(d_0)$ being the uncertainty in d_0. There is also a requirement on the longitudinal impact parameter of the electron track (with respect to the primary vertex), z_0, of $|z_0 \sin\theta| < 0.5\,\text{mm}$. The impact parameters d_0 and z_0, along with other quantities involved in the perigee parametrization used in track reconstruction, are illustrated in Figure 4.21.

4.5.3.2 Signal muons

Signal muons must satisfy $|\eta| < 2.4$, in addition to the Tight isolation WP defined in [138]. They also need to fulfill the following requirements on the impact parameters of the associated tracks: $d_0/\sigma(d_0) < 3$ and $|z_0 \sin\theta| < 0.5\,\text{mm}$.

4.5.3.3 Signal jets

In addition to the requirements described for baseline jets, the signal jets also need to satisfy $|\eta| < 2.4$. The same goes for b-tagged jets.

[6] $p_{\text{T}}^{\mu} > 0.7 \sum p_{\text{T}}^{\text{jet tracks}}$, where p_{T}^{μ} and $p_{\text{T}}^{\text{jet tracks}}$ are the transverse momenta of the muon and the tracks associated with the jets, respectively.

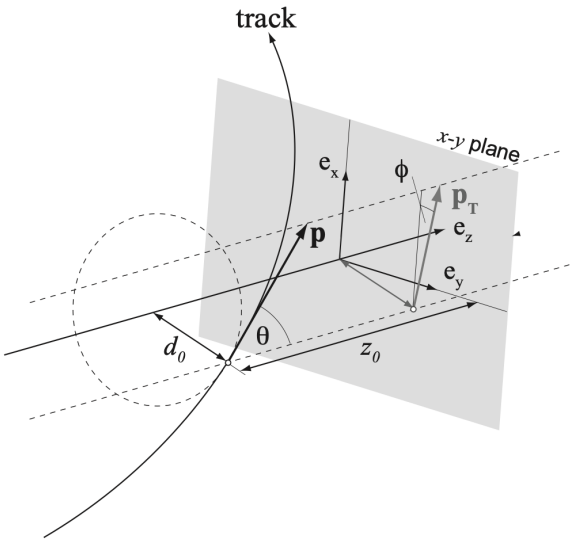


Figure 4.21: A schematic involving the five perigee parameters used in track reconstruction: the *transverse impact parameter*, d_0, the *longitudinal impact parameter*, z_0, the *azimuthal angle*, ϕ_0, of the track at the point (d_0, z_0), the *polar angle* of the track, θ, and the *signed charge over momentum*, q/p, representing the direction of curvature [148].

4.5.4 Missing transverse energy

The missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, is the magnitude of the negative vector sum of the transverse momenta of all the reconstructed objects (electrons, muons, photons and jets), in addition to the soft (low p_{T}) tracks which are associated with the primary vertex, but not with a reconstructed physics object. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated using the physics objects after they have been calibrated. See section 3.2.2.1 for a more formal definition of missing transverse energy.

4.5.5 Triggers

In order to select relevant collision events that have two high-p_{T} leptons, we require the events to have fired one of the *unprescaled*[7] *dilepton triggers*. This includes *dielectron*, *dimuon* and *electron-muon* triggers, both symmetric and

[7] In practice, there is a limit to the rate at which we can read out and store the full information about the collision events. In cases where the selection criteria of certain triggers are too inclusive, i.e., would read out too much data too often, we can apply *prescaling* to the relevant triggers. E.g., a trigger with a prescale factor of 1/10, only stores one out of ten events which fire the trigger in question. In that way, we can still make use of the trigger selection. However, we lose the full event information about nine out of ten events, and we need to multiply the resulting event rate by 10 to obtain the actual number of events which have fired the trigger. By only making use of *unprescaled* triggers, we have the full information

asymmetric in the p_{T}-thresholds of the 1st and 2nd lepton, in the range from 8 to 22 GeV. Since the triggers are not fully efficient starting from the online p_{T}-threshold (the threshold required for the event to be stored in the first place), we apply a so-called offline p_{T}-cut on top (at the analysis level) to ensure that the trigger efficiencies have reached their maximum values (the trigger efficiency plateau).

In the analysis targeting $\tilde{\chi}_1^{\pm}\ \tilde{\chi}_1^{\mp}$ production (Chapter 6), the offline p_{T}-thresholds are set to 25 and 20 GeV for the p_{T}-leading and -sub-leading leptons, respectively, whereas the analyses targeting $\tilde{\chi}_1^{\pm}\ \tilde{\chi}_2^{0}$ production (Chapter 7 and 8), require both the leading and sub-leading leptons to satisfy $p_{\mathrm{T}} > 25$ GeV.

For more general information about the trigger and data acquisition (TDAQ) system, see section 4.2.1, and for lists of the specific triggers used in the different years of data taking, see Appendix B.

4.5.6 Scale factors

The efficiencies of the signal lepton triggers, reconstruction, identification, isolation and b-tagging may differ slightly between data and MC. Therefore we apply *scale factors*, SFs, to the MC event weights, for each of the procedures mentioned above, that correct for these discrepancies. The scale factor for a specific quantity is determined as the ratio of efficiency measured in data, ϵ_{data}, to the efficiency found in simulated events, ϵ_{MC},

$$SF_i \equiv \frac{\epsilon_{i,\mathrm{data}}}{\epsilon_{i,\mathrm{MC}}}, \tag{4.1}$$

where i represents an index in the set of measured and reconstructed quantities mentioned above.

about all events that fire the trigger, which avoids blowing up the statistical uncertainties related to the application of prescale factors.

Part II

Search for Supersymmetry

Chapter 5

Analysis workflow

This second part of the thesis presents searches for supersymmetic particles in the data recorded by the ATLAS detector during the LHC Run 2. As a member of the ATLAS collaboration, I was part of an international working group searching for electroweak production of sparticles in final states with two leptons. The various tasks of the analyses were distributed among the team members, where progress was presented and discussed in weekly working group meetings.

In the chapters that follow, I will mainly focus on the parts of the analyses that I contributed to myself, which center on estimation of expected signal sensitivity with a given integrated luminosity, and definition of signal regions that optimize the sensitivity to particluar SUSY scenarios. Since the analyses presented in this thesis follow more or less the same workflow from start to finish, this first analysis chapter aims to give an overview of the main steps taken by a full, conventional analysis chain, before the actual work is presented in Chapter 6 through 9.

5.1 Signal model

The first step is to decide on which signal to look for. In our case, we wanted to investigate whether charginos and/or neutralinos have been produced in the proton-proton collisions. We then need to know how to recognize these particles in the data, i.e., the signature we expect them to produce in the detectors.

In the case of the Minimal Supersymmetric Standard Model (MSSM), there are more than 100 free parameters, which affect the composition of the charginos and neutralinos, their lifetimes, how they decay, etc. This practically makes it intractable to explore all the possibilities in which SUSY can be realized in nature. Therefore, we need to make some simplifying assumptions on the values of the parameters and/or relations between them, in order to reduce the number of free parameters to a more feasible set. Typically, we design and optimize searches for so-called *simplified models*. The analyses presented in this thesis consider simplified models where the decay chains of the sparticles are fully determined, and the only two free parameters of the models are the mass of the sparticles produced in the proton collision, which are assumed to be the next-to-lightest SUSY particles (NLSPs), and the mass of the lightest SUSY particle (LSP). Another alternative could be to fix the LSP mass (e.g., a massless LSP), and consider the branching fraction for NLSP decays to different SM particles as a free parameter instead.

If a given simplified model has a decay chain that allows different lepton or jet multiplicities (number of particles/objects) in the final state, searches will

typically also be carried out separately for each final state multiplicity.

5.2 Optimization of expected signal sensitivity

Once we know the signal model we are searching for and what kind of signatures it is expected to leave in the detectors, we can start to define a set of object and event selection criteria that optimize the sensitivity to the signal in question. Effectively, we want to increase the "signal-to-noise" ratio, where SM background processes that produce the same final state particles/objects as the signal, take the role of the "noise" in the event selection.

5.2.1 Monte Carlo

In this thesis, all sensitivity and signal region optimization studies are carried out using background estimates taken purely from Monte Carlo (MC) simulations. For the final statistical analysis, however, we can get more accurate modeling for some of the background contributions, such as the fake and non-prompt lepton backgrounds, using data-driven methods. But the data-driven estimates require a lot more work, compared to taking the estimates straight from simulation, and therefore MC samples are much more convenient to use when searching for approximately optimal event and object selections.

The way we optimize the signal sensitivity is to look at histograms of simulated event variables, such as the lepton and jet multiplicities, their transverse momenta, their combined masses, their angular separations, and so on, and identify variables that behave differently for signal and background. E.g., SUSY signal events typically tend to have higher values of missing transverse energy compared to the SM background processes. By requiring that the events have large $E_{\mathrm{T}}^{\mathrm{miss}}$, we can effectively cut away most of the background events and still keep most of the signal events.

5.2.2 Expected significance

The figure of merit that we want to optimize is the signal sensitivity or *expected significance* of measuring a signal-induced excess of events on top of the SM background. Given an expected number of background and signal events, B and S, respectively, the significance is a measure of how consistent, or rather how inconsistent, a measurement of $S + B$ is with the background-only hypothesis, where $S = 0$, given an estimated uncertainty on the number of background events.

In the final statistical analysis, we use a profile likelihood ratio test statistic, and may run pseudo-experiments in order to map out the expected test-statistic distribution for repeated experiments. However, in the process of optimizing the signal regions, it is overly tedious to search for optimal cuts by re-running the full statistical calculation every time we try a new selection or cut. Instead, we make use of a heuristic function to estimate the approximate significance, which is considered a satisfactory compromise for this purpose.

The heuristic function we make use of is called `BinomialExpZ`[1] [149, 150, 151], as implemented numerically in RooStats [153], and takes three input arguments: the expected number of *signal* (S) and *background* (B) events, along with the *relative uncertainty on the background* ($\Delta B/B$) in the signal region. This method assumes the background uncertainty has a Gaussian distribution, and we choose a flat 30% systematic uncertainty as a rough, conservative estimate in all studies, i.e., $\Delta B/B = 0.3$. The function calculates a significance, or Z-value, in one-tailed Gaussian standard deviations, from a hypothesis test of background-only vs. signal-plus-background. In References [149, 150, 151], this is referred to as the Z_{N}-value, where the subscript N refers to the *normal* (Gaussian) distribution assumed for the mean background estimate B.

Analytically, we may express the calculation of the Z-value as follows:

$$Z_{\mathrm{N}}(S, B, \Delta B/B) = \sqrt{2}\,\mathrm{erf}^{-1}(1 - 2p_{\mathrm{N}}), \tag{5.2}$$

where erf is the error function, and the p_{N}-value is calculated by

$$p_{\mathrm{N}} = \int_0^\infty db\, G(b; B, \Delta B) \sum_{n=S+b}^{\infty} P(n; b), \tag{5.3}$$

where $G(b; B, \Delta B)$ denotes a Gaussian distribution with expectation value B and standard deviation ΔB, and $P(n; b)$ denotes a Poisson distribution with expectation value b. The sum calculates the p-value of a Poisson distribution for values equal to or more extreme than $n = S + b$, where there essentially is a

[1] The term *binomial* in `BinomialExpZ` is derived from the so-called "on/off" problem in gamma-ray astronomy, which in HEP is encountered when we have a number of signal-plus-background events in a signal region, SR ("on signal source"), and a number of background-only events in a signal-free or sideband/control region, CR ("off signal source") [149, 150, 151]. The joint probability $P(n_{\mathrm{on}}, n_{\mathrm{off}}; \mu_{\mathrm{on}}, \mu_{\mathrm{off}})$ for a given measurement of n_{on} and n_{off} events in the two regions can either be modeled as a product of Poisson probabilities with separate expectation values μ_{on} and μ_{off}, or alternatively as a product of a combined Poisson probability with expectation value $\mu_{\mathrm{tot}} = \mu_{\mathrm{on}} + \mu_{\mathrm{off}}$ for measuring the total yield n_{tot}, with a binomial probability characterized by the ratio of the two Poisson means, $\lambda \equiv \mu_{\mathrm{off}}/\mu_{\mathrm{on}}$. The binomial parameter, i.e., the probability of an event falling into the SR or on-region, is then $\rho = \mu_{\mathrm{on}}/\mu_{\mathrm{tot}} = 1/(1 + \lambda)$. Under the background-only hypothesis, i.e., $\mu_{\mathrm{on}} = \mu_S + \mu_B = \mu_B$, we get $\lambda = \mu_{\mathrm{off}}/\mu_B \equiv \tau$. Since all information about the on/off-division of events is captured in the binomial probability, the p-value for the test of the background-only hypothesis can be calculated as a one-tailed binomial probability sum:

$$p_{\mathrm{Bi}} = \sum_{j=n_{\mathrm{on}}}^{n_{\mathrm{tot}}} \mathrm{Bi}(j|n_{\mathrm{tot}}; \rho). \tag{5.1}$$

RooStats calculates this sum numerically by use of the incomplete beta function [152].

Using the relation that $\tau = \mu_{\mathrm{off}}/\mu_B$, where the point estimates of μ_{off} and μ_B can be expressed as n_{off} and $\hat{\mu}_B = B$, with uncertainties $\sqrt{n_{\mathrm{off}}}$ and ΔB, respectively, we now have a relation between the formulation of the on/off problem in terms of τ and the "Gaussian-mean background" problem in terms of the relative uncertainty $\Delta B/B$, namely $\tau = B/(\Delta B)^2$. This is in fact what the `BinomialExpZ` function does internally. It translates the relative background uncertainty to a value for τ, and calculates the p-value, $p_{\mathrm{N}} = p_{\mathrm{Bi}}$, in terms of the incomplete beta function.

Bayesian integral over a Gaussian prior on the expected background yield B, over all possible background yields b. Numerically, the Poisson sum is calculated by use of the regularized, incomplete beta function [152], which allows non-integer values for S and b, which is frequently encountered when using scaled MC events to estimate expected event yields.

The convention in HEP is to exclude signal models at 95% C.L. The Z-value obtained from `BinomialExpZ` is strictly speaking an estimate of the expected "discovery significance" (Z_{disc}), i.e., how consistent $S+B$ is with B-only, given an estimate of the uncertainty in the background, ΔB. In order to evaluate exclusion potential (Z_{excl}), assuming no excess above B, we would instead need to consider how consistent B-only is with $S + B$, given an estimate of the uncertainty in the signal-plus-background yield, $\Delta(S + B)$. Under the assumption that $\Delta(S + B) \approx \Delta B$, we may consider the Z-value as a rough estimate of expected exclusion significance as well. This assumption may not always hold in practice, but in general the Z-value has proven to also be a useful measure of exclusion sensitivity, which is shown in comparisons of expected exclusion limits, obtained from Z-values, with the observed exclusion limits presented in Chapter 6 and 7.

The 95% percentile of a one-tailed Gaussian deviation translates to $Z \approx 1.64$. Thereby, we roughly expect to be able to exclude signal hypotheses (sparticle masses) at 95% C.L. (p-value ≤ 0.05) if the Z-value exceeds 1.64.

5.2.3 Signal regions

Since the event kinematics may vary quite a bit for different signal hypothesis, most notably w.r.t. the size of the mass-difference (or mass-splitting) between the sparticles involved in the simplified model, we optimize multiple sets of event selection which target different mass-splitting regions. We typically start by defining a selection that optimizes our sensitivity to *large mass-splittings*, or the *high-mass region* in terms of the pair-produced particles we can reach, where the production cross-section for the sparticles is the limiting factor rather than the mass-splitting itself. To cover the opposite case, we also design a selection which targets *small mass-splittings*, or the *low-mass region*, where low-p_{T} objects in the final state is the limiting factor. The overlap between the two regions also shows behavior that is different compared to each of the two extremes, so a third set of selections is designed, called the *intermediate-mass region*, to cover the transition between the low-p_{T} and high-p_{T} final state regimes.

Additionally, we may also add a category to select events with initial state radiation, where one of the incoming protons radiate a jet or a gluon that recoils against the rest of the event, and can boost the final state objects of small mass-splitting scenarios to higher transverse momenta. Such a selection is considered for the analysis in Chapter 7.

Yet another event selection may be defined to target off-shell decays of the next-to-lightest supersymmetric particle (NLSP) to the LSP, where the mass of an intermediate particle is too heavy to be produced in the given model, but can still contribute via suppressed off-mass-shell production. This scenario is considered in the analysis presented in Chapter 8.

Each such set of selection criteria, which is used to optimize the sensitivity to a certain signal scenario, we call a *signal region*, or SR for short.

5.3 Background modeling

When we have defined a set of signal regions, we want to make sure that we have modeled the background in these regions sufficiently well.

5.3.1 Control regions

To this end, we define *control regions*, or CRs, from a trade-off of three main objectives: (i) to select events dominated by a specific background process, (ii) to be reasonably close to the SR(s), but at the same time have negligible signal contamination, and (iii) to have high statistics (large number of events) in order to constrain the background normalization of a given process in the SR(s). We usually make dedicated CRs for the main background processes in each SR.

Then, we fit normalization factors for the main backgrounds to the experimental data in the CRs and SRs simultaneously, to correct for possible mis-modeling in the normalization of these backgrounds. Another advantage of doing such a fit, is the cancellation of a number of systematic uncertainties related to the normalization of the backgrounds.

5.3.2 Validation regions

After the simultaneous fit of the background to data in the CRs and SRs, the resulting background estimates are extrapolated to and validated in regions closer to the SRs. These are called *validation regions* or VRs. After validation, the background estimates from the CRs are finally extrapolated to the SRs.

In cases where a specific background process is found to be sufficiently well modeled, either directly from MC (e.g., with normalization factors consistent with 1 within uncertainties) or by construction from a data-driven estimate, we may skip the fit to data in a CR, and simply validate the background modeling in a VR directly.

5.3.3 Data-driven estimates

If the shapes of the background distributions in MC are not matching the corresponding distributions in data closely enough, we may instead need to use data-driven methods to extract the shapes of the distributions directly from the experimental data. This can be done by defining CRs dominated by the background process in question, using appropriate methods to estimate the sizes and shapes of the background contributions in the regions and subtract off any contamination in the CRs from other backgrounds that we are more confident in the modeling of (e.g., from MC). Then we take the remainder as our data-driven estimate of this background. The CR estimates are then extrapolated to VRs

to check the validity in regions which are independent from where they were extracted, and finally they are extrapolated to the SRs.

5.4 Systematic uncertainties

In addition to *statistical uncertainties* on the number of events from limited-sized datasets, there are also a number of other uncertainties which we refer to as *systematic uncertainties*. We can broadly put them in three categories: (i) *experimental uncertainties* related to the reconstruction and calibration of experimental measurements, (ii) *theoretical uncertainties* on parameters that enter the Monte Carlo simulations, and (iii) *uncertainties from data-driven background estimates*.

5.4.1 Experimental uncertainties

A large number of systematic uncertainties arise all the way from resolution effects in the various sub-detector systems, reconstruction of the momenta of particle tracks from hits in multiple detector layers, the assignment of energy deposits in the calorimeters to different reconstructed objects, calibration of the energy of objects across different energy scales and sub-detectors, and so on. Every type of reconstructed particle or object, such as electrons and photons, muons, hadronic jets, taus and $E_{\mathrm{T}}^{\mathrm{miss}}$, have their associated uncertainties, which we need to account for in the statistical analysis, in order to get results which are as unbiased as possible.

So-called kinematic uncertainties, related to the reconstruction and calibration of energy scale and resolution of the individual objects, are provided by combined performance groups in terms of $1\,\sigma$ up- and down-variations, which can be applied to the analysis by the use of dedicated software tools. We produce one set of analysis ntuples for each systematic variation, which then have propagated the effect of the given systematic variation through to the final analysis output.

There are also uncertainties related to the reconstruction efficiencies of different aspect of the events, such as particle identification, flavor tagging, triggers, and so on, which can be applied as $1\,\sigma$ up- and down-variations in terms of event weights. For these efficiency uncertainties, it suffices to store the weights in the nominal (main) ntuple, without the need for an extra copy of the ntuple for each variation.

See Section 8.5.1 for more details on the implementation in the full Run 2 analysis.

5.4.2 Theoretical uncertainties

In order to produce the MC simulations, we need to put in numbers for quantities such as the parton distribution function (PDF) of the colliding protons, the coupling strengths, the cross-sections for particles involved in the interactions, etc., which are theoretical parameters with associated uncertainties. The cross-

sections are, e.g., affected by which order in perturbation theory they have been calculated to, and the renormalization and factorization used.

Estimates for $1\,\sigma$ up- and down-variations on the PDF, coupling strengths and the renormalization and factorization scales are provided by the physics modeling group in terms of event weights.

See Section 8.5.2 for more details on the implementation in the full Run 2 analysis.

5.4.3 Data-driven background modeling uncertainties

When we extract a background estimate directly from the data, we need some way of quantifying the uncertainties on these estimates. This can, e.g., be done by estimating the same background by two different data-driven methods, and take the difference in the final estimates as the uncertainty on the method we choose to use in the end. Another way of estimating the uncertainty is to vary the control region selections systematically up and down, and take the corresponding difference as a measure of the uncertainty on the data-driven background estimate. One can also combine different sources/estimates of uncertainties in quadrature, if several uncorrelated sources are expected to contribute.

5.5 Statistical analysis

For the final statistical interpretation we use the HistFitter framework [154], which has become the primary tool for statistical data analysis in searches for supersymmetry performed by ATLAS.

We supply HistFitter with the set of all analysis ntuples, including both the nominal estimates and all the systematic variations, which in some cases are given in terms of simple event weights to multiply the nominal estimates with and in other cases in terms of additional ntuples. HistFitter then makes histograms for each control, validation and signal region, before making use of the HistFactory package to define probability density functions (PDFs) for these regions.

The likelihood function used to do fits of the background to data in the CRs and SRs, consists of a Poisson factor for each CR and SR, typically along with factors putting Gaussian constraints on the nuisance parameters, the latter representing the systematic variations in the fit. The fits are performed using the RooFit [155] package.

Finally, HistFitter makes use of RooStats [153] to perform hypothesis tests, first for checking the compatibility of the data with the background-only hypothesis, and then, if no significant excess has been observed, for setting limits on the signal model in question. The test statistic considered by the LHC experiments is the profile likelihood ratio, where the CLs [156] method is used for the purpose of exclusion. The calculations and interpretations of the hypothesis tests are carried out in a frequentist fashion, i.e., by calculating frequentist test statistics and p-values for the hypotheses.

For more detailed information about the HistFitter analysis for the full Run 2 dataset, see Section 8.6.

5.6 Overview of analysis chapters

The following chapters present studies and results that are part of analysis efforts performed by the ATLAS working group on searches for electroweak production of sparticles in final states with two leptons. The results presented in Chapter 6 and 7 are sensitivity studies that each laid ground for respective ATLAS publications. The analysis presented in Chapter 8 represents my contributions to an ATLAS paper that was, at the time of writing, just recently published, whereas the multivariate analysis presented in Chapter 9 is only considered a feasibility study, and will not be part of an ATLAS publication.

The focus of these chapters will primarily be on the parts of the analyses that I have been directly involved in myself. The main outcomes of the publications will also be presented, but for a full review you will be referred to the respective papers.

Chapter 6 describes the first study I was involved in, just after the start of the LHC Run 2 in 2015. The task was to check if we expected to surpass our sensitivity to chargino-pair production with decays via sleptons or sneutrinos into two leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$ in the final state, with the luminosity we expected to have for the summer conferences in 2016. This chapter therefore focuses mostly on the sensitivity studies that led up to the following publication, but also presents the final exclusion limits obtained for this signal model with $13.3\,\mathrm{fb}^{-1}$ of data.

Chapter 7 describes my next involvement, where I moved on to do a similar sensitivity study, but then for chargino-neutralino production with decays via SM gauge bosons into two leptons, two jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ in the final state, with the luminosity we expected to have for the spring conferences of 2017. The work also included optimization of the signal regions used in the previous iteration of this analysis, to accommodate the larger amount of data. The published results from the first two years of Run 2, which amounted to $36.1\,\mathrm{fb}^{-1}$ of data collected by ATLAS, is also presented.

Chapter 8 describes the work I have done on the analysis of the full Run 2 dataset. The targeted signal model is again the chargino-neutralino production with decays via SM gauge bosons into final states with two leptons, two jets and $E_{\mathrm{T}}^{\mathrm{miss}}$. This time I was more involved in the whole process from maintaining common analysis code, producing new MC samples, developing signal, control and validation regions, to estimating the systematic unertainties and performing the statistical interpretation in the end. Results corresponding to $139\,\mathrm{fb}^{-1}$ of integrated luminosity are presented.

Finally, in Chapter 9, I describe a feasibility study on the use of machine learning techniques in an attempt to increase the expected sensitivity to the same simplified model as discussed in the two previous chapters, with the luminosity of the Run 2 dataset. The studies only show expected sensitivity in the multivariate analysis signal regions, and do only unblind data in loose validation regions.

Chapter 6

Search for $\tilde{\chi}_1^{\pm}\,\tilde{\chi}_1^{\mp}$ production

At 10:40 in the morning on June 3, 2015, the LHC finally declared stable beams, after the accelerator team and the associated experiments had spent the last 27 months making upgrades and preparations for proton-proton collisions at the unprecedented center-of-mass energy of 13 TeV [157]. This marked the start of the LHC Run 2 and the exploration of so far uncharted territory. The approximate doubling of the collision energy, up from 7 and 8 TeV in Run 1, gave hopes of discovering new and heavier particles, that had previously been out of reach.

In particular, there was great anticipation tied to whether supersymmetric particles would be discovered in this new energy range. From a naturalness point of view, the supersymmetric partners of the heavier SM particles, like the higgsino, wino, bino, stop and stau, are expected to be found at the TeV-scale in order to keep the loop corrections to the higgs mass at an acceptable level. Otherwise, SUSY would not be as natural and attractive a solution to the hierarchy problem as many people in the community have hoped for. To many, the LHC Run 2 is a make-or-break for whether supersymmetry is likely to be realized in nature or not.

With these questions in the back of our minds, we began the search for signs of supersymmetric signals in the fresh datasets, with anticipation and excitement.

6.1 Signal model

My first task was to look for production of chargino-pairs, $\tilde{\chi}_1^{\pm}\,\tilde{\chi}_1^{\mp}$, decaying to the lightest neutralino, $\tilde{\chi}_1^0$, via a charged slepton, $\tilde{\ell}$, or a sneutrino, $\tilde{\nu}$. This process leads to events with two leptons (electrons or muons) and missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, in the final state – the latter arising because the neutrinos, ν, and the $\tilde{\chi}_1^0$ will escape detection. A diagram of this process is shown in Figure 6.1.

The signal process we are looking for in this case is a simplified model, where a set of restricting assumptions have been made on the masses and possible decays of the sparticles involved. The chargino is assumed to be pure wino, whereas the the lightest neutralino is taken to be pure bino and the lightest SUSY particle. We assume that R-parity is conserved, which makes the $\tilde{\chi}_1^0$ stable and unable to decay further to SM particles. The sleptons are taken to be the superpartners of the left-handed leptons, i.e. $\tilde{e}_{\mathrm{L}}$, $\tilde{\nu}_{e,\mathrm{L}}$, $\tilde{\mu}_{\mathrm{L}}$ and $\tilde{\nu}_{\mu,\mathrm{L}}$, with a mass halfway between the $\tilde{\chi}_1^{\pm}$ and the $\tilde{\chi}_1^0$, $m_{\tilde{\ell}_{\mathrm{L}},\tilde{\nu}_{\mathrm{L}}} = (m_{\tilde{\chi}_1^{\pm}} + m_{\tilde{\chi}_1^0})/2$.

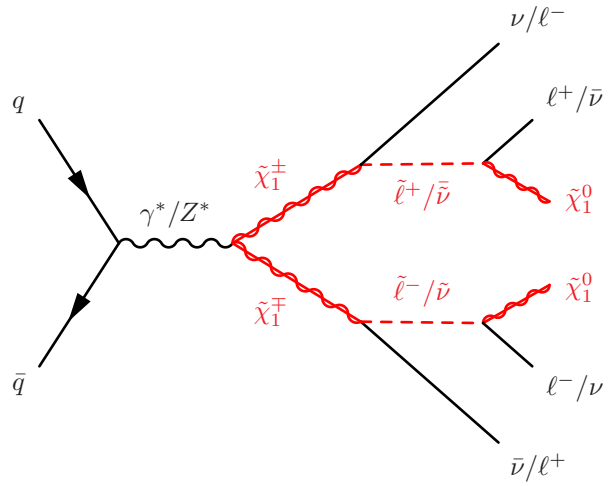


Figure 6.1: Chargino-pair production with decays via sleptons or sneutrinoes into two leptons and two lightest neutralinos.

6.2 Event and object selection

The event and object selections which are common to all analyses presented in the thesis, were introduced in Section 4.5. There is, however, an additional selection on forward jets used for the publication of the final analysis in Reference [158], that is unique to this chapter, and is explained below. However, I did not apply a selection on forward jets in my sensitivity studies.

As there are no hadronic jets involved in the hard-scatter process of this simplified model, we would like to reject (veto) events with significant, high-p_{T} jet activity. Because we collide protons, there will to some extent always be QCD-processes involved, be it elastic scattering or soft pileup-activity from other quark, anti-quark and/or gluon interactions taking place in the same bunch crossing. Therefore, we need to have a set of signal jet definitions as well, in order to register whether high-p_{T} jets most likely have been produced in the hard-scatter event.

We define three categories of hadronic signal jets: (i) *central light-flavor jets*, which correspond to the jet definitions in Section 4.5.2.3 and 4.5.3.3; (ii) *central b-jets*, which correspond to the b-jet definitions in Section 4.5.2.4 and 4.5.3.3; and unique to this analysis, a third category of (iii) *forward jets*, which means they are measured in the forward regions of the ATLAS detector with $2.4 < |\eta| < 4.5$, and additionally need to satisfy $p_{\mathrm{T}} > 30\,\mathrm{GeV}$.

6.3 Sensitivity studies

As a member of the working group in ATLAS, searching for electroweak production of SUSY particles in events with two or three leptons plus $E_{\mathrm{T}}^{\mathrm{miss}}$, my role

was to investigate how sensitive we expected to be to chargino-pair production with decays via sleptons, with the amount of data we expected to have in time for the summer conferences in 2016. The outcome of this study would help us understand whether we expected to surpass the sensitivity of Run 1 with the data collected by ATLAS during the spring of 2016, or whether we would need to wait for more data. It would essentially tell us whether this particular signal model was worth pursuing for the first publication.

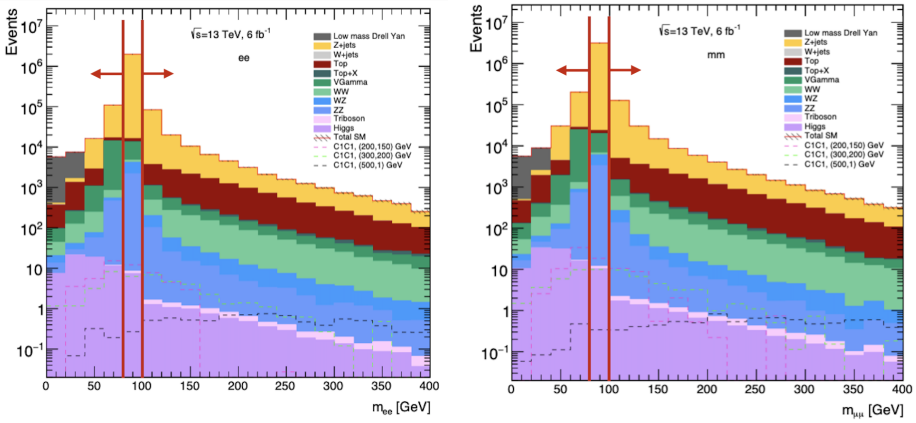
For these studies, we used Monte Carlo (MC) simulated samples both for the signal process and for all relevant SM background processes that can give two leptons in the final state. The MC samples are described in more detail in Section 4.3.4 and 4.3.5, respectively.

The signal regions considered for these sensitivity studies are given in table 6.1, and are based on the ones used in the previous 8 TeV-iteration of this analysis [159]. This entails selecting events with two oppositely charged leptons, split in two categories according to whether the two leptons are of the same lepton flavor, e^+e^- or $\mu^+\mu^-$ (SF), or of different lepton flavors, $e^+\mu^-$ or μ^+e^- (DF). The reason for this separation is the distinct difference in background composition between the two categories. In addition, the invariant mass of the two leptons, $m_{\ell\ell}$, is required to not be within 10 GeV of the Z-mass, in order to reject leptons from backgrounds involving Z-bosons, such as $Z+$ jets, ZZ and WZ events. Furthermore, a lower $E_{\mathrm{T}}^{\mathrm{miss}}$ cut on 40 GeV throws away a significant part of the SM backgrounds in general, while retaining most of the SUSY events, and finally we also veto any signal jets. Then we are left with the so-called *stransverse* mass, m_{T2}, which is defined in Section 3.2.2.2.

The stransverse mass turns out to be the most sensitive variable to this signal model. It is very efficient at rejecting background processes involving the W-boson, like WW diboson production, and processes involving top quarks, like single-top and $t\bar{t}$ production, with subsequent $t \to W + b$ decays, since leptons $+ E_{\mathrm{T}}^{\mathrm{miss}}$ from a W-boson result in a characteristic kinematic endpoint at the W-mass. The signal models, on the other hand, give rise to a similar endpoint at values of m_{T2} equal to the mass-difference between $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_1^0$. This makes the stransverse mass an excellent discriminant between signal and background for $m_{\mathrm{T2}} > m_W \equiv 80.4\,\mathrm{GeV}$. For these studies, I consider four different regions given by $m_{\mathrm{T2}} > 90, 120, 150$ and 180 GeV, respectively, which target increasing mass-differences between $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_1^0$.

Distributions of the variables that are involved in the signal region definitions are shown for the same flavor leptons in Figure 6.2 through 6.5, and for different flavor leptons in Figure 6.6. The red lines mark the cut-values, and the arrows point towards the region(s) that are kept. The cuts are applied sequentially, in the order the figures are shown, where the later figures have the cuts from previous figures applied. E.g., the $E_{\mathrm{T}}^{\mathrm{miss}}$-distributions are shown after the cuts on $m_{\ell\ell}$ have been applied, and so on.

At the time, we did not know how much data would be collected by the ATLAS detector in time for the publication. I therefore show expected sensitivities for two integrated luminosity scenarios: $6\,\mathrm{fb}^{-1}$ in Figure 6.7 and $10\,\mathrm{fb}^{-1}$ in Figure 6.8. In each case, an approximate one-tailed Gaussian Z_{N}-significance is shown on

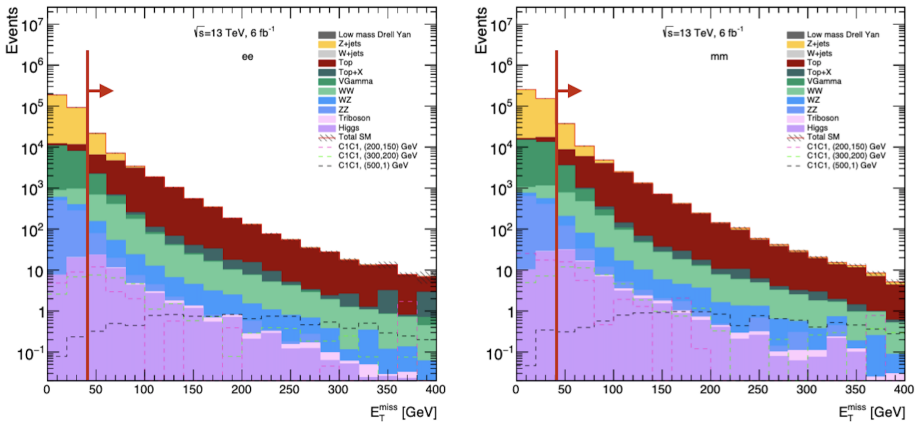


(a) e^+e^-

(b) $\mu^+\mu^-$

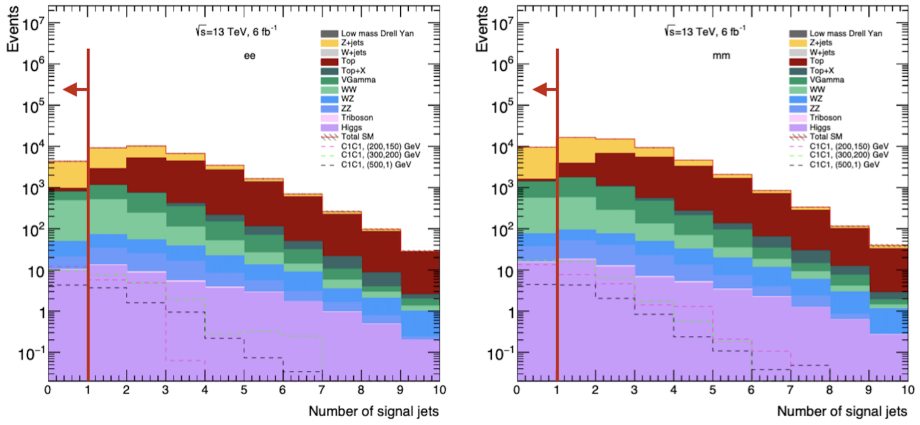
(c) Enlarged legend. C1 is short for the lightest chargino, $\tilde{\chi}_1^{\pm}$, where "C1C1, (200, 150) GeV" denotes the chargino-pair production signal where the mass of the lightest chargino is $m_{\tilde{\chi}_1^{\pm}} = 200\,\text{GeV}$ and the mass of the lightest neutralino (the LSP) is $m_{\tilde{\chi}_1^0} = 150\,\text{GeV}$.

Figure 6.2: SFOS leptons before cut on $m_{\ell\ell}$. The red lines indicate where the cuts on $m_{\ell\ell}$ are going to be applied, and the arrows point towards the regions that will be kept, i.e., the events between the red lines will be rejected.



(a) e^+e^- (b) $\mu^+\mu^-$

Figure 6.3: SFOS leptons after cut on $m_{\ell\ell}$ has been applied. The red lines indicate that we are going to make a cut at $E_{\mathrm{T}}^{\mathrm{miss}} = 40\,\mathrm{GeV}$, and the arrow points towards the region (to the right of the line) that will be kept, i.e., we keep events with $E_{\mathrm{T}}^{\mathrm{miss}} > 40\,\mathrm{GeV}$.

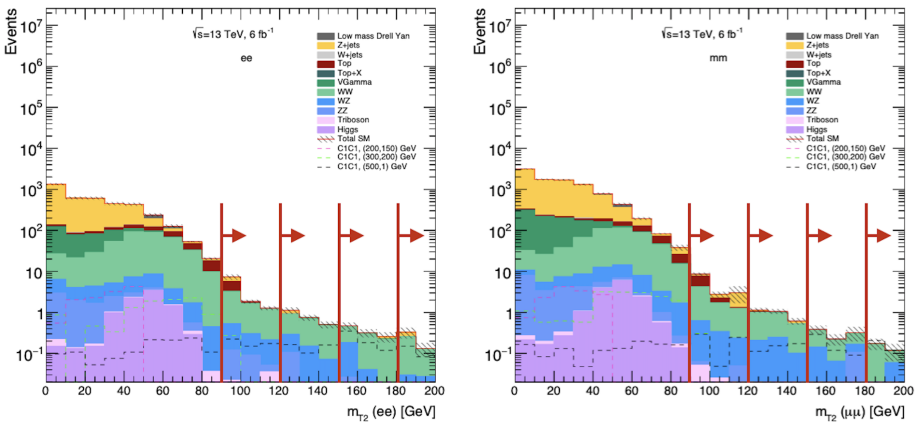


(a) e^+e^- (b) $\mu^+\mu^-$

Figure 6.4: SFOS leptons after the cuts on $m_{\ell\ell}$ and $E_{\mathrm{T}}^{\mathrm{miss}}$ have been applied. The red lines and arrows indicate that we are only going to keep events with no signal jets.

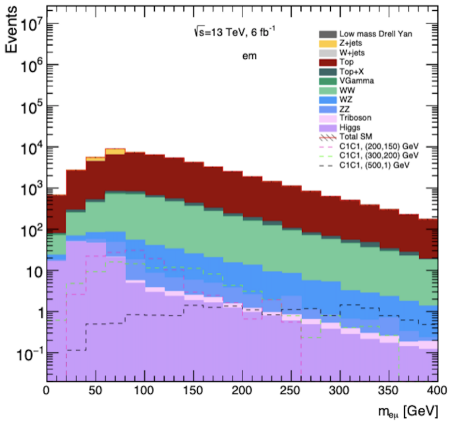
Table 6.1: Signal region definitions that target $\tilde{\chi}_1^{\pm}\tilde{\chi}_1^{\mp}$ production with decays via sleptons to final-states with two opposite-sign (OS) leptons and missing E_{T}. There are separate selections for same-flavor (SF) and different-flavor (DF) lepton events, i.e., $e^+e^-/\mu^+\mu^-$ and $e^+\mu^-/\mu^+e^-$, respectively.

	SR2ℓ	
leptons (e, μ)	SFOS	DFOS
$\boldsymbol{\lvert m_{\ell\ell} - m_Z \rvert}$ **[GeV]**	> 10	–
$\boldsymbol{E_{\mathrm{T}}^{\mathrm{miss}}}$ **[GeV]**	> 40	
central light jets	0	
$\boldsymbol{m_{\mathrm{T2}}}$ **[GeV]**	$> 90, 120, 150, 180$	

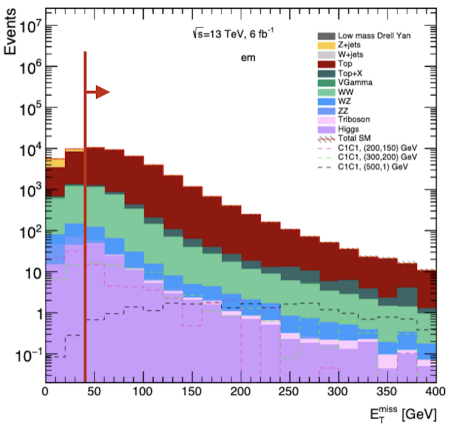


(a) e^+e^- (b) $\mu^+\mu^-$

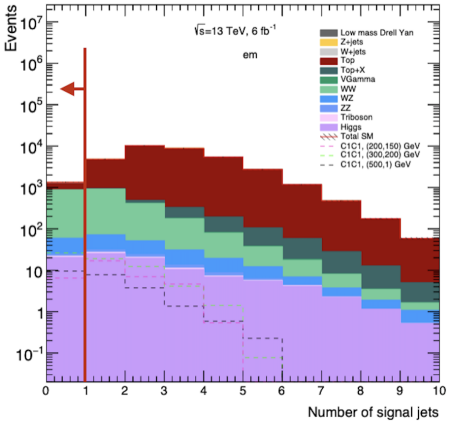
Figure 6.5: SFOS leptons after the cuts on $m_{\ell\ell}$, $E_{\mathrm{T}}^{\mathrm{miss}}$ and the signal jet veto have been applied. The red lines and arrows indicate the four cuts on m_{T2}, which make up four separate signal regions: (i) $m_{\mathrm{T2}} > 90\,\mathrm{GeV}$, (ii) $m_{\mathrm{T2}} > 120\,\mathrm{GeV}$, (iii) $m_{\mathrm{T2}} > 150\,\mathrm{GeV}$ and (iv) $m_{\mathrm{T2}} > 180\,\mathrm{GeV}$. The four SRs are not mutually exclusive, and cannot be statistically combined.



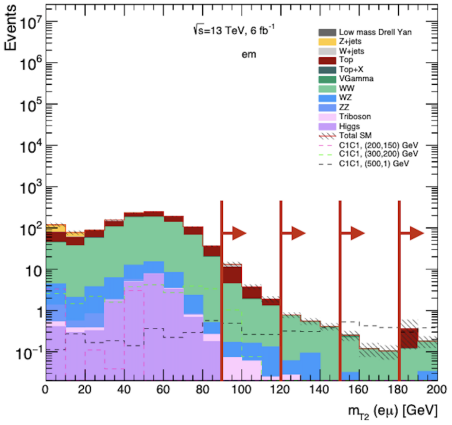
(a) No cut on $m_{\ell\ell}$ for DF events, because the Z-boson does not decay to DF leptons



(b) Before appliying the cut $E_{\mathrm{T}}^{\mathrm{miss}} > 40\,\mathrm{GeV}$



(c) After the cut on $E_{\mathrm{T}}^{\mathrm{miss}}$, but before the cut on the number of signal jets, which is a veto on all signal jets



(d) After the cuts on $E_{\mathrm{T}}^{\mathrm{miss}}$ and the number of signal jets, but before the cuts on m_{T2}

Figure 6.6: Distributions of events with DFOS leptons, with the value and direction of each cut indicated in the figures, in the order the cuts are applied (from top left to bottom right). The final cuts, applied to m_{T2} in (d), define the four separate signal regions for DFOS events: (i) $m_{\mathrm{T2}} > 90\,\mathrm{GeV}$, (ii) $m_{\mathrm{T2}} > 120\,\mathrm{GeV}$, (iii) $m_{\mathrm{T2}} > 150\,\mathrm{GeV}$ and (iv) $m_{\mathrm{T2}} > 180\,\mathrm{GeV}$. The four SRs are not mutually exclusive, and cannot be statistically combined.

the z-axis, in the mass-plane of $\tilde{\chi}_1^{\pm}$ on the x-axis and $\tilde{\chi}_1^0$ on the y-axis, for all the simulated signal mass hypotheses that has a non-zero expected significance. There is one significance plot for each of the four m_{T2}-regions listed in Table 6.1, and plotted for SFOS leptons in Figure 6.5 and for DFOS leptons in Figure 6.6d. For the calculation of the Z_{N}-significances (see Section 5.2.2 for more details), a 30% flat systematic uncertainty has been assumed on the background yield. A specific $\tilde{\chi}_1^{\pm}$,$\tilde{\chi}_1^0$ mass hypothesis can be excluded at 95% confidence level (CL) if the significance exceeds a value of 1.64. This would correspond to a test-statistic that falls in the most extreme 5% of a one-tailed Gaussian distribution, that can be generated from repeated (pseudo-)experiments of the signal-plus-background hypothesis. We can therefore put a lower limit on the $\tilde{\chi}_1^{\pm}$,$\tilde{\chi}_1^0$-masses by tracing a line along the 1.64 contour in these plots, and interpret this line as the expected lower limit on the $\tilde{\chi}_1^{\pm}$,$\tilde{\chi}_1^0$-masses at 95% CL. The lower-limit bounds on these masses obtained by ATLAS at 8 TeV [159], are indicated in the plots as a curved, dashed line. Mass points lying inside of this contour have therefore already been ruled out for this signal model.

Already with $6\,\mathrm{fb}^{-1}$ of integrated luminosity, as shown in Figure 6.7, each of the points lying inside of the previous exclusion contour reach well beyond expected exclusion significance in at least one of the four m_{T2}-regions. Going beyond the already excluded mass-points, the $m_{\mathrm{T2}} > 90\,\mathrm{GeV}$ region shows some sensitivity to the low mass-splitting (150, 50) GeV mass-point ($\Delta m = 100\,\mathrm{GeV}$), with a significance of 0.4, which is still quite far from exclusion sensitivity. The $m_{\mathrm{T2}} > 120\,\mathrm{GeV}$ regions shows exclusion sensitivity to the (500, 1) GeV mass-point, while the $m_{\mathrm{T2}} > 150, 180\,\mathrm{GeV}$ regions are able to increase the sensitivity to the higher $\tilde{\chi}_1^{\pm}$-masses. The $m_{\mathrm{T2}} > 150\,\mathrm{GeV}$ is expected to exclude all the three mass-points with a $\tilde{\chi}_1^{\pm}$-mass of 500 GeV, while the $m_{\mathrm{T2}} > 180\,\mathrm{GeV}$ region gains even more sensitivity to the higher-mass-splitting scenarios ($\Delta m \geq 400\,\mathrm{GeV}$), at the cost of becoming less sensitive to the lower and medium mass-splitting cases ($\Delta m < 400\,\mathrm{GeV}$).

Moving on to the higher-luminosity scenario with $10\,\mathrm{fb}^{-1}$, shown in Figure 6.8, the sensitivity has increased slightly across the board. Most notably, the $m_{\mathrm{T2}} > 180\,\mathrm{GeV}$ region has reached expected exclusion sensitivity to $\tilde{\chi}_1^{\pm}$-masses of 600 GeV, namely the (600, 1) and (600, 100) GeV points.

The conclusion was thereby that we already expected to improve these limits with $6\,\mathrm{fb}^{-1}$ of integrated luminosity, which was a conservative estimate of the amount of data we expected to have in time for the summer conferences in 2016. The improvements in sensitivity are due to the increase in center-of-mass energy, up from 8 TeV in Run 1 to 13 TeV in Run 2. These results therefore supported the plan to target this signal model in the first Run 2 publication.

6.4 First results with Run 2 data

By the summer of 2016, ATLAS had recorded about $13\,\mathrm{fb}^{-1}$ of integrated luminosity from proton-proton collisions, which also passed the data quality requirements necessary for use in physics analysis.

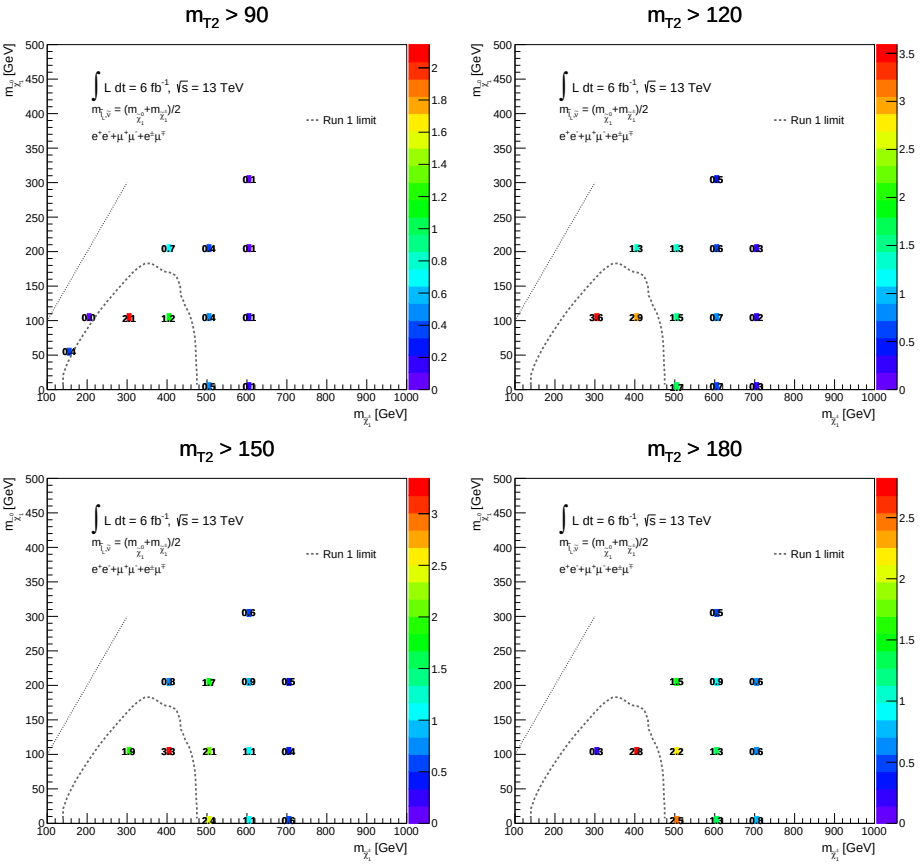


Figure 6.7: Expected sensitivity of SR2ℓ with $6\,\mathrm{fb}^{-1}$, where the z-axis shows the expected Z-value in terms of a one-tailed Gaussian standard deviation. Only the mass-points with a Z-value greater than 0 are shown in the plots. The straight, dotted line indicates the line where $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_1^0$ have equal masses, whereas the curved, dashed line indicates the exclusion limit obtained by ATLAS in the LHC Run 1 at 8 TeV. Mass-points with a $Z \geq 1.6$ are expected to be excluded with the given integrated luminosity.

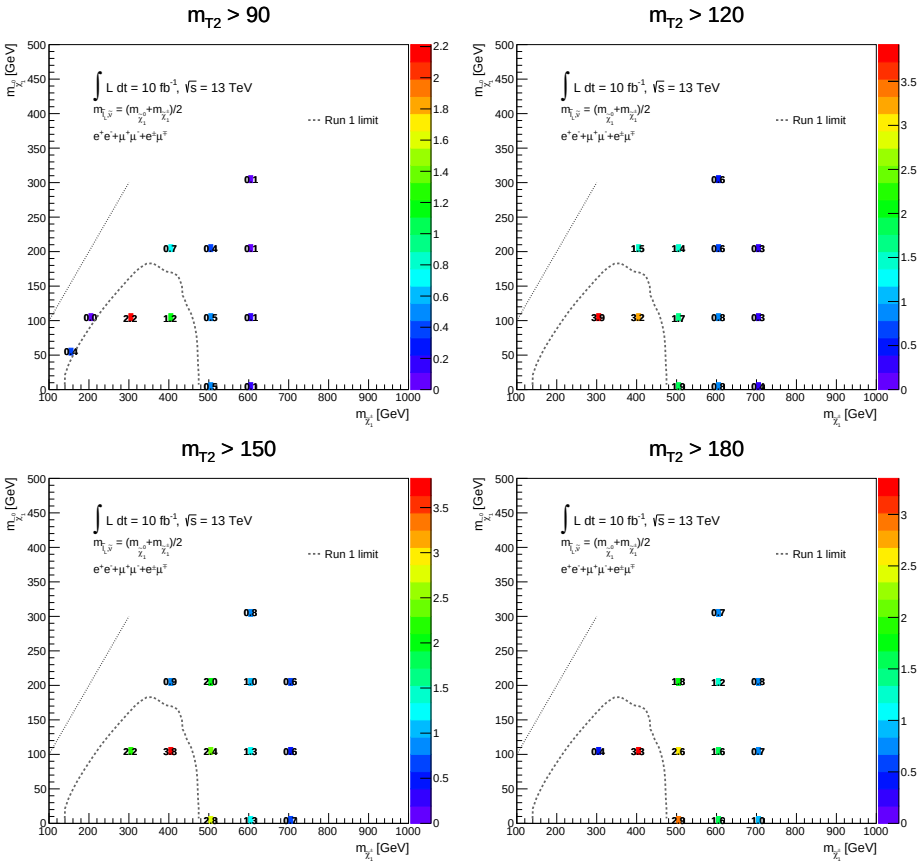


Figure 6.8: Expected sensitivity of SR2ℓ with $10\,\mathrm{fb}^{-1}$, where the z-axis shows the expected Z-value in terms of a one-tailed Gaussian standard deviation. Only the mass-points with a Z-value greater than 0 are shown in the plots. The straight, dotted line indicates the line where $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_1^0$ have equal masses, whereas the curved, dashed line indicates the exclusion limit obtained by ATLAS in the LHC Run 1 at 8 TeV. Mass-points with a $Z \geq 1.6$ are expected to be excluded with the given integrated luminosity.

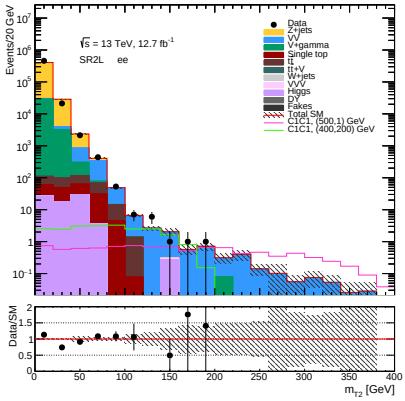
6.4.1 Preliminary results

Figure 6.9 shows three plots of m_{T2} in the signal region SR2ℓ of Table 6.1, with a preliminary dataset of 12.7 fb^{-1}. The figure is split into subfigures according to the three lepton flavor combinations (a) $e^{\pm}e^{\mp}$, (b) $\mu^{\pm}\mu^{\mp}$ and (c) $e^{\pm}\mu^{\mp}$, with all cuts applied except for those on m_{T2}. All backgrounds are taken directly from MC, except for the contribution from fake and non-prompt leptons, which was derived by use of the data-driven Matrix Method (see Section 4.4). No fits to data in control regions have been carried out for these plots, and only statistical uncertainties are taken into account in the error bars for the data points and in the hatched areas, where the latter shows the total uncertainty of all the stacked background sources. The data is found to be consistent with the background, as no significant excess is seen above the SM prediction. Simulated signal models with masses of (400, 200) and (500, 1) GeV have also been overlaid in order to visualize the expected contribution from these scenarios, if they were realized in nature.

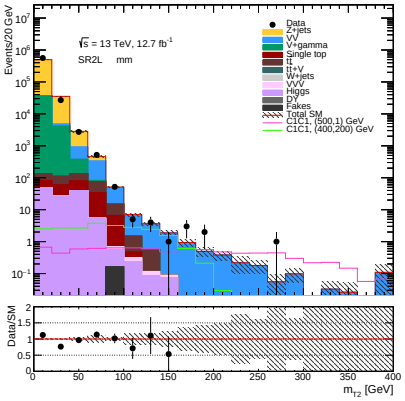
6.4.2 Published results

Corresponding m_{T2}-distributions for the full analysis are split into SF and DF events and shown in Figure 6.10a and 6.10b, respectively. The number of estimated background events before (*MC exp.*) and after the simultaneous fit to data in the CRs and SRs (*Fitted*), are along with the number of observed events shown for each SR in Table 6.2. The SRs are denoted SR2ℓA, -B and -C, according to their stransverse mass cuts, $m_{\mathrm{T2}} > 90, 120$ and 150 GeV, respectively. The total background is further split into five categories: diboson processes with same flavor and different flavor lepton final-states, denoted *VVSF* and *VVDF*, respectively, processes involving top quarks such as tW, $t\bar{t}$ and $t\bar{t}V$, denoted *Top Quark*, the remaining smaller background processes taken directly from MC (not fitted to the data) such as Higgs, $Z+$jets, $V\gamma$ and triboson, denoted *Others*, and finally, the data-driven estimates of fake and non-prompt lepton backgrounds, denoted as *Reducible*. The plots in Figure 6.10 and the yields Table 6.2 include both statistical and systematic uncertainties on the expected/fitted number of background events in the SRs.

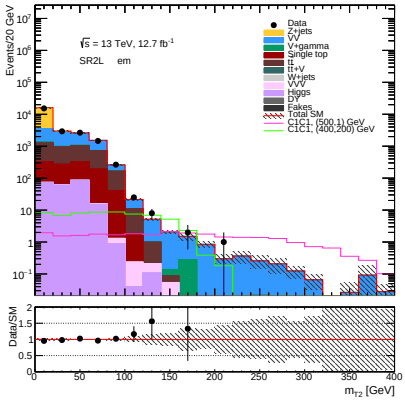
The dominant systematic uncertainties affecting the number of events are listed for each SR in Table 6.3. The statistical uncertainties arise from the finite number of events in the MC samples used to derive the background predictions, and range from 5% in SR2ℓA (DF) to 25% in SR2ℓC (DF), relative the number of background events in the respective SRs. The jet energy scale (JES) and resolution (JER) are the dominant experimental uncertainties, and range from 3% in SR2ℓC (DF) to 18-28% in SR2ℓC (SF). Other experimental uncertainties, such as those related to measurements and modeling of leptons, $E_{\mathrm{T}}^{\mathrm{miss}}$, triggers and pile-up, are found to be negligible. Theory uncertainties related to the renormalization, factorization and resummation scales used to generate the diboson MC samples also have a significant impact on this analysis, ranging from 15% in SR2ℓA (DF) to 66% in SR2ℓC (DF). Uncertainties have been assigned



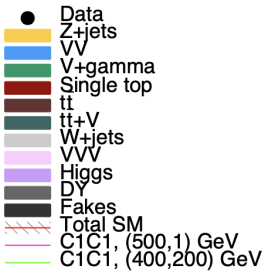
(a) $e^{\pm}e^{\mp}$



(b) $\mu^{\pm}\mu^{\mp}$

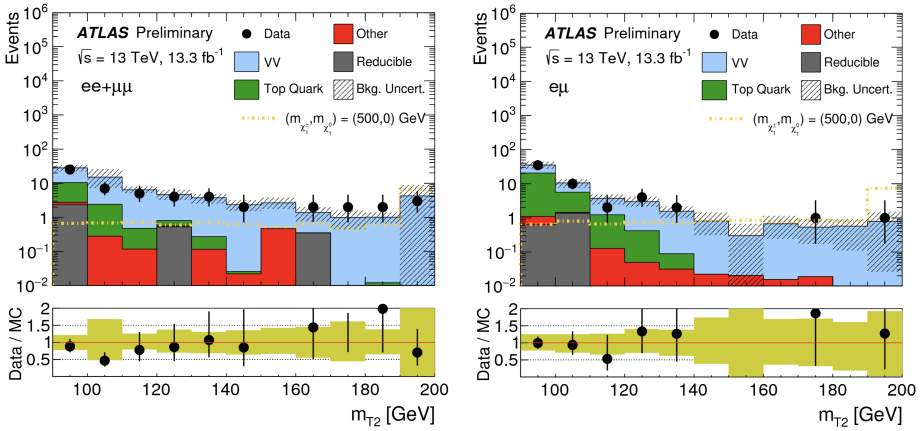


(c) $e^{\pm}\mu^{\mp}$



(d) Enlarged legend. C1 is short for the lightest chargino, $\tilde{\chi}_1^{\pm}$, where "C1C1, (500, 1) GeV" denotes the chargino-pair production signal where the mass of the lightest chargino is $m_{\tilde{\chi}_1^{\pm}} = 500\,\text{GeV}$ and the mass of the lightest neutralino (the LSP) is $m_{\tilde{\chi}_1^0} = 1\,\text{GeV}$.

Figure 6.9: m_{T2}-distributions of data compared to background in SR2ℓ with $12.7\,\text{fb}^{-1}$, where all cuts except the ones on m_{T2} have been applied. The plots show the distributions of opposite sign (a) dielectron, (b) dimuon and (c) electron-muon events. The pink line shows the expected contribution from the simulated signal with $m_{\tilde{\chi}_1^{\pm}} = 500\,\text{GeV}$ and $m_{\tilde{\chi}_1^0} = 1\,\text{GeV}$ ($\Delta m \approx 500\,\text{GeV}$), and correspondingly the green line which shows contributions from the signal with $m_{\tilde{\chi}_1^{\pm}} = 400\,\text{GeV}$ and $m_{\tilde{\chi}_1^0} = 200\,\text{GeV}$ ($\Delta m = 200\,\text{GeV}$). The bottom plot shows the ratio of data over the sum of stacked backgrounds in each histogram bin. In both the top and the bottom plot, the error bars on the data points and the hatched area, which represents the total uncertainty on the stacked backgrounds, correspond to statistical uncertainties only.



(a) SF events in SR2ℓ

(b) DF events in SR2ℓ

Figure 6.10: The m_{T2}-distributions of (a) SF and (b) DF events in the signal region SR2ℓ, with 13.3 fb^{-1}. Diboson backgrounds (VV) have been fitted and normalized to data in dedicated CRs. The background uncertainties, shown as hatched areas in the top plot and in gold areas in the bottom plot, include both statistical and systematic uncertainties. From Reference [158].

to the data-driven background estimates for the fake and non-prompt lepton backgrounds as well. A 30% systematic uncertainty is assigned to account for potential differences in the light-flavor, heavy-flavor and conversion probabilities between the CRs and SRS, and another 15% uncertainty to account for possible pile-up dependence. Even though the modeling uncertainties on the reducible background estimates are large, the total contribution from fakes are so small in the SRs that the effect on the total background uncertainties are negligible, nonetheless.

6.4.3 Exclusion limits

Since there are no significant deviations from the SM expectations found in the SRs, we can proceed to set model-independent exclusion limits on any form of BSM signal above the SM backgrounds, given the observed data. Table 6.4 lists the observed exclusion limits on the visible signal cross-section, $\langle \epsilon\sigma \rangle^{95}_{\mathrm{obs}}$, and the number of signal events S^{95}_{obs} in each of the SRs, at 95% CL. The number of signal events that we expected to exclude at 95% with 13.3 fb^{-1}, in addition to the p-value under the background-only hypothesis, CL_B, are also given in the table. The largest deviation from the background-only hypothesis is observed in SR2ℓA, with CL_B, i.e., only 17% of repeated experiments are expected to give a more extreme value.

We also want to set model-dependent limits in the context of the simplified SUSY model introduced in Section 6.1. In Figure 6.11 the expected exclusion

Table 6.2: Yield tables for the observed and expected number of events in the various SRs. For the backgrounds, the expected event yields are shown both before (MC exp.) and after a simultaneous fit of the diboson backgrounds to data in SRs and dedicated CRs (Fitted). Both statistical and systematic uncertainties are included in the table. From Reference [158].

Region	SR2ℓA (SF)	SR2ℓA (DF)	SR2ℓB (SF)	SR2ℓB (DF)	SR2ℓC (SF)	SR2ℓC (DF)
Observed events	56	55	19	8	9	2
Fitted bkg events	70 ± 12	57.6 ± 8.5	20.7 ± 5.0	8.5 ± 3.6	10.2 ± 3.3	3.1 ± 2.3
Fitted VVSF events	56 ± 11	$--$	19.1 ± 5.0	$--$	9.5 ± 3.2	$--$
Fitted VVDF events	$--$	30.1 ± 7.9	$--$	7.8 ± 3.6	$--$	3.0 ± 2.3
Fitted Top Quark events	10.6 ± 4.0	25.1 ± 8.7	0.42 ± 0.16	0.44 ± 0.24	0.01 ± 0.00	0.01 ± 0.01
Fitted Others events	1.51 ± 0.46	0.90 ± 0.08	0.71 ± 0.29	0.23 ± 0.03	0.53 ± 0.26	0.13 ± 0.03
Fitted Reducible events	2.32 ± 0.80	1.53 ± 0.52	0.48 ± 0.17	0.00 ± 0.29	0.16 ± 0.06	0.00 ± 0.23
MC exp. SM events	56 ± 14	56 ± 12	15.8 ± 5.0	8.1 ± 3.3	7.7 ± 2.9	3.0 ± 2.1
MC exp. VVSF events	41.2 ± 10.2	$--$	14.2 ± 4.7	$--$	7.0 ± 2.8	$--$
MC exp. VVDF events	$--$	28.7 ± 6.3	$--$	7.4 ± 3.3	$--$	2.8 ± 2.1
MC exp. Top Quark events	10.6 ± 4.1	25.1 ± 8.9	0.42 ± 0.17	0.44 ± 0.24	0.01 ± 0.00	0.01 ± 0.01
MC exp. Others events	1.51 ± 0.48	0.90 ± 0.09	0.71 ± 0.30	0.23 ± 0.03	0.53 ± 0.27	0.13 ± 0.03
MC exp. Reducible events	2.32 ± 0.80	1.53 ± 0.52	0.48 ± 0.17	0.00 ± 0.29	0.16 ± 0.06	0.00 ± 0.23

Table 6.3: Systematic uncertainties in the various SR2ℓ regions. From Reference [158].

Systematic	SR2ℓA (SF)	SR2ℓA (DF)	SR2ℓB (SF)	SR2ℓB (DF)	SR2ℓC (SF)	SR2ℓC (DF)
Statistical uncertainty	8%	5%	9%	12%	12%	25%
Jet Energy Scale/Resolution	13-23%	12%	16-26%	3-8%	18-28%	3%
Theory	15%	21%	21%	39%	28%	66%
Reducible closure	$< 1\%$	$< 1\%$	$< 1\%$	–%	$< 1\%$	–%

Table 6.4: Table showing the model-independent exclusion limits, for each SR, on the observed visible signal cross-section $\langle \epsilon\sigma \rangle_{\text{obs}}^{95}$, and the observed and expected limit on the number of signal events, S_{obs}^{95} and S_{exp}^{95}, respectively, all at 95% CL. The CL_B, i.e., the one-tailed Gaussian p-value under the background-only hypothesis, is also shown. From Reference [158].

Signal Region	$\langle \epsilon\sigma \rangle_{\text{obs}}^{95}$[fb]	S_{obs}^{95}	S_{exp}^{95}	CL_B
SR2ℓ-A	1.89	25.1	35^{+13}_{-10}	0.17
SR2ℓ-B	1.24	16.5	$17.0^{+7.1}_{-2.8}$	0.41
SR2ℓ-C	0.87	11.6	$12.7^{+3.0}_{-4.1}$	0.36

contour from $10\,\text{fb}^{-1}$ in the $m_{\text{T2}} > 150\,\text{GeV}$ region is compared with the observed exclusion contour from $13\,\text{fb}^{-1}$ publication [158], both taken at 95% CL. The expected exclusion plot in Figure 6.11 has been obtained by use of the HistFitter package to calculate the expected significance for each signal point and to interpolate between them, in order to get smoother contours. These estimates are slightly more optimistic than those obtained in the previous and simpler

significance plot for the same m_{T2}-region. The observed exclusion contour from the publication is obtained by the same signal regions as described in section 6.3, except the additional veto on forward jets, there is no requirement on $E_{\mathrm{T}}^{\mathrm{miss}}$, and the $m_{\mathrm{T2}} > 180\,\mathrm{GeV}$ region is not part of the analysis.

In the published analysis, the expected significance for each mass-point is calculated for each of the three m_{T2} SR-cuts, and the SR with the highest expected significance is chosen for that point. Figure 8 in the publication shows which region gave the best significance for each mass-point. The $m_{\mathrm{T2}} > 90\,\mathrm{GeV}$ region (called SR2ℓA in the publication) gave highest expected sensitivity to the points with $\Delta m \leq 100\,\mathrm{GeV}$, the region with $m_{\mathrm{T2}} > 120\,\mathrm{GeV}$ (SR2ℓB) gave highest sensitivity for $100\,\mathrm{GeV} \leq \Delta m \leq 300\,\mathrm{GeV}$, whereas the region with $m_{\mathrm{T2}} > 150\,\mathrm{GeV}$ (SR2ℓC) gave the highest expected significance for points with $\Delta m \geq 500\,\mathrm{GeV}$.

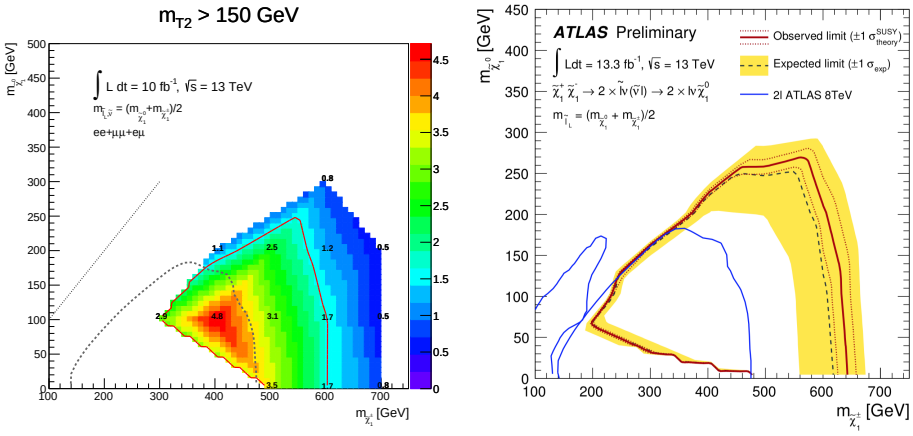
The expected exclusion contour in Figure 6.11a, which only considers the $m_{\mathrm{T2}} > 150\,\mathrm{GeV}$ cut, should then be comparable in the high-mass region of Figure 6.11b, i.e. for high $\tilde{\chi}_1^{\pm}$-mass where the sensitivity is largely limited by the production cross-section of the $\tilde{\chi}_1^{\pm}$. The expected contour from the sensitivity study with $10\,\mathrm{fb}^{-1}$, extends up to $m_{\tilde{\chi}_1^{\pm}}$ of 600 GeV, whereas the expected contour from the full, published analysis with $13.3\,\mathrm{fb}^{-1}$, which includes systematic uncertainties and fits to data in CRs, reaches a slightly higher limit of approximately 625 GeV. The corresponding exclusion reach for $\tilde{\chi}_1^0$-masses are 240 and 270 GeV, respectively. Given that the expected limits of the publication were found using $13.3\,\mathrm{fb}^{-1}$ compared to $10\,\mathrm{fb}^{-1}$ considered in my results, the expected exclusion from the sensitivity study turned out to be in good agreement with the final and more advanced results from the publication.

The CMS collaboration has also published interpretations for the same simplified model, with $35.9\,\mathrm{fb}^{-1}$ of pp collision data collected by the CMS detector [161]. The expected and observed exclusion contours are shown in Figure 6.12. With $35.9\,\mathrm{fb}^{-1}$ of integrated luminosity, they managed to exclude $\tilde{\chi}_1^{\pm}$-masses up to 800 GeV and $\tilde{\chi}_1^0$-masses up to 330 GeV.

6.5 Summary

My sensitivity studies showed that we expected to go beyond the Run 1 sensitivity to chargino-pair production with decays via sleptons into two leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$, already with $6\,\mathrm{fb}^{-1}$ of Run 2 data at 13 TeV. The dataset available for the summer conferences of 2016 amounted to $13.3\,\mathrm{fb}^{-1}$, which significantly improved the exclusion limits in the mass-plane of the lightest neutralino and the lightest chargino compared to the ATLAS limits of Run 1. The lower limit on the mass of the lightest chargino was raised from roughly 475 GeV to 625 GeV for a massless lightest neutralino.

ATLAS has since then published updated results for the same simplified model with the full Run 2 dataset of $139\,\mathrm{fb}^{-1}$ [162]. A plot of m_{T2} in the new signal regions are shown in Figure 6.13a, which make use of binned SRs to take advantage of the m_{T2}-shape of the targeted simplified model. The corresponding



(a) Expected exclusion limits with $m_{\mathrm{T2}} > 150\,\mathrm{GeV}$ and $10\,\mathrm{fb}^{-1}$

(b) Expected and observed exclusion limits with $m_{\mathrm{T2}} > 90, 120, 150\,\mathrm{GeV}$ and $13.3\,\mathrm{fb}^{-1}$

Figure 6.11: (a) Expected exclusion contour with 10 fb^{-1} (red line) compared to (b) the expected and observed exclusion contours with 13.3 fb^{-1} from the publication in Reference [158]. The plot in (a) uses only significances obtained from the $m_{\mathrm{T2}} > 150\,\mathrm{GeV}$ region. The Run 1 limit for non-compressed mass-splitting scenarios is denoted by the curved, dashed line [159], whereas the line of equal masses for $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_1^0$ is denoted with a straight, dotted line. The plot in (b) uses the highest significance found for each mass-point out of the three SRs with $m_{\mathrm{T2}} > 90, 120, 150\,\mathrm{GeV}$. The Run 1 limits for compressed mass-splitting searches, which make use of initial state radiation (ISR) jets to boost low-p_{T} leptons, are highlighted by the leftmost blue, curved line [160], whereas the rightmost blue line illustrates the Run 1 limits for non-compressed searches [159].

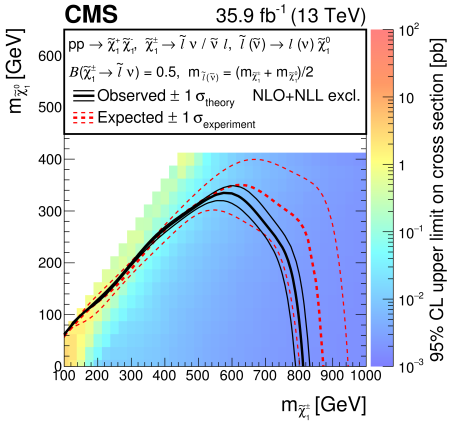
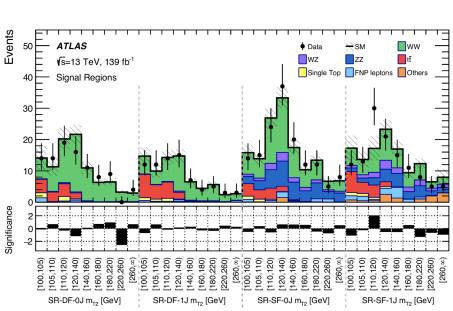
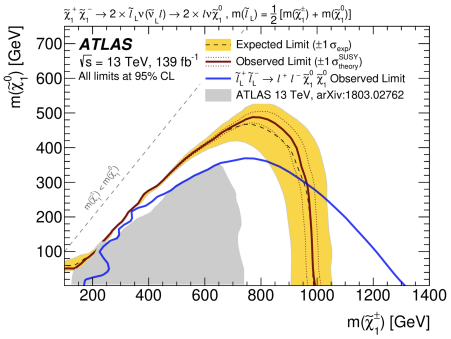


Figure 6.12: Expected and observed exclusion contours with 35.9 fb^{-1} from the CMS collaboration. The figure is taken from Reference [161].



(a) Signal regions binned in m_{T2}



(b) Exclusion limits

Figure 6.13: The signal regions split into all four combinations of SF/DF and 0/1 jet selections, each of them binned in m_{T2}, are shown in (a), and the corresponding exclusion limits obtained from them, with the full Run 2 dataset of ATLAS, are shown in (b). From Reference [162].

expected and observed exclusion limits with $139\,\mathrm{fb}^{-1}$ are shown in Figure 6.13b, where the lower limit on the lightest chargino is increased all the way up to roughly 1 TeV for a massless lightest neutralino.

In the fall of 2016, I moved on to work on a new final-state, namely two leptons and jets in addition to missing transverse energy. In particular, the search for this final-state is motivated by pair-production of the lightest chargino along with the next-to-lightest neutralino, with decays via the SM gauge bosons W and Z. Chapter 7 presents the work I did on signal region optimization, along with the published ATLAS results for this final-state with the full pp-collision dataset of 2015+2016, which amounted to $36.1\,\mathrm{fb}^{-1}$.

Chapter 7

Search for $\tilde{\chi}_1^{\pm}\ \tilde{\chi}_2^0$ production: Data from 2015 and 2016

At 06:02 in the morning on December 5, the LHC experiments recorded the last collision events of 2016 – the end of a remarkably successful year for the LHC [163]. From April to the end of October, ATLAS and CMS recorded about $40\,\text{fb}^{-1}$ of integrated luminosity, which amounts to 60% more data than the original expectation of $25\,\text{fb}^{-1}$. The projected and recorded luminosities are shown as functions of time in Figure 7.1. In total, all four LHC experiments recorded roughly 6.5 million billion (6.5×10^{15}) collisions during this period. This became possible because of the extraordinary availability of the LHC and its injectors, which meant that the LHC was in "collision mode" almost 50% of the time.

In the final four weeks of the 2016 data-taking, the LHC switched collision mode from colliding protons on protons to colliding protons on lead ions. This took place with the record-breaking center-of-mass energy of 8.16 TeV for asymmetric collisions.

Out of the $40\,\text{fb}^{-1}$ *pp*-dataset recorded by ATLAS, $36.1\,\text{fb}^{-1}$ were in the end found to pass all the data quality requirements needed in order to be used for physics analysis. The following chapter describes the search for pair-production of the lightest chargino, $\tilde{\chi}_1^{\pm}$, along with the next-to-lightest neutralino, $\tilde{\chi}_2^0$, using the full 2015 and 2016 dataset.

7.1 Signal model

For the full 2015 and 2016 dataset, I moved on to consider another electroweak SUSY scenario, namely the pair-production of the lightest chargino, $\tilde{\chi}_1^{\pm}$, along with the next-to-lightest neutralino, $\tilde{\chi}_2^0$. We assume that $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ are mass-degenerate, i.e., $m_{\tilde{\chi}_1^{\pm}} = m_{\tilde{\chi}_2^0}$, and that the lightest neutralino, $\tilde{\chi}_1^0$, is the lightest SUSY particle (LSP). Further we assume that $\tilde{\chi}_1^{\pm}$ decays with 100% branching fraction (BF) to a W boson and a $\tilde{\chi}_1^0$, where the W decays hadronically (i.e., into a quark-anti-quark pair), and that $\tilde{\chi}_2^0$ decays with 100% BF to a Z boson and a $\tilde{\chi}_1^0$, where the Z decays leptonically (i.e., into two opposite sign electrons or muons). This simplified model is illustrated in figure 7.2.

The SM gauge bosons are in this chapter further assumed to decay *on-shell*, which means that the mass-difference between $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ and the lightest SUSY particle, $\tilde{\chi}_1^0$, is greater than or equal to the mass of each of the gauge bosons. Since the Z boson is heavier than the W (roughly 91.2 and 80.4 GeV, respectively) the condition for on-shell gauge boson decays therefore becomes $\Delta m(\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0, \tilde{\chi}_1^0) \equiv m_{\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0} - m_{\tilde{\chi}_1^0} \geq m_Z$.

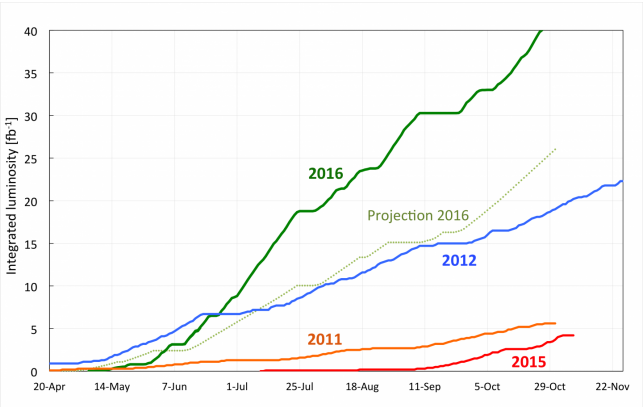


Figure 7.1: The projected luminosity for the 2016 data-taking (dotted, green line) compared to the actual amount recorded in 2016 (solid, green line). The integrated luminosity provided by the LHC in the previous years of Run 1 (2011 and 2012) and Run 2 (2015) are also shown by separate lines. From Reference [163].

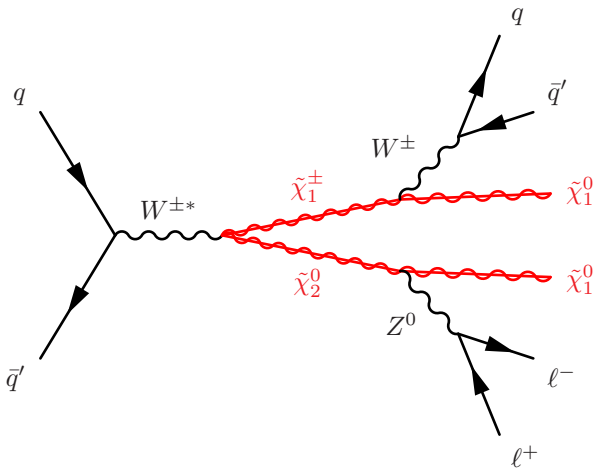


Figure 7.2: Chargino-neutralino production with decays via the W and Z bosons into two leptons, two jets and two lightest neutralinos (LSPs).

As in the $\tilde{\chi}_1^{\pm}\,\tilde{\chi}_1^{\mp}$ case, we assume that R-parity is conserved, so that $\tilde{\chi}_1^0$ is stable and cannot decay further to SM particles. The weakly interacting $\tilde{\chi}_1^0$ therefore escapes detection, and leads to a significant amount of missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, in the event.

7.2 Event and object selection

The object definitions used in the following studies are the same as those described in section 6.2, except that the forward jet veto has been abandoned in favor of using the Jet Vertex Tagger (JVT) to reject pile-up events instead.

7.3 Signal region optimization

The previous ATLAS search for this model was published back in 2014 [159], using $20.3\,\mathrm{fb}^{-1}$ of data from pp-collisions at $\sqrt{s} = 8\,\mathrm{TeV}$. The signal region (SR) called SR-Zjets, given in Table 1 of Reference [159], was used to target the signal model in Figure 7.2, and was also used as a starting point for the studies described in this section.

7.3.1 Updated SR definitions for $\sqrt{s} = 13\,\mathrm{TeV}$

Table 7.1 shows my updated set of cuts for the same model, here called SR2ℓ-jets, that tries to accommodate the higher collision energy and larger statistics obtained in the new dataset. The following cuts have been kept as they were in the 8 TeV analysis: (i) events are required to have exactly two same-flavor (SF) opposite-sign (OS) leptons (electrons or muons), which pass the signal lepton requirements[1]; (ii) at least two non-b-tagged jets measured in the central parts of the detector, with $|\eta| < 2.4$, which pass the signal jet requirements; (iii) no central b-tagged jets; (iv) the dilepton invariant mass, $m_{\ell\ell}$, must be within 10 GeV of the Z-mass, i.e., consistent with an on-shell Z-boson; (v) the dilepton p_{T}, i.e., the p_{T} of the Z, is required to be at least 80 GeV, so that the reconstructed Z has a significant recoil against the rest of the event; and finally, (vi) the anglular separation between the two leptons is required to be in the range $0.3 < \Delta R < 1.5$ (in units of radians), which indicates the two leptons are boosted in a common direction.

The cuts that are modified compared to the Run 1 analysis, are as follows. Instead of an asymmetric W-mass window requirement for the W-candidate jets, of $[50, 100]$ GeV, where $m_W \approx 80.379\,\mathrm{GeV}$, we make it a symmetric $m_W \pm 20\,\mathrm{GeV}$ mass-window. The reason for the wider W-mass window compared to that of the Z-mass, is that the jet energy resolution is larger than the energy and p_{T} resolution of the leptons. The final, and most sensitive variable to this signal

[1] Events with 2 signal leptons and $\geq$ 3 baseline leptons, are rejected, i.e., the events should have exactly 2 baseline leptons, where they both also need to pass the signal lepton requirements.

Table 7.1: Signal region definitions that target $\tilde{\chi}_1^\pm\tilde{\chi}_2^0$-production via hadronically decaying W and a leptonically decaying Z bosons, to final-states with two opposite-sign leptons, at least two jets and $E_{\mathrm{T}}^{\mathrm{miss}}$.

	SR2ℓ-jets
leptons (e, μ)	SFOS
central light jets	≥ 2
central b-jets	0
$\lvert m_{\ell\ell} - m_Z \rvert$ [GeV]	< 10
$\lvert m_{jj} - m_W \rvert$ [GeV]	< 20
$p_{\mathrm{T},\ell\ell}$ [GeV]	> 80
$\Delta R_{\ell\ell}$	$(0.3, 1.5)$
$E_{\mathrm{T}}^{\mathrm{miss,rel}}$ [GeV]	$> 80, 150, 200, 250$

model, is the so-called 'relative $E_{\mathrm{T}}^{\mathrm{miss}}$", or $E_{\mathrm{T}}^{\mathrm{miss,rel}}$, which is defined as

$$E_{\mathrm{T}}^{\mathrm{miss,rel}} = \begin{cases} E_{\mathrm{T}}^{\mathrm{miss}} & \text{if } \Delta\phi_{\ell,j} \geq \pi/2 \\ E_{\mathrm{T}}^{\mathrm{miss}} \times \sin\Delta\phi_{\ell,j} & \text{if } \Delta\phi_{\ell,j} < \pi/2 \end{cases}, \tag{7.1}$$

where $\Delta\phi_{\ell,j}$ is the azimuthal angle $\Delta\phi$ between the $p_{\mathrm{T}}^{\mathrm{miss}}$ (the direction of $E_{\mathrm{T}}^{\mathrm{miss}}$) and the closest signal lepton or jet. The aim of this variable is to suppress the contribution from "fake $E_{\mathrm{T}}^{\mathrm{miss}}$", which arises from significantly mis-measured jets or leptons. In addition to the $E_{\mathrm{T}}^{\mathrm{miss,rel}}$ cut of $> 80\,\mathrm{GeV}$, which was applied in the previous analysis, I added three additional regions with $E_{\mathrm{T}}^{\mathrm{miss,rel}}$ cuts of > 150, 200 and 250 GeV, in order to gain sensitivity to higher $\tilde{\chi}_1^\pm$,$\tilde{\chi}_2^0$-masses.

7.3.2 Optimization of cuts by Z-value

Plots of the five last variables in Table 7.1 are shown in Figure 7.3 and 7.4. The vertical lines indicate where the cuts are applied to that variable, with the arrows pointing towards the region(s) that is (are) kept. The cuts are applied sequentially, so that each plot has the cuts on all preceding variables applied. The distribution of $E_{\mathrm{T}}^{\mathrm{miss,rel}}$, the final variable of SR2ℓ-jets, is shown in Figure 7.4, and has all SR-cuts applied except for a cut on $E_{\mathrm{T}}^{\mathrm{miss,rel}}$. The colors of the horizontal lines and arrows correspond to the color of the signal mass-point they are most relevant for. Three different signal mass-points are superimposed in the plots using dashed lines. These are the $(m_{\tilde{\chi}_1^\pm,\tilde{\chi}_2^0}, m_{\tilde{\chi}_1^0}) = (400, 200)$, (450, 50) and (600, 0) GeV mass-points, shown in orange, purple and turquoise, with mass-splittings $\Delta m = 200$, 400 and 600 GeV, respectively.

The expected significance (Z_{N}, see Section 5.2.2) for each of the three benchmark signal points are shown in the lower panel of Figure 7.3 and 7.4, as a function of lower-cut value. That is, each bin in the lower panel shows the expected significance when all events to the left of that bin, in the upper plot, are rejected, and the number of signal and background events of the bin in question

and all bins to the right are summed up and given as inputs to the calculation of the Z-value.

The Z-values in the lower panels are therefore only helpful for guiding choices on where to place lower-cuts, and are not representative for choosing upper-cuts or ranges (combination of lower- and upper-cuts). Neither of the plots in Figure 7.3 show any sensitivity to signal at all ($Z_{\mathrm{N}} \approx 0$), because the signals are still completely swamped in background. The Z_{N}-values are actually slightly negative in these plots. This is an artefact of the `BinomialExpZ` calculation, which, e.g., calculates a continuous p-value from a binomial distribution with a fractional number of events, using the regularized incomplete beta function. Negative values occur in cases with very few signal events (typically less than one) in combination with a large relative background uncertainty. In the analyses of this thesis, I have set the relative background uncertainty to be 0.3 (30%). In the case of one background event, Z_{N} goes negative when the number of signal events go below roughly 0.4. In the case of 10 background events, Z_{N} goes negative when the number of signal events go below roughly one. In these cases, the `BinomialExpZ` calculation is a bit too conservative with respect to calculations using only integer limits for summing p-values from the binomial distribution. For a more detailed discussion, see Reference [164].

For choosing the cut-ranges of $m_{\ell\ell}$, m_{jj}, $p_{\mathrm{T},\ell\ell}$ and $\Delta R_{\ell\ell}$ in Figure 7.3, the upper panels, i.e., the distributions of the variables themselves, are more helpful for comparing the behavior of the various signal models and the backgrounds. The cuts are chosen so that regions with peaks in the signal distributions are kept, and regions with large background contributions and small contributions from signal are rejected.

The sudden spike in W + jets contributions to the [300, 320) GeV bin of (b) is most likely due to one event with a large event weight. This is known to occur for MC events generated by Sherpa in rare parts of phase space. In any case, this event is rejected by the m_{jj} cut.

Based on Figure 7.3c alone, it appears that it would be beneficial to cut even harder on $p_{\mathrm{T},\ell\ell}$ than the lower-cut at 80 GeV, which is indicated on the plot. However, it was found that the subsequent cuts on $\Delta R_{\ell\ell}$ and $E_{\mathrm{T}}^{\mathrm{miss,rel}}$ manage to reject much of the same background as a harder cut on $p_{\mathrm{T},\ell\ell}$ would, and we settled for 80 GeV, which is consistent with the previous publication from Run 1.

Figure 7.4 shows the $E_{\mathrm{T}}^{\mathrm{miss,rel}}$-distributions of the expected signal and background after all cuts of SR2ℓ-jets have been applied, except for the cuts on $E_{\mathrm{T}}^{\mathrm{miss,rel}}$ itself. By comparison with the distributions of Figure 7.3, it is clear that the preceding cuts have been very effective in reducing the background contributions to the event selection, so that we become sensitive to the various benchmark signal points shown in the plot. We also see that $E_{\mathrm{T}}^{\mathrm{miss,rel}}$ is a very good discriminator for separation of signal and background, where background contributions tend towards lower values and signal tends towards larger values.

We can also notice that the expected significance for each of the signal models peak at different cut-values. The larger the mass-splitting, the higher is the optimal cut. However, some caution should be taken in following these

significance plots too high in cut-value. The reason for this is that a proper significance calculation would require at least three expected signal events in order to exclude that model at 95% CL, given zero observed events. We would also like to have a certain amount of background statistics left in the signal region, typically no less than 1 event, in order to have confidence in the background modeling. Such considerations have not been taken into account in these approximate significance estimates, and is one of the reasons why we do not necessarily choose to cut at the peak of these distributions.

Another consideration to make, when choosing the cut-values, is that we only look at a few benchmark points in the optimization studies, in order to monitor how different mass-hypotheses behave under different cuts. There are, however, many more signal-points that get evaluated using these four, or so, signal regions, and we should try to find a compromise between optimal sensitivity to one specific benchmark point and a good coverage over a collection of mass-hypotheses.

7.3.3 Expected significance with 36.5 fb^{-1}

The approximate expected significance, with the anticipated luminosity of 36.5 fb^{-1}, across the whole signal grid, is shown in each of the four $E_{\text{T}}^{\text{miss,rel}}$ SRs in Figure 7.5. Similarly to the m_{T2}-regions for the chargino-pair production in Figure 6.7 and 6.8, we also here see that the lower $E_{\text{T}}^{\text{miss,rel}}$-cuts are favorable for lower mass-splitting scenarios, while the opposite is true for higher cuts on $E_{\text{T}}^{\text{miss,rel}}$.

The upper left plot of Figure 7.5, shows the Z_{N}-significance for each signal mass point in the SR with $E_{\text{T}}^{\text{miss,rel}} > 80\,\text{GeV}$. Only the signal mass hypothesis with (250, 50) GeV, which reaches $Z_{\text{N}} = 2.2$, has $Z_{\text{N}} > 1.6$, and is thereby expected to be excluded at 95% CL.

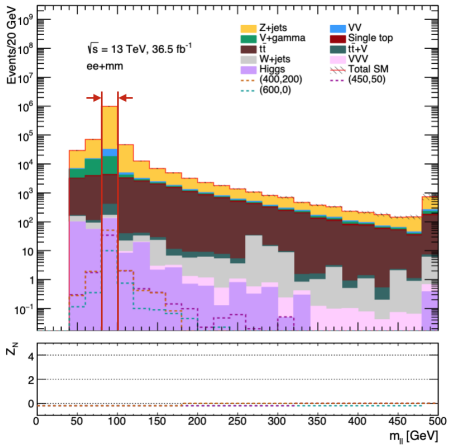
The upper right plot of Figure 7.5, however, which shows the expected significances for the SR with $E_{\text{T}}^{\text{miss,rel}} > 150\,\text{GeV}$, results in expected exclusion for 13 signal mass points, up to $m_{\tilde{\chi}_1^\pm/\tilde{\chi}_2^0} = 500\,\text{GeV}$ and $m_{\tilde{\chi}_1^0} = 150\,\text{GeV}$.

Next, the SR with $E_{\text{T}}^{\text{miss,rel}} > 200\,\text{GeV}$, shown on the lower left, reaches expected exclusion sensitivity for 18 signal mass points, all the way up to $m_{\tilde{\chi}_1^\pm/\tilde{\chi}_2^0} = 550\,\text{GeV}$ and $m_{\tilde{\chi}_1^0} = 200\,\text{GeV}$. It thereby increases the sensitivity to intermediate and high mass-splitting points. However, the sensitivity to the mass points on the $\Delta m = 150 and 200\,\text{GeV}$ diagonal, e.g., the (250, 100) and (250, 50) GeV points, are reduced from $Z_{\text{N}} = 1.1$ to 0 and $Z_{\text{N}} = 3.1$ to 1.2, respectively.

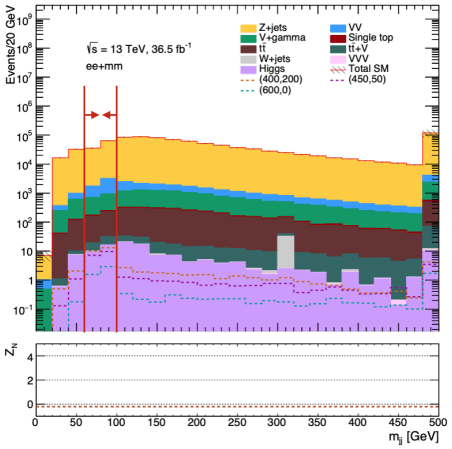
The final SR, with $E_{\text{T}}^{\text{miss,rel}} > 250\,\text{GeV}$, shown on the lower right, continues to improve sensitivity to the high- and the upper part of the intermediate-Δm regions, but less so compared to the increase from $E_{\text{T}}^{\text{miss,rel}} > 150\,\text{GeV}$ to $E_{\text{T}}^{\text{miss,rel}} > 200\,\text{GeV}$. The sensitivity to mass points with $\Delta m \leq 300\,\text{GeV}$ are reduced in this case.

The best performing SRs in the different mass-splitting regions then turned out to be:

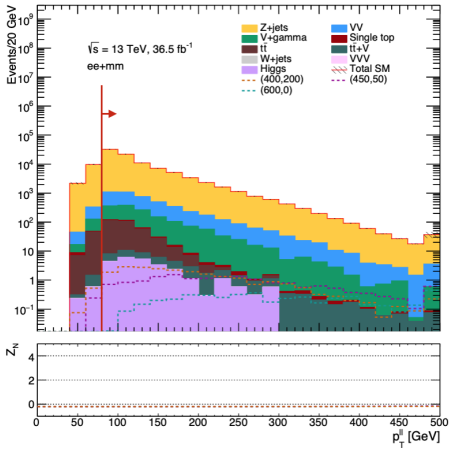
- $E_{\text{T}}^{\text{miss,rel}} > 150\,\text{GeV}$ for $\Delta m \leq 200\,\text{GeV}$,



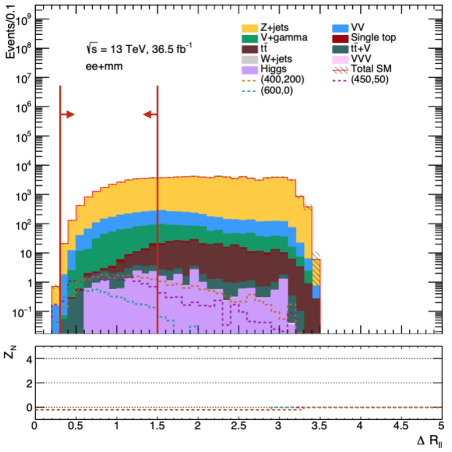
(a) Dilepton invariant mass, $m_{\ell\ell}$



(b) Dijet invariant mass, m_{jj}



(c) Transverse momentum of the dilepton system, $p_{\mathrm{T},\ell\ell}$



(d) Angular separation between the two leptons, $\Delta R_{\ell\ell}$

Figure 7.3: Plots of sequentially applied cuts in SR2ℓ-jets. See Figure 7.4 for an enlarged view of the plot legend.

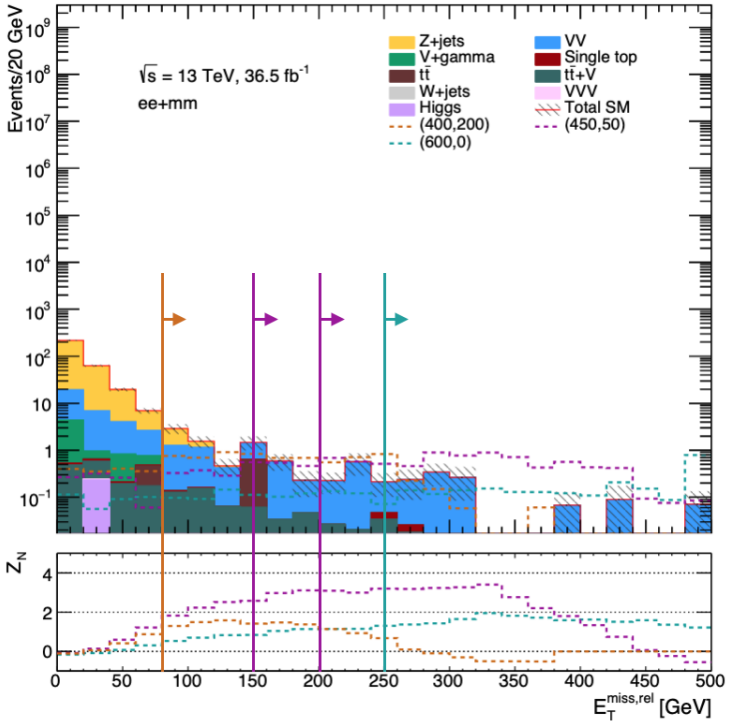


Figure 7.4: Distribution of $E_{\mathrm{T}}^{\mathrm{miss,rel}}$ in SR2ℓ-jets, with all cuts applied except the cut on $E_{\mathrm{T}}^{\mathrm{miss,rel}}$. The vertical lines indicate the four cut-values applied to $E_{\mathrm{T}}^{\mathrm{miss,rel}}$, where the arrows point towards the region being kept, and the colors signify which benchmark signal point they are most relevant for. The benchmark points correspond to the following mass-hypotheses: $(m_{\tilde{\chi}_1^{\pm},\tilde{\chi}_2^0}, m_{\tilde{\chi}_1^0}) = (400, 200)$ GeV in orange, $(450, 50)$ GeV in purple and $(600, 0)$ GeV in turquoise.

- $E_{\mathrm{T}}^{\mathrm{miss,rel}} > 200\,\mathrm{GeV}$ for $250\,\mathrm{GeV} \leq \Delta m \leq 350 - 400\,\mathrm{GeV}$, and
- $E_{\mathrm{T}}^{\mathrm{miss,rel}} > 250\,\mathrm{GeV}$ for $\Delta m \geq 350 - 400\,\mathrm{GeV}$.

The SR with $E_{\mathrm{T}}^{\mathrm{miss,rel}} > 80\,\mathrm{GeV}$ did not achieve better sensitivity than the other SRs for any of the grid points.

In Figure 7.6, the expected significance with $36.5\,\mathrm{fb}^{-1}$ for the whole grid is shown for the most aggressive $E_{\mathrm{T}}^{\mathrm{miss,rel}}$-region, and compared with the observed exclusion limits with $36.1\,\mathrm{fb}^{-1}$ [165]. An expected 95% C.L. exclusion contour in the left plot (not drawn) would correspond to a significance value of 1.6. Looking at the left plot, we expect to exclude $\tilde{\chi}_1^{\pm}$,$\tilde{\chi}_2^0$-masses up to around 550 GeV, given a massless $\tilde{\chi}_1^0$, and $\tilde{\chi}_1^0$-masses up to 200 GeV, given a $\tilde{\chi}_1^{\pm}$,$\tilde{\chi}_2^0$-mass of 500 GeV. This corresponds very well with the expected exclusion contour in the high-mass region of the right plot, where the expected significances have been calculated

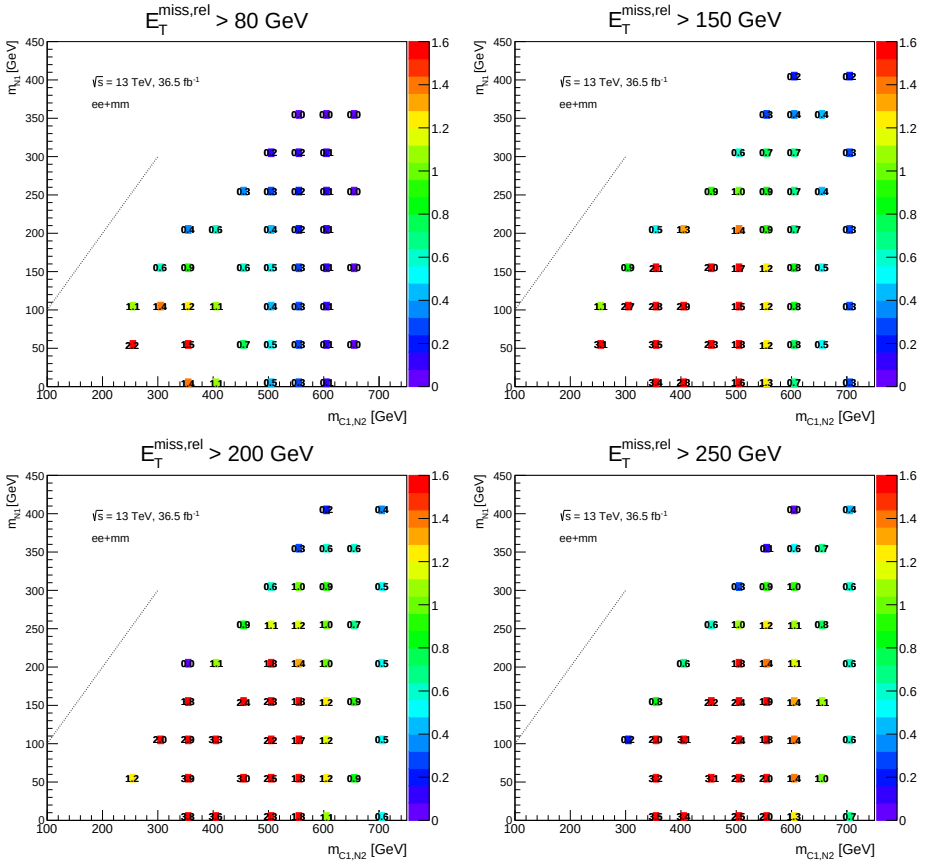
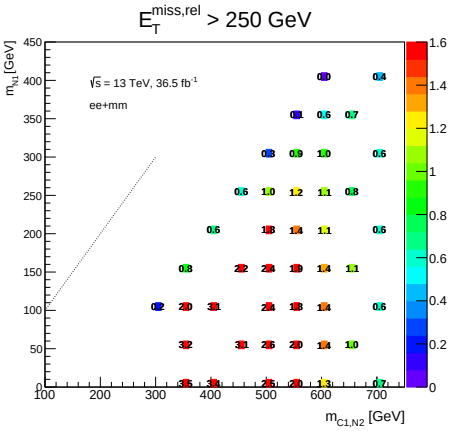
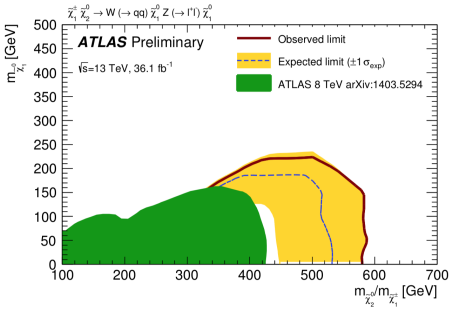


Figure 7.5: Expected exclusion significance, Z_{N}-value, shown on the z/color-axis, in the mass-plane of $m_{\tilde{\chi}_1^\pm,\tilde{\chi}_2^0}$ on the x-axis and $m_{\tilde{\chi}_1^0}$ on the y-axis. The results correspond to $36.5\,\mathrm{fb}^{-1}$ of integrated luminosity, using the SR2ℓ-jets region in Table 7.1, where each plot uses a distinct lower-cut value on $E_{\mathrm{T}}^{\mathrm{miss,rel}}$, as indicated above each plot. The dotted, black lines indicate the lines of equal masses, $m_{\tilde{\chi}_1^\pm,\tilde{\chi}_2^0} = m_{\tilde{\chi}_1^0}$.



(a) SR2ℓ-jets with $E_{\mathrm{T}}^{\mathrm{miss,rel}} > 250\,\mathrm{GeV}$ and $36.5\,\mathrm{fb}^{-1}$



(b) 2ℓ+jets SRs of Ref. [165] with $36.1\,\mathrm{fb}^{-1}$

Figure 7.6: The expected exclusion significance, Z_{N}, in the mass-plane of $m_{\tilde{\chi}_1^\pm,\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$ with $36.5\,\mathrm{fb}^{-1}$, using the SR2ℓ-jets region selection with $E_{\mathrm{T}}^{\mathrm{miss,rel}} > 250\,\mathrm{GeV}$, is shown in (a). The expected and observed exclusion contours at 95% CL with $36.1\,\mathrm{fb}^{-1}$ of data, from Reference [165] (conference paper with preliminary results), are shown in (b). In (a), the dotted, black line indicates the line of equal masses, $m_{\tilde{\chi}_1^\pm,\tilde{\chi}_2^0} = m_{\tilde{\chi}_1^0}$, and the signal points with $Z_{\mathrm{N}} > 1.6$ are expected to be excluded at 95% CL.

using the more thorough and robust CLs-method, including properly estimated systematic uncertainties. The observed exclusion contour covers even higher masses than the expected exclusion contour with $36.1\,\mathrm{fb}^{-1}$. This is because the observed number of events in data was lower than the background expectation.

7.3.4 Published analysis

For the analysis of the 2015+2016 data, there were analysts from two different institutes within the same working group who showed interest in studying the same signal model and final-state, and had slightly different approaches to the problem. I based my studies on the approach taken by the previous search for electroweak production of $\tilde{\chi}_1^\pm$ $\tilde{\chi}_2^0$ in final-states with two leptons and jets. The other institute was also working on searches for production of the same sparticles in strong interactions, with two leptons and typically more hadronic activity compared to the electroweak interactions. Whereas I made use of the $E_{\mathrm{T}}^{\mathrm{miss,rel}}$ variable, which is a combination of $E_{\mathrm{T}}^{\mathrm{miss}}$ and the angular separation, $\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}}, X)$, between the direction of $p_{\mathrm{T}}^{\mathrm{miss}}$ and other objects, X, their approach was to separate these two variables and consider them independently.

In order to harmonize our efforts and end up with one common result, we compared the sensitivities of the two approaches, and concluded that they roughly

achieved equivalent results. We then decided to go for the approach of treating the two quantities separately, for a simpler and more intuitive interpretation of the analysis.

The final SR definitions [166], which were further refined after the harmonization efforts, are listed in Table 7.2. The SR2-int and SR2-high regions, which target intermediate and high mass-splittings, respectively, are quite similar to my initially proposed cuts in Table 7.1. The event topology is similarly selected by requiring events with at least two non-b-tagged jets, and two SFOS leptons and two highest-p_{T} jets with invariant masses close to the Z and W masses, respectively. The p_{T} of the dilepton (Z) system must be $> 80\,\mathrm{GeV}$, where the two leptons have a rather low opening-angle between them of $\Delta R_{\ell\ell} < 1.8$ (which is slightly less constraining compared to $0.3 < \Delta R_{\ell\ell} < 1.5$ in Table 7.1). SR2-int also requires $E_{\mathrm{T}}^{\mathrm{miss}} > 150\,\mathrm{GeV}$ and SR2-high $E_{\mathrm{T}}^{\mathrm{miss}} > 250\,\mathrm{GeV}$, where $E_{\mathrm{T}}^{\mathrm{miss,rel}} \approx E_{\mathrm{T}}^{\mathrm{miss}}$ in the intermediate and high-mass-splitting cases, where the jet mis-measurements and fake $E_{\mathrm{T}}^{\mathrm{miss}}$ contributions are much lower compared to the low-mass-splitting case.

SR2-int and SR2-high additionally introduce a p_{T} cut on the dijet (W) system of $> 100\,\mathrm{GeV}$, a requirement of a rather low angular separation also between the two p_{T}-leading jets of $\Delta R_{jj} < 1.5$, i.e., so that they are boosted in a common direction, and finally requiring the angular separation between the $p_{\mathrm{T}}^{\mathrm{miss}}$ and the W (dijet system) in the transverse plane to be $0.5 < \Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}}, W) < 3.0$.

The most prominent differences between the SRs in Figure 7.7 and Table 7.2 are the selections related to the low-mass-splitting scenarios, where the mass-splitting between $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$ is approximately equal to the mass of the Z boson, and the signal therefore becomes kinematically similar to the SM diboson backgrounds. Here, two sets of orthogonal regions are defined, one for events with exactly two non-b-tagged jets, where both are assumed to originate from the W, called SR2-low-2J, and one region for events with 3-5 non-b-tagged jets, called SR2-low-3J. In the latter region, the two jets which are closest to the $p_{\mathrm{T}}^{\mathrm{miss}}$ in the transverse plane (smallest $\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{jet})$) are assumed to originate from the W boson, and the remaining signal jets of the event are taken to be initial-state-radiation (ISR), which recoil against the pair-produced sparticles and boost the otherwise low-p_{T} final-state particles to higher p_{T}. As the two low-mass-splitting SRs are orthogonal, they are combined in the statistical analysis, but their background contributions are predicted and validated independently.

The dominant source of background in all the SRs is diboson processes, which make up nearly all of the background contributions in the SR2-int and SR2-high regions. In the SR2-low regions, the background composition is more mixed, with larger contributions from other sources such as $Z+\mathrm{jets}$, processes involving top quarks and fake and non-prompt (FNP) leptons.

Given sub-optimal modeling of the fake $E_{\mathrm{T}}^{\mathrm{miss}}$ contributions from $Z+\mathrm{jets}$ processes in MC simulations, this background is instead estimated by a data-driven approach. Using $\gamma+\mathrm{jets}$ events in data to measure the shape of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution, the resulting $E_{\mathrm{T}}^{\mathrm{miss}}$ template is normalized to data in an SR-like control region with $E_{\mathrm{T}}^{\mathrm{miss}} < 100\,\mathrm{GeV}$. See Reference [166] for more details about

Table 7.2: Signal region definitions used for the preliminary conference results in Reference [165] and the final published analysis with 36.1 fb^{-1} in Reference [166].

2ℓ+jets signal region definitions				
	SR2-int	SR2-high	SR2-low-2J	SR2-low-3J
$n_{\text{non-}b\text{-tagged jets}}$	≥ 2		2	3–5
$m_{\ell\ell}$ [GeV]	81–101		81–101	86–96
m_{jj} [GeV]	70–100		70–90	70–90
$E_{\text{T}}^{\text{miss}}$ [GeV]	> 150	> 250	> 100	> 100
p_{T}^{Z} [GeV]	> 80		> 60	> 40
p_{T}^{W} [GeV]	> 100			
m_{T2} [GeV]	> 100			
$\Delta R_{(jj)}$	< 1.5			< 2.2
$\Delta R_{(\ell\ell)}$	< 1.8			
$\Delta\phi_{(p_{\text{T}}^{\text{miss}},Z)}$			< 0.8	
$\Delta\phi_{(p_{\text{T}}^{\text{miss}},W)}$	0.5–3.0		> 1.5	< 2.2
$E_{\text{T}}^{\text{miss}}/p_{\text{T}}^{Z}$			$0.6 - -1.6$	
$E_{\text{T}}^{\text{miss}}/p_{\text{T}}^{W}$			< 0.8	
$\Delta\phi_{(p_{\text{T}}^{\text{miss}},\text{ISR})}$				> 2.4
$\Delta\phi_{(p_{\text{T}}^{\text{miss}},\text{jet1})}$				> 2.6
$E_{\text{T}}^{\text{miss}}/p_{\text{T}}^{\text{ISR}}$				0.4–0.8
$\lvert\eta(Z)\rvert$				< 1.6
$p_{\text{T}}^{\text{jet3}}$ [GeV]				> 30

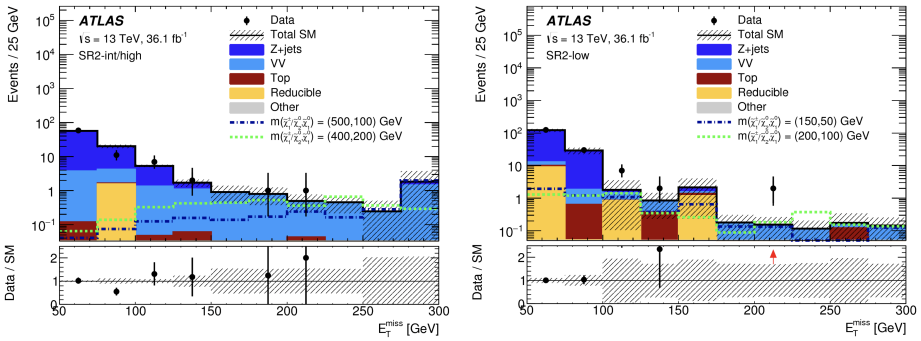
the γ + jets template method.

The contributions from FNP leptons are estimated by the data-driven Matrix Method, whereas the remaining sources of background, including diboson and top quark processes, are taken directly from MC.

In addition to the VRs used to validate the modeling of the Z + jets and FNP leptons backgrounds, a "tight" and "loose" VR selection are defined for each SR, in order to validate the overall modeling of the backgrounds in these SRs. See Reference [166] for more information about the VR definitions. No CRs are used to constrain the backgrounds in the likelihood fit of this analysis, i.e., only SRs are involved in the fit.

The dominant systematic uncertainties in the SRs come from the modeling of diboson events, which range from approximately 30 to 40%, and from uncertainties associated with the data-driven Z + jets estimates, which amount to about 42% in SR2-int, 71% in SR-high and 64% in the combined SR2-low.

Figure 7.7 shows the $E_{\text{T}}^{\text{miss}}$-distributions for events in the four SRs defined in Table 7.2. Among the most noteworthy features in these plots are perhaps the lack of observed events in the SR2-high region with $E_{\text{T}}^{\text{miss}} > 250\,\text{GeV}$, in Figure 7.7a, and the excesses of observed events in the $E_{\text{T}}^{\text{miss}} \in [100, 125)$ and $[200, 225)$ GeV bins of Figure 7.7b.



(a) SR2-int and SR2-high (b) SR2-low-2J and SR2-low-3J

Figure 7.7: Distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ for events in (a) SR2-int and SR2-high, and in (b) SR2-low-2J and SR2-low-3J. From Reference [166].

Table 7.3: The expected number of background events (Total SM), with systematic uncertainties, along with the observed number of events in each of the SRs. The lower part of the table shows the total SM background split into contributions from sub-components. From Reference [166].

SR2-	int	high	low (combined)
Observed	2	0	11
Total SM	$4.1^{+2.6}_{-1.8}$	$1.6^{+1.6}_{-1.1}$	$4.2^{+3.4}_{-1.6}$
VV	4.0 ± 1.8	1.6 ± 1.1	1.7 ± 1.0
Top	0.15 ± 0.11	0.04 ± 0.03	0.8 ± 0.4
FNP	$0.0^{+0.2}_{-0.0}$	$0.0^{+0.1}_{-0.0}$	$0.7^{+1.8}_{-0.7}$
Z + jets	$0.0^{+1.8}_{-0.0}$	$0.0^{+1.2}_{-0.0}$	$1.0^{+2.7}_{-1.0}$
Other	–	–	–

The number of expected background events (Total SM) and the number of observed events in the SRs of Table 7.2 are shown in Table 7.3. The total SM background is also split into the sub-components from dibosons (VV), top quarks, fake and non-prompt lepton sources (FNP), Z + jets and other background processes. As one can see from Figure 7.7a, there are no observed events in SR2-high, where 1.6 was expected. In SR2-low, however, a sizable excess of 11 event are observed, where 4.2 events were expected. This translates to a p-value of 0.06, which corresponds to a $1.6\,\sigma$ (one-tailed Gaussian) excess, as can be seen in Figure 7.8.

In the absence of observed excesses in SR2-int and SR2-high, model-independent upper limits are set on the visible cross-section for any beyond-SM (BSM) signal, $\langle \epsilon\sigma \rangle^{95}_{\mathrm{obs}}$, and on the expected and observed upper limits on the number of BSM signal events, S^{95}_{exp} and S^{95}_{obs}, respectively, all calculated at 95% C.L., which are

Table 7.4: Model-independent limits on beyond-SM signal with 36.1 fb^{-1}. From Reference [166].

Signal channel	Region	N_{obs}	N_{exp}	$\langle\epsilon\sigma\rangle^{95}_{\text{obs}}$[fb]	S^{95}_{obs}	S^{95}_{exp}	$p(s{=}0)$	Z
2ℓ+jets	SR2-int	2	$4.1^{+2.6}_{-1.8}$	0.13	4.5	$5.6^{+2.2}_{-1.4}$	0.5	0
	SR2-high	0	$1.6^{+1.6}_{-1.1}$	0.09	3.1	$3.1^{+1.4}_{-0.1}$	0.5	0
	SR2-low	11	$4.2^{+3.4}_{-1.6}$	0.43	15.7	12^{+4}_{-2}	0.06	1.6

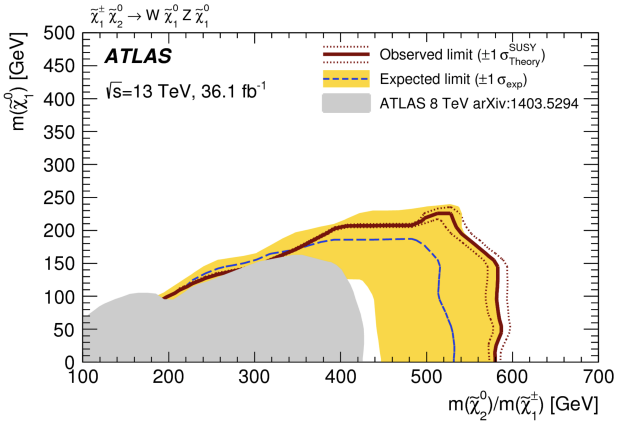


Figure 7.8: Expected and observed exclusion limits on the simplified SUSY model in the mass-plane of $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$, where both the hadronic (2ℓ+jets) and leptonic (3ℓ) decay modes of the W have been combined to produce stronger exclusion limits in the same mass-plane. From Reference [166].

shown in Table 7.4. The p-value under the background-only hypothesis is also shown in the next-to-last column ($p(s=0)$), and the corresponding Z-value is given in the last column. The p-values are cut at 0.5 wherever the observed number is lower than the expected number of events, in order to avoid negative Z-values.

Finally, expected and observed, model-dependent lower limits on the masses of $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$ in the simplified model are illustrated in Figure 7.8. As opposed to the exclusion plot in Figure 7.6b, from the conference paper of Reference [165], which shows the contributions from the 2ℓ+jets channel only, the exclusion contours of Figure 7.8, from the journal publication, are derived from a combination of both the hadronic (2ℓ+jets) and leptonic (3ℓ) decay modes of the W boson. This results in slightly modified expected and observed exclusion contours, where the 3ℓ channel in particular adds sensitivity in the low-mass-splitting region.

Figure 7.9 shows corresponding results from the CMS collaboration for the same simplified model, with 35.9 fb^{-1} of integrated luminosity collected by the

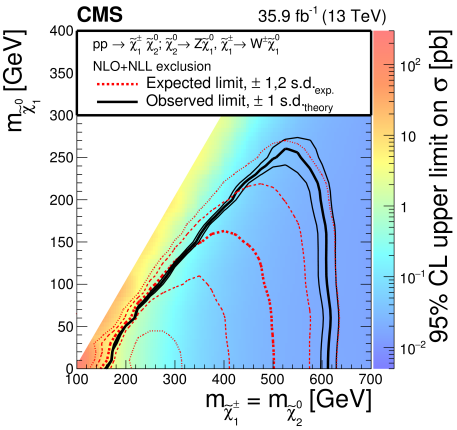


Figure 7.9: Expected and observed exclusion contours with 35.9 fb^{-1} of integrated luminosity from the CMS collaboration. The figure is taken from Reference [167].

CMS detector. The limits are based on the 2ℓ+jets channel only. The expected and observed exclusion contours are consistent with the ATLAS results. Both experiments observe less events than expected in the high-mass-splitting regime, and are thereby able to set stronger exclusion limits than expected.

7.4 Summary

My proposed SR selections for targeting pair-production of $\tilde{\chi}_1^\pm$ $\tilde{\chi}_2^0$, with decays via W and Z bosons into $2\ell + \text{jets} + E_\text{T}^\text{miss}$ in the final state, using the full dataset taken by ATLAS in 2015 and 2016, are listed in Table 7.1 and were shown to be correspondingly sensitive to the simplified signal model as the SRs used for the publications [165, 166], listed in Table 7.2.

The high- and intermediate-mass-splitting regions, SR2-high and SR2-int, yielded less observed events than were expected, and thereby set stronger exclusion limits on the $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$ masses than the expectation. The lower-limit on the mass-degenerate $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ was raised from a mass of 425 GeV with the 8 TeV ATLAS data, up to 580 GeV with the 13 TeV data, for a massless $\tilde{\chi}_1^0$ LSP. Similarly, the 8 TeV results from ATLAS excluded $\tilde{\chi}_1^0$-masses up to 160 GeV for $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$-masses of 350 GeV, which are extended up to 220 GeV for $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$-masses of 500 GeV with 13 TeV data taken by ATLAS. The low-mass-splitting region, SR2-low, on the other hand, resulted in a $1.6\,\sigma$ excess in the observed data compared to the expectation from background contributions.

Similar local excesses, of up to $3.0\,\sigma$, were also observed in the low-mass-splitting regions of Reference [168], which utilized so-called recursive jigsaw reconstruction to target the event topologies of the same simplified SUSY model as presented in this chapter. This caused quite a bit of excitement in SUSY

circles, and anticipation of more data to either confirm the observed excesses as a real signal or rather as a statistical fluctuation.

In the following chapter, Chapter 8, I present the analysis of the full Run 2 dataset of ATLAS, which amounted to $139\,\text{fb}^{-1}$, in search for the same signal model.

Chapter 8

Search for $\tilde{\chi}_1^{\pm}\,\tilde{\chi}_2^0$ production: The complete Run 2 dataset

On December 3, 2018, the final collisions of the LHC Run 2 took place, which ended three and a half years of impressively high performance delivered by the particle accelerators and injectors [169]. The LHC had then provided 160 fb^{-1} of proton-proton collisions since June 2015, where the instantaneous luminosity reached up to $2 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$ during 2017 and 2018, which is twice the beam intensity the LHC was designed to deliver at the interaction points. The LHC had also shown great versatility by colliding protons on lead-ions, lead on lead, and on October 12, 2017, for the first time, fully stripped Xenon-ions were collided, in order to gain more insight into the exotic Quark Gluon Plasma phase that existed in the very early universe.

Figure 8.1 shows the accumulated luminosity per year for the full Run 1 and Run 2. The final year, 2018, provided more integrated luminosity than any of the preceding years, which is a testament to the ever improving performance and the stability of the LHC.

The distributions of the estimated average number of pileup interactions per bunch crossing is shown for each Run 2 year in Figure 8.2. ATLAS recorded collisions with up to 70 interactions per bunch crossing both in 2017 and 2018, due to the LHC team's ability to squeeze the proton beams more than the preceding years of data-taking. The year-averaged mean number of pileup interactions seen by the ATLAS detector was 13.4 in 2015, 25.1 in 2016, 37.8

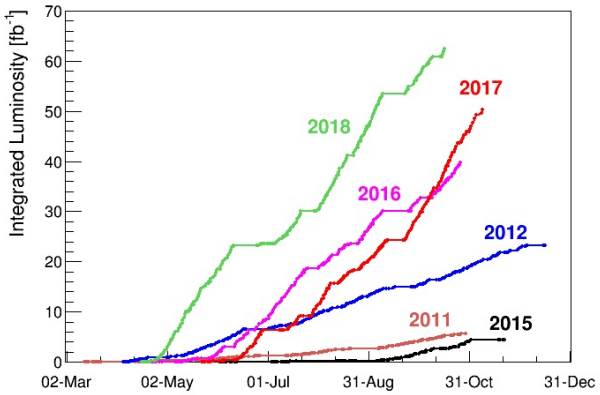


Figure 8.1: Integrated luminosity accumulated per year of the LHC Run 2 (2015 to 2018), as well as the LHC Run 1 (2011 and 2012). From Reference [170].

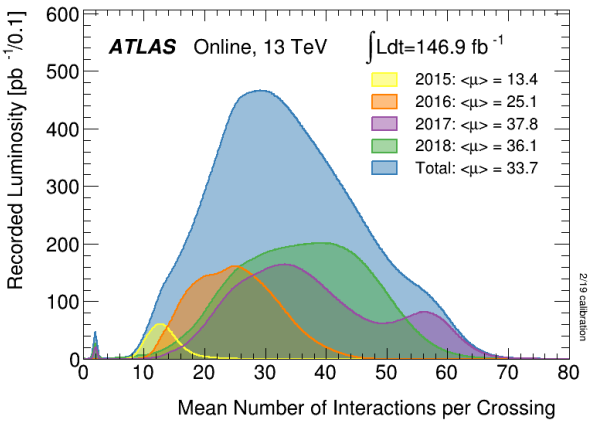


Figure 8.2: Integrated luminosity recorded by ATLAS as a function of the mean number of interactions per bunch crossing (pile-up), split by year of the LHC Run 2. From Reference [171].

in 2017 and 36.1 in 2018. By combining the data collected in the full Run 2 dataset, we get a total average of 33.7 pileup interactions per bunch crossing.

8.1 Signal model and scope of work

For the analysis with the complete Run 2 dataset, I continued working on the same signal model as described in Section 7.1, namely search for pair-production of $\tilde{\chi}_1^{\pm}$ $\tilde{\chi}_2^0$ in final-states with two leptons, jets and $E_{\mathrm{T}}^{\mathrm{miss}}$, with intermediate decays via W and Z bosons.

The initial plan was for me to mainly work on a multivariate analysis, using methods such as boosted decision trees or neural networks, to investigate the possibility of gaining sensitivity compared to the more traditional cut-based analysis methods. However, it turned out that we were missing people in the working group that could develop the cut-based analysis with the full Run 2 dataset, which was considered the main priority, and should establish a baseline to compare the multivariate analysis against. My focus was therefore quickly shifted towards development of the cut-based analysis, with the multivariate analysis as more of a lower-priority side-project.

Also this time, I was involved in defining and optimizing signal regions, with accompanying control and validation regions. In addition, I became more involved in the development of common analysis software within the working group, production of common analysis ntuples along with requests and production of Monte Carlo samples for background and signal processes.

From January to December 2019, I was asked to be one of two analysis contacts for the working group targeting electroweak and strong production of the above mentioned SUSY scenario. My role was then to keep in touch with

the conveners of the SUSY subgroups managing searches for electroweak and strong production in ATLAS, as well as taking part in leading the efforts within the working group itself.

8.2 Event and object selection

The objects selected for the full Run 2 analysis are in most respects equal to those selected for the 2015+2016 analysis, but there are a few updates that will be addressed here. The changes are mostly related to new object *working points* (WPs), which correspond to optimized selection efficiencies for targeted regions of phase space.

In order to improve the efficiency of high-quality leptons in a high-pile-up environment, new isolation WPs were designed for both electrons [147, 172] and muons [173]. We therefore moved to the recommended *Tight* isolation requirements for high-p_{T} electrons and muons, which make use of a multivariate track-to-vertex-association score to determine how many tracks come from the same interaction vertex as the leptons in question.

New $E_{\mathrm{T}}^{\mathrm{miss}}$ WPs also became available, with new sets of requirements put on the objects that go into the vector p_{T} sum [174]. We studied the impact of the so-called *Loose* and *Tight* $E_{\mathrm{T}}^{\mathrm{miss}}$ WPs [175] on our signal sensitivity. The Loose WP definition was more similar to the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation used for the 2015+2016 analysis, but as the impact of moving from one WP to the other was shown not to significantly affect our sensitivity, we decided to go with the general recommendation of using the Tight $E_{\mathrm{T}}^{\mathrm{miss}}$ WP, which was designed to be more robust against pile-up.

8.3 $E_{\mathrm{T}}^{\mathrm{miss}}$ significance

$E_{\mathrm{T}}^{\mathrm{miss}}$, the negative vector sum of the momenta and/or energies of reconstructed physics objects and low-momentum tracks, is one of the most important variables for distinguishing SUSY processes from SM backgrounds. Usually, a hard cut on $E_{\mathrm{T}}^{\mathrm{miss}}$ is very efficient at rejecting the backgrounds, while keeping most of the expected signal. However, as discussed earlier, so-called "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ may arise from significantly mis-measured jets or leptons. Hence, SM backgrounds such as $Z+\mathrm{jets}$, which in reality should not yield both two leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$ in the same process, may contribute to significant amounts of reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ in the signal regions.

For the early Run 2 analysis, targeting production of a $\tilde{\chi}_1^{\pm}\,\tilde{\chi}_2^0$-pair with decays via W and Z into $2\ell+\mathrm{jets}+E_{\mathrm{T}}^{\mathrm{miss}}$ in the final state, as discussed in Chapter 7, this issue was addressed by considering $E_{\mathrm{T}}^{\mathrm{miss,rel}}$ instead, which effectively rejects any $E_{\mathrm{T}}^{\mathrm{miss}}$ component which is parallel with the jet or lepton that is closest to the direction of the reconstructed $p_{\mathrm{T}}^{\mathrm{miss}}$, in the transverse plane.

For the full Run 2 analysis we consider yet another approach to mitigate the effect of artificially large contributions from the important $E_{\mathrm{T}}^{\mathrm{miss}}$-variable. This

involves making an estimate of the relative uncertainty in the $E_{\mathrm{T}}^{\mathrm{miss}}$-calculation. We call this variable the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance.

A simple way to estimate this significance is to take the square root of the sum of hard (reconstructed objects) or total (soft + hard) transverse energy in the event, commonly labeled H_{T} and E_{T}, respectively, as an estimate of the uncertainty in the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$. This proxy significance is then defined as $E_{\mathrm{T}}^{\mathrm{miss}}/\sqrt{H_{\mathrm{T}}}$ or $E_{\mathrm{T}}^{\mathrm{miss}}/\sqrt{E_{\mathrm{T}}}$. However, these definitions assume an $E_{\mathrm{T}}^{\mathrm{miss}}$-resolution which is purely calorimeter-dependent.

In order to get a better handle on the actual uncertainties involved in the calculation of the $E_{\mathrm{T}}^{\mathrm{miss}}$, an object-based $E_{\mathrm{T}}^{\mathrm{miss}}$ significance [176] has been developed that uses the energy/momentum resolutions of all contributions to the calculation. This includes both the hard inner detector tracks of jets and leptons, in addition to the soft tracks not associated to reconstructed physics objects, which also enter into the $E_{\mathrm{T}}^{\mathrm{miss}}$-calculation.

The object-based $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ is formally defined as a hypothesis test, where the likelihood of the measured 2-dimensional $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$, given a non-zero true $\mathbf{p}_{\mathrm{T}}^{\mathrm{inv}}$ from invisible objects, is compared to the corresponding likelihood, given a true invisible $\mathbf{p}_{\mathrm{T}}$ of zero:

$$(E_{\mathrm{T}}^{\mathrm{miss,sig}})^2 = 2\ln\left(\frac{\max_{\mathbf{p}_{\mathrm{T}}^{\mathrm{inv}}\neq\mathbf{0}}\mathcal{L}(\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}|\mathbf{p}_{\mathrm{T}}^{\mathrm{inv}})}{\max_{\mathbf{p}_{\mathrm{T}}^{\mathrm{inv}}=\mathbf{0}}\mathcal{L}(\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}|\mathbf{p}_{\mathrm{T}}^{\mathrm{inv}})}\right). \tag{8.1}$$

If we assume that the measurement of each object or energy deposit is independent, that the measurements have Gaussian errors and that the total momentum should sum to zero (momentum conservation), we can express this as

$$(E_{\mathrm{T}}^{\mathrm{miss,sig}})^2 = 2\ln\left(\frac{\mathcal{L}(\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}|\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}})}{\mathcal{L}(\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}|\mathbf{0})}\right) \tag{8.2}$$

$$= (\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}})^{\mathrm{T}}\left(\sum_i \mathbf{V}_i\right)^{-1}(\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}), \tag{8.3}$$

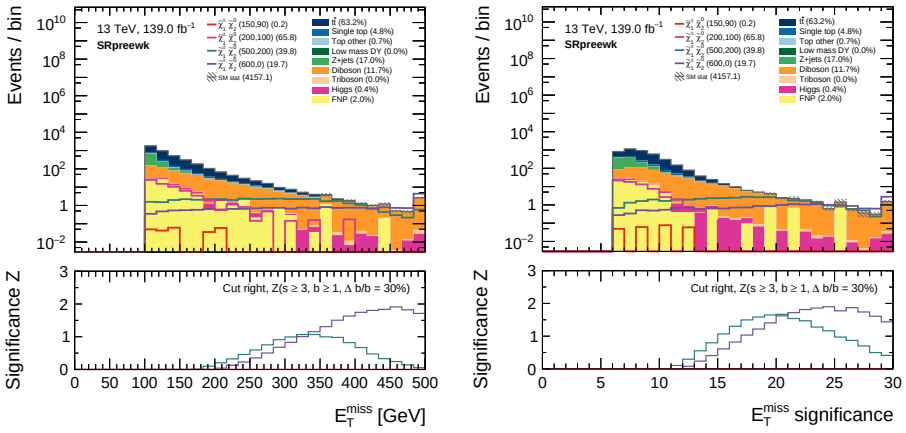
where $\mathbf{V}_i$ is the 2×2 covariance matrix of measurement i.

Alternatively, we can rewrite the above expression in a form that resembles the simpler event-based versions of $E_{\mathrm{T}}^{\mathrm{miss}}$ significance:

$$E_{\mathrm{T}}^{\mathrm{miss,sig}} = \frac{|\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}|}{\sigma(\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}})}, \tag{8.4}$$

where $\sigma(\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}})$ represents the magnitude of the uncertainty in the $E_{\mathrm{T}}^{\mathrm{miss}}$ and incorporates the individual object-based measurement resolutions, with possible correlations between measurements in the two dimensions of the transverse plane.

In Figure 8.3, the sensitivity to four benchmark signal mass-hypotheses are shown in the lower panels, using the regular $E_{\mathrm{T}}^{\mathrm{miss}}$ itself on the left and the object-based $E_{\mathrm{T}}^{\mathrm{miss}}$ significance on the right. As we can see from the plots, the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance is a little more sensitive to the benchmark signal scenarios, in particular for the medium mass-splitting point, at $(m_{\tilde{\chi}_1^\pm,\tilde{\chi}_2^0}, m_{\tilde{\chi}_1^0}) = (500, 200)$ GeV. This



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

(b) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in a preselection region

Figure 8.3: (a) Regular $E_{\mathrm{T}}^{\mathrm{miss}}$ vs. (b) object based $E_{\mathrm{T}}^{\mathrm{miss}}$ significance for stacked SM backgrounds in filled histograms compared to various combinations of signal masses shown by superimposed lines. Preselection cuts (SRpreewk) requiring exactly two same-flavor opposite-sign electrons or muons, at least 2 signal jets, at most one b-tagged jet, $m_{\ell\ell} \in (71, 91)\,\mathrm{GeV}$, $m_{jj} \in (60, 111)\,\mathrm{GeV}$, $E_{\mathrm{T}}^{\mathrm{miss}} > 100\,\mathrm{GeV}$ and $E_{\mathrm{T}}^{\mathrm{miss,sig}} > 6$ have been applied to the dataset before plotting. Notice, in particular, how the Z-value (in the lower panels) of the intermediate mass-splitting point with $(m_{\tilde{\chi}_1^{\pm}, \tilde{\chi}_2^0}, m_{\tilde{\chi}_1^0}) = (500, 200)\,\mathrm{GeV}$, shown in turquoise, increases significantly in the right plot compared to the plot on the left. The lower panels show the calculated Z-value as a function of lower cut, i.e., events to the right of a given cut in the upper panel are kept. The relative uncertainty on the background ($\Delta B/B$) used in the calculation is 30%. At least three signal events and at least one background event is required in order for a given cut/bin to get a non-zero Z-value.

is mostly the case due to the effective suppression of fake $E_{\mathrm{T}}^{\mathrm{miss}}$ contributions, which generally have larger uncertainties in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation compared with the contributions from real $E_{\mathrm{T}}^{\mathrm{miss}}$. This effectively shifts the higher values of $E_{\mathrm{T}}^{\mathrm{miss}}$ from non-diboson backgrounds, most prominently $Z + \mathrm{jets}$ and $t\bar{t}$, to lower values in $E_{\mathrm{T}}^{\mathrm{miss}}$ significance, as compared to the diboson backgrounds, without noticably affecting the signal contributions. Thus, we can increase the signal significance (Z-value), especially for intermediate mass-splitting scenarios, by using $E_{\mathrm{T}}^{\mathrm{miss}}$ significance instead of $E_{\mathrm{T}}^{\mathrm{miss}}$ alone. From the lower panel of Figure 8.3, we can see that the maximum Z-value for the $(500, 200)$ GeV sample (turquoise) increases from roughly 1.1 using $E_{\mathrm{T}}^{\mathrm{miss}}$ in (a), to 1.7 when using $E_{\mathrm{T}}^{\mathrm{miss}}$ significance in (b) instead.

8.4 Conventional signal regions

We consider the $m_{\tilde{\chi}_1^\pm/\tilde{\chi}_2^0}, m_{\tilde{\chi}_1^0}$-plane in three distinct parts, according to the mass-difference between the two sparticles, as the final-state kinematics change significantly from the small-Δm to the large-Δm portion of the phase space.

The large-Δm regions are typically where the SUSY signals are most distinguishable from the SM backgrounds. Therefore, I start by describing the event and object selections optimized for maximum sensitivity to the large mass-splittings, along with regions to validate the backgrounds in these signal regions, before continuing on to the intermediate and small mass-splitting regions, where the SUSY signals become increasingly difficult to distinguish from the backgrounds.

8.4.1 High-Δm regions

The kinematics of large mass-splitting scenarios are characterized by high-p_T final-state objects and large $E_\mathrm{T}^\mathrm{miss}$. The high-mass signal regions, called SR-High, which are defined in Table 8.1, therefore target large values of $E_\mathrm{T}^\mathrm{miss}$ and $E_\mathrm{T}^\mathrm{miss,sig}$, along with relatively small angular jet separation, ΔR_{jj}, from the collimated decay products of the W. In order to make use of shape differences between signal and backgrounds in $E_\mathrm{T}^\mathrm{miss,sig}$ and ΔR_{jj}, we define four SR-High bins: $\Delta R_{jj} \in (0, 0.8)$ called SR-High-8 and $\Delta R_{jj} \in (0.8, 1.6)$ called SR-High-16, which both are further split into two $E_\mathrm{T}^\mathrm{miss,sig}$ bins, where $E_\mathrm{T}^\mathrm{miss,sig} \in (18, 21)$ and $E_\mathrm{T}^\mathrm{miss,sig} \geq 21$, with region names suffixed by -a and -b, respectively. Plots after the $E_\mathrm{T}^\mathrm{miss,sig} > 18$ cut has been applied, are shown for SR-High-8 and SR-High-16 in Figure 8.4a and 8.4b, respectively.

In order to select event topologies consistent with the signal model, we require that the events have two same-flavor (SF) leptons with opposite-sign (OS) electric charges, and at least two hadronic jets. The leptons need to have $p_\mathrm{T} > 25\,\mathrm{GeV}$ in order to be on the trigger efficiency plateau, and jets to have $p_\mathrm{T} > 30\,\mathrm{GeV}$ to suppress events where the jets come from other hadronic activity, such as elastic scattering and pile-up events.

All high-mass regions must also have $E_\mathrm{T}^\mathrm{miss} > 100\,\mathrm{GeV}$ in order to suppress SM backgrounds in general, and have $m_\mathrm{T2} > 80\,\mathrm{GeV}$ in order to suppress backgrounds involving the W-boson, such as WW and $t\bar{t}$. The two highest-p_T jets in each event are required to make up an invariant mass between 60 and 110 GeV, to be consistent with coming from a W-boson. By requiring the number of b-tagged jets (classified at 77% efficiency) to be less than or equal to 1, we reject most of the backgrounds involving top quarks, such as $t\bar{t}$ and tW.

After the cuts on $E_\mathrm{T}^\mathrm{miss,sig}$ and m_T2, the backgrounds left in in SR-High are dominated by ZZ-contributions. This allows us to relax the dilepton invariant mass-window from $m_{\ell\ell} \in (80, 100)\,\mathrm{GeV}$, as is used for the intermediate and low-mass regions, to $m_{\ell\ell} \in (70, 110)\,\mathrm{GeV}$, to increase statistics without decreasing the signal purity, thereby increasing the sensitivity. A similar argument can be

made for the inclusion of up to one b-tagged jet, as the rate of ZZ decaying into b-quarks is rather low.

We can see from Figure 8.4a that we have significant sensitivity to the $(500, 200)$ GeV mass-point with the cut at $E_{\mathrm{T}}^{\mathrm{miss,sig}} > 18$, while there is more to gain for the $(600, 0)$ GeV point by cutting harder, e.g., around $E_{\mathrm{T}}^{\mathrm{miss,sig}} > 21$. However, we can also notice the differences in shape of the distributions between the falling background and the two mentioned signal mass-points with peaks at different positions above $E_{\mathrm{T}}^{\mathrm{miss,sig}} > 18$. By binning this region in the two $18 < E_{\mathrm{T}}^{\mathrm{miss,sig}} < 21$ and $E_{\mathrm{T}}^{\mathrm{miss,sig}} > 21$ bins, we can exploit part of these differences in shape, in addition to the differences in total event yields.

The lower significance panel in Figure 8.4a shows similar behavior as well, but there is more background and less signal in general in the $\Delta R_{jj} \in (0.8, 1.6)$ region than there was in the more boosted $\Delta R_{jj} \in (0, 0.8)$ region. Since the two regions are orthogonal (non-overlapping), we can add them both simultaneously to the likelihood fit in the end, and sensitivity in any region will contribute to the overall sensitivity to the various signal points.

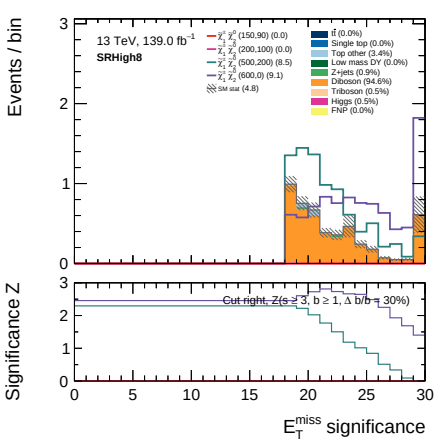
The background modeling in the SR-High bins are validated in two validation regions, which are defined in Table 8.1. VR-High is designed to validate the m_{jj} sidebands, with dijet invariant masses off the W-resonance and with $m_{jj} > 20$ GeV to stay within the bounds of the MC background simulations, whereas VR-High-R also includes events with $\Delta R_{jj} > 1.6$. The $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ distributions in the two VRs are plotted in Figure 8.5a and 8.5b, respectively. The VRs are highly dominated by diboson contributions, as are the SRs, and there does not seem to be any systematic mis-modeling in the yields or shapes of the backgrounds in these regions.

The default jet reconstruction algorithm uses a clustering radius of $R = 0.4$ to reconstruct single jets. In cases where the W of the event is highly energetic, also referred to as a "boosted W", its decay products become highly collimated. At sufficiently high boost, the two daughter jets become so collimated that the reconstruction algorithm is unable to tell them apart, and consequently reconstructs them as a single jet. A separate signal region, called SR-High-4 has been defined to cover the cases where there is only one signal jet in the event, but where this jet has a mass consistent with the W-mass. The distributions of $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and the mass of the leading jet in SR-High-4 are plotted in Figure 8.6a and 8.6b, respectively. The level of background in this region is low, and the signals mentioned earlier in this section are clearly visible above the background in the high $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ region. It is also striking to see the clear peak in the leading jet mass at the pole mass of the W boson.

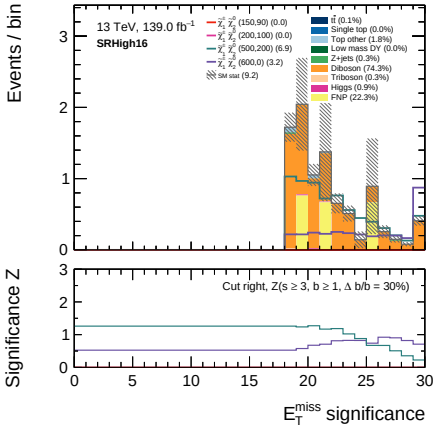
A dedicated validation region, called VR-High-4, is checking the background modeling in these one-jet cases, in the sidebands of the W-resonance. Plots of $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and the mass of the p_{T}-leading jet are shown in Figure 8.7a and 8.7b, respectively. These plots show fairly good overall modeling of the backgrounds, which also here are dominated by dibosons. SR-High-4 and VR-High-4 are defined in Table 8.1 as well.

Table 8.1: Signal and validation regions (SRs and VRs, respectively) targeting high-Δm scenarios. Common-High denotes the cuts that are common to all the regions listed below.

		Common-High			
Leptons		2 SFOS, $p_{\mathrm{T}} > 25\,\mathrm{GeV}$			
$\boldsymbol{E_{\mathrm{T}}^{\mathrm{miss}}}$ **[GeV]**		> 100			
$\boldsymbol{m_{\mathrm{T2}}}$ **[GeV]**		> 80			
$\boldsymbol{m_{\ell\ell}}$ **[GeV]**		$(71, 111)$			
$\boldsymbol{n_{b\text{-tags}}}$		≤ 1			
	SR-High	**VR-High**	**VR-High-R**	**SR-High-4**	**VR-High-4**
$\boldsymbol{n_{\mathrm{jets}}}$	≥ 2	≥ 2		1	1
$\boldsymbol{m_{j(j)}}$ **[GeV]**	$(60, 110)$	$> 20, \notin (60, 110)$	> 20	$(60, 110)$	$> 20, \notin (60, 110)$
$\boldsymbol{\Delta R_{jj}}$	$(0, 0.8, 1.6)$	< 1.6	> 1.6	–	–
$\boldsymbol{E_{\mathrm{T}}^{\mathrm{miss,sig}}}$	$(18, 21, \infty)$	$(18, \infty)$		$(12, \infty)$	$(12, \infty)$

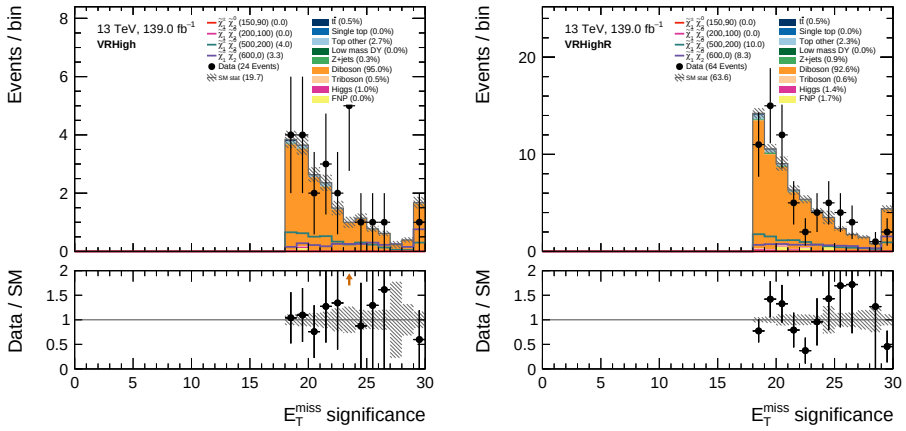


(a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-High-8 ($\Delta R_{jj} \in (0, 0.8)$)



(b) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-High-16 ($\Delta R_{jj} \in (0.8, 1.6)$)

Figure 8.4: $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ for various signal hypotheses (lines) compared to the SM background (stacked, filled histograms) in the (a) SR-High-8 and (b) SR-High-16 regions. The lower panels show the calculated Z-value as a function of lower cut, i.e., events to the right of a given cut in the upper panel are kept. The relative uncertainty on the background ($\Delta B/B$) used in the calculation is 30%. At least three signal events and at least one background event is required in order for a given cut/bin to get a non-zero Z-value. (The Z-values shown in the lower panel for bins to the left of the SR, i.e., $E_{\mathrm{T}}^{\mathrm{miss,sig}} < 18$, are superfluous, as they all get the same integrated number of signal and background events as the left-most bin in the SR.)



(a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in VR-High ($\Delta R_{jj} < 1.6$) (b) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in VR-High-R ($\Delta R_{jj} > 1.6$)

Figure 8.5: $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in data (black dots) compared to the SM background (stacked, filled histograms) in the VR-High and VR-High-R regions. The statistical uncertainty of the data points are indicated by black error bars, whereas the statistical uncertainty on the combined SM backgrounds are shown by hatched areas. The lower panels show the ratio of the number of data to background events for each bin in the upper panel.

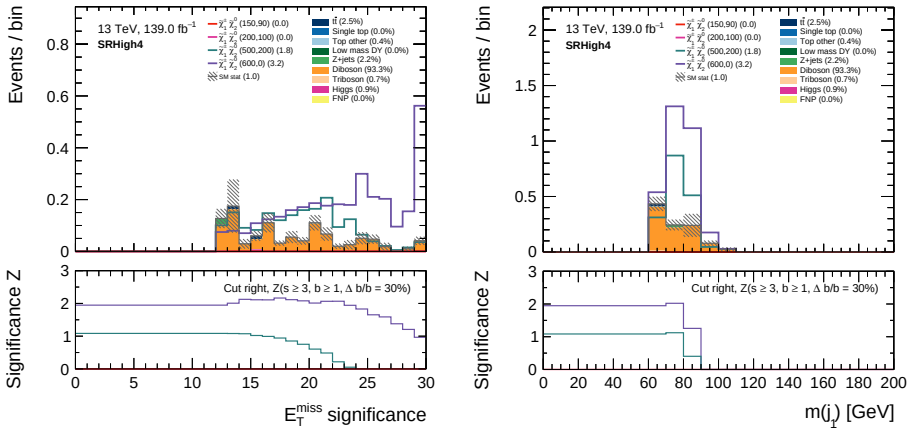
8.4.2 Intermediate-Δm regions

As we move to signal models where the $\tilde{\chi}_1^0$ is closer in mass to the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ compared to the high-Δm regions, i.e., with smaller Δm, there is not as much $E_{\mathrm{T}}^{\mathrm{miss}}$ being produced in the events. The intermediate Δm regions aim to capture the change in kinematics that follow from the smaller mass-splittings, yet without being significantly constrained by low final-state momenta. This predominantly means targeting higher LSP masses.

In this region of phase space, the background is no longer dominated by pure ZZ-contributions, and therefore we move from allowing one b-tagged jet in SR-High to vetoing b-jets altogether. By the same argument, we also move to a narrower dilepton invariant mass window, only allowing $m_{\ell\ell}$ of $\pm 10\,\mathrm{GeV}$ away from the Z-mass, instead of $\pm 20\,\mathrm{GeV}$ for SR-High. These are the standard selections that we also use for SR-Low, to be discussed in Section 8.4.3.

The signal regions targeting intermediate Δm are called SR-Int and are defined in Table 8.2. The main difference w.r.t. the SR-High regions is that they cover lower values of $E_{\mathrm{T}}^{\mathrm{miss,sig}}$, namely $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ between 12 and 18. Two bins in $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ are defined to capture shape differences between the signal and the background. SR-Int-a covers $E_{\mathrm{T}}^{\mathrm{miss,sig}} \in (12, 15)$, whereas SR-Int-b covers $E_{\mathrm{T}}^{\mathrm{miss,sig}} \in (15, 18)$. Additionally, both of the SR-Int bins require a hard jet, where the leading (highest-p_{T}) jet of the event must have $p_{\mathrm{T}} > 60\,\mathrm{GeV}$.

Figure 8.8 shows plots of $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and m_{jj} for various benchmark signal



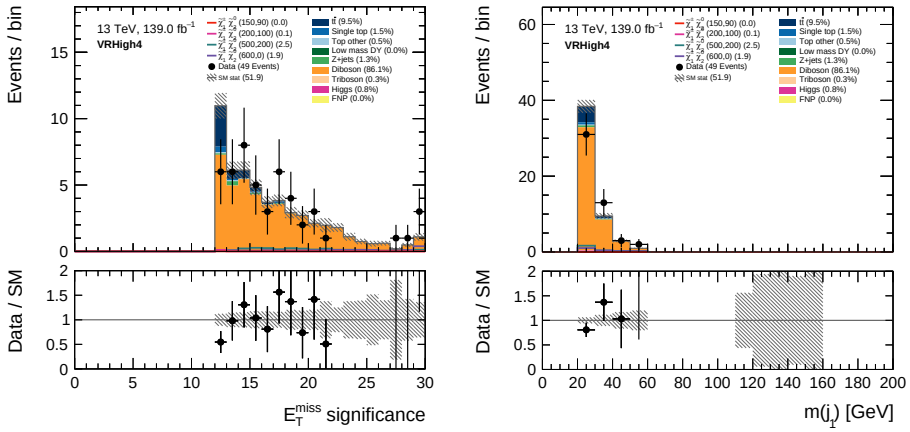
(a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-High-4 ($R = 0.4$) (b) m_{j_1} in SR-High-4 ($R = 0.4$)

Figure 8.6: (a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and (b) the mass of the leading jet for various signal hypotheses (lines) compared to the SM background (stacked, filled histograms) in the SR-High-4 region. The lower panels show the calculated Z-value as a function of lower cut, i.e., events to the right of a given cut in the upper panel are kept. The relative uncertainty on the background ($\Delta B/B$) used in the calculation is 30%. At least three signal events and at least one background event is required in order for a given cut/bin to get a non-zero Z-value. (The Z-values shown in the lower panel for bins to the left of the SR, i.e., $E_{\mathrm{T}}^{\mathrm{miss,sig}} < 12$ and $m_{j_1} < 60\,\mathrm{GeV}$, are superfluous, as they all get the same integrated number of signal and background events as the left-most bin in the SR.)

mass-points compared to the expected SM background in SR-Int. By comparison with the plots of SR-High in Figure 8.4 and SR-High-4 in Figure 8.6, we can see that SR-Int has more background and is most sensitive to the $(500, 200)\,\mathrm{GeV}$ mass-point, with $\Delta m = 300\,\mathrm{GeV}$, among the benchmark signal points.

The background in SR-Int is dominated by diboson processes (WZ and ZZ), with small contributions from top and Z + jets. We define control regions for diboson and $t\bar{t}$, which allow dedicated normalization factors to be fitted to the regions dominated by these processes. CR-VZ looks in the m_{jj}-sideband regions of the W-resonance, in the already diboson-dominated phase space, and drops the p_{T}-cut on the leading jet. In order to obtain a control region pure in top-processes, CR-tt requires at least one b-tagged jet, and looks in the lower $E_{\mathrm{T}}^{\mathrm{miss,sig}}$-sideband with values between 9 and 12, to get sufficient statistics. Figure 8.9a and 8.9b show $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in CR-ZZ and CR-tt, respectively, which confirm that the regions are dominated by dibosons in the first case and by $t\bar{t}$ in the second, and that the regions have reasonably high statistics.

The modeling of the SM background in SR-Int is validated in VR-Int, which has a reversed p_{T}-cut on the leading jet compared to SR-Int, i.e., looks in the lower sideband at $p_{\mathrm{T}}(j_1) < 60\,\mathrm{GeV}$. Plots of $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and m_{jj} in VR-Int are



(a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in VR-High-4 ($R = 0.4$) (b) m_{j_1} in VR-High-4 ($R = 0.4$)

Figure 8.7: (a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and (b) the mass of the leading jet in data (black dots) compared to the SM background (stacked, filled histograms) in the VR-High-4 region. The statistical uncertainty of the data points are indicated by black error bars, whereas the statistical uncertainty on the combined SM backgrounds are shown by hatched areas. The lower panels show the ratio of the number of data to background events for each bin in the upper panel.

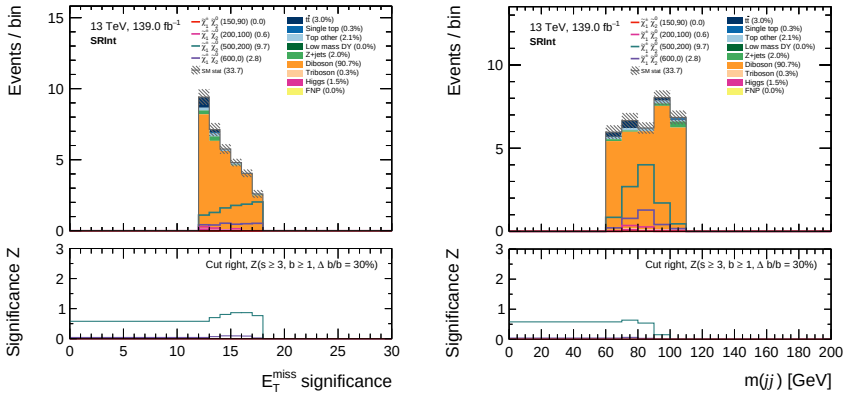
Table 8.2: Signal, validation and control regions (SR, VR and CRs, respectively) targeting intermediate-Δm scenarios. Common-Int denotes the cuts that are common to all the regions listed below.

	Common-Int			
Leptons	2 SFOS, $p_{\mathrm{T}} > 25\,\mathrm{GeV}$			
$\boldsymbol{E_{\mathrm{T}}^{\mathrm{miss}}}$ **[GeV]**	> 100			
$\boldsymbol{m_{\mathrm{T2}}}$ **[GeV]**	> 80			
$\boldsymbol{m_{\ell\ell}}$ **[GeV]**	$\in (81, 101)$			
$\boldsymbol{n_{\mathrm{jets}}}$	≥ 2			
	SR-Int	**VR-Int**	**CR-VZ**	**CR-tt**
$\boldsymbol{n_{b\text{-tags}}}$	0	0	0	≥ 1
$\boldsymbol{m_{jj}}$ **[GeV]**	$(60, 110)$	$(60, 110)$	$> 20, \notin (60, 110)$	> 20
$\boldsymbol{p_{\mathrm{T}}(j_1)}$ **[GeV]**	> 60	< 60	–	> 60
$\boldsymbol{E_{\mathrm{T}}^{\mathrm{miss,sig}}}$	$(12, 15, 18)$	$(12, 18)$	$(12, 18)$	$(9, 12)$

shown in Figure 8.10a and 8.10b, respectively. The modeling looks decent, taking the relatively low statistics of the region into account.

8.4.3 Low-Δm regions

The low-Δm regions target mass-splittings approximately equal to, or somewhat larger than, the masses of the SM gauge bosons that mediate the sparticle decays.



(a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-Int (b) m_{jj} in SR-Int

Figure 8.8: (a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and (b) m_{jj} for various signal hypotheses (lines) compared to the SM background (stacked, filled histograms) in the SR-Int region. The lower panels show the calculated Z-value as a function of lower cut, i.e., events to the right of a given cut in the upper panel are kept. The relative uncertainty on the background ($\Delta B/B$) used in the calculation is 30%. At least three signal events and at least one background event is required in order for a given cut/bin to get a non-zero Z-value. (The Z-values shown in the lower panel for bins to the left of the SR, i.e., $E_{\mathrm{T}}^{\mathrm{miss,sig}} < 12$ and $m_{jj} < 60\,\mathrm{GeV}$, are superfluous, as they all get the same integrated number of signal and background events as the left-most bin in the SR.)

The signal in this region of phase space is characterized by soft decay products, and therefore becomes kinematically similar to the SM backgrounds.

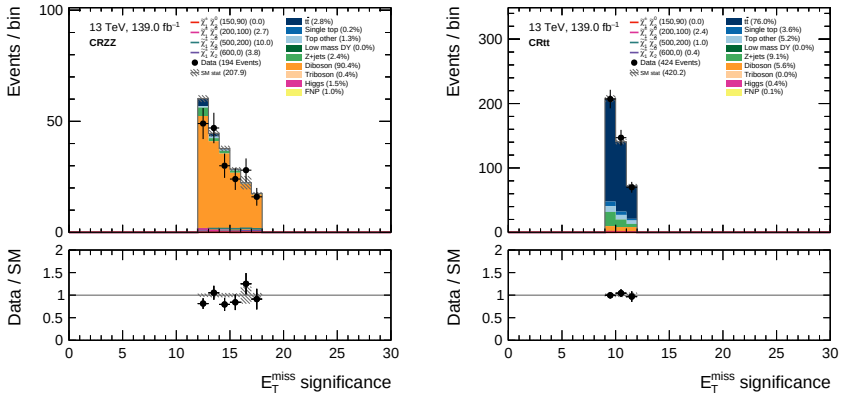
The SR-Low regions, defined in Table 8.3, continue to go even lower in $E_{\mathrm{T}}^{\mathrm{miss,sig}}$, and are separated in two bins: SR-Low-a with $E_{\mathrm{T}}^{\mathrm{miss,sig}} \in (6, 9)$ and SR-Low-b with $E_{\mathrm{T}}^{\mathrm{miss,sig}} \in (9, 12)$.

This region of phase space cuts heavily into SM territory, and especially $Z+\mathrm{jets}$ becomes a large source of background. By restricting the number of signal jets to exactly two, instead of two or more, the $Z+\mathrm{jets}$ background is reduced. Additionaly, restricting the angular separation of the leptons, $\Delta R_{\ell\ell} < 1$, the collimated leptons from the energetic Zs produced in the SUSY events are kept, whereas both $Z+\mathrm{jets}$ and $t\bar{t}$ are further reduced.

A plot of $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-Low is shown in Figure 8.11a. From the lower panel of the plot, we see that this region shows some sensitivity to the $(200, 100)\,\mathrm{GeV}$ mass-point, but that it is far from reaching expected exclusion significance.

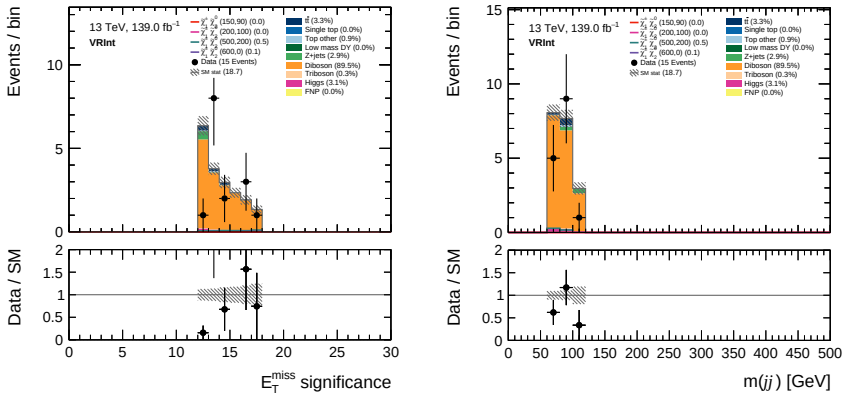
The modeling of the backgrounds in SR-Low is checked in VR-Low, which is a higher-end sideband in $\Delta R_{\ell\ell}$ to SR-Low. Figure 8.12 shows a plot of $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in VR-Low, which indicates fairly good modeling of the background, given the limited statistics.

An additional low-mass region, called SR-Low-2, is designed to pick up some



(a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in CR-ZZ (b) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in CR-tt

Figure 8.9: $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in data (black dots) compared to the SM background (stacked, filled histograms) in the (a) CR-ZZ and (b) CR-tt regions. The statistical uncertainty of the data points are indicated by black error bars, whereas the statistical uncertainty on the combined SM backgrounds are shown by hatched areas. The lower panels show the ratio of the number of data to background events for each bin in the upper panel.



(a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in VR-Int (b) m_{jj} in VR-Int

Figure 8.10: (a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and (b) m_{jj} in data (black dots) compared to the SM background (stacked, filled histograms) in the VR-Int region. The statistical uncertainty of the data points are indicated by black error bars, whereas the statistical uncertainty on the combined SM backgrounds are shown by hatched areas. The lower panels show the ratio of the number of data to background events for each bin in the upper panel.

Table 8.3: Signal, validation and control regions (SRs, VRs and CRs, respectively) targeting low-Δm scenarios. Common-Low denotes the cuts that are common to all the regions listed below.

			Common-Low		
Leptons			2 SFOS, $p_{\mathrm{T}} > 25\,\mathrm{GeV}$		
$\boldsymbol{E_{\mathrm{T}}^{\mathrm{miss}}}$ **[GeV]**			> 100		
$\boldsymbol{m_{\ell\ell}}$ **[GeV]**			$\in (81, 101)$		
$\boldsymbol{n_{\mathrm{jets}}}$			2		
$\boldsymbol{n_{b\text{-tags}}}$			0		
	SR-Low	**VR-Low**	**SR-Low-2**	**VR-Low-2**	**CR-Z**
$\boldsymbol{m_{jj}}$ **[GeV]**	$(60, 110)$	$(60, 110)$	$(60, 110)$	$> 20, \notin (60, 110)$	$> 20, \notin (60, 110)$
$\boldsymbol{m_{\mathrm{T2}}}$ **[GeV]**	> 80	> 80	< 80	< 80	> 80
$\boldsymbol{\Delta R_{\ell\ell}}$	< 1	$(1, 1.4)$	< 1.6	< 1.6	–
$\boldsymbol{\Delta\phi(p_{\ell\ell}, E_{\mathrm{T}}^{\mathrm{miss}})}$	–	–	< 0.6	< 0.6	–
$\boldsymbol{E_{\mathrm{T}}^{\mathrm{miss,sig}}}$	$(6, 9, 12)$	$(6, 12)$	$(6, 9)$	$(6, 9)$	$(6, 9)$

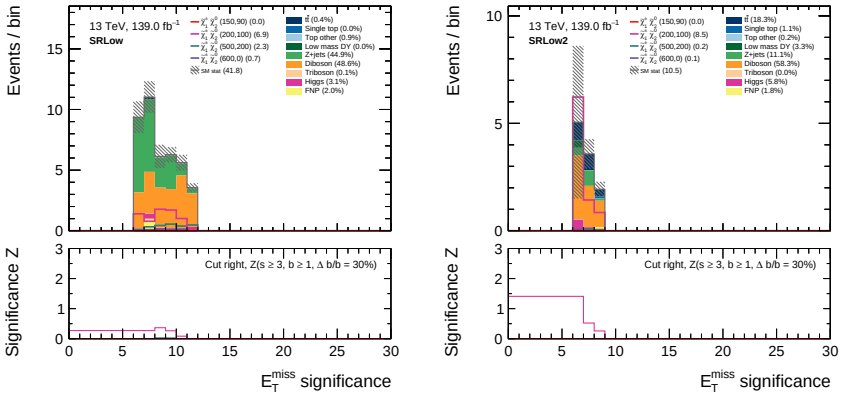
remaining sensitivity for $m_{\mathrm{T2}} < 80\,\mathrm{GeV}$, in cases where there is angular alignment of the transverse momentum of the two-lepton system with the $E_{\mathrm{T}}^{\mathrm{miss}}$. This behavior is suppressed in the processes that produce the fake $E_{\mathrm{T}}^{\mathrm{miss}}$ in $Z+\mathrm{jets}$ events. This region only covers $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ between 6 and 9, and extends the cut on the lepton angular separation from 1 to 1.6. A plot of $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-Low-2 is shown in Figure 8.11b, and shows far better sensitivity to the $(200, 100)\,\mathrm{GeV}$ mass-point than the corresponding plot for SR-Low.

The background modeling in SR-Low-2 is validated in the m_{jj}-sidebands of the W-resonance. Figure 8.12b indicates that the backgrounds are reasonably well modeled for $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in VR-Low-2 as well.

Since $Z+$jets is the dominant background for the SR-Low-regions, a dedicated control region is defined to allow for a normalization factor for the $Z+\mathrm{jets}$ background to be fitted to data. The control region CR-Z considers the lower $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ bin of SR-Low and the m_{jj}-sidebands off the W-resonance. Figure 8.13 shows plots of $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and m_{jj} in CR-Z.

8.4.4 Off-shell regions

When the mass-splitting between the $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ and the $\tilde{\chi}_1^0$ becomes so small that it goes below the (pole) masses of the W and Z bosons that mediate the decays, the bosons must be produced as virtual particles with lower masses, i.e., off their mass-shells. Then, there is no longer a clear dijet resonance at the W-mass to separate signal events from events with other jet activity, and therefore we do not consider an m_{jj} cut in the off-shell case. On the other hand, because we require exactly two leptons, the dilepton invariant mass still makes a solid discriminant. In fact, it now becomes the most powerful feature of the event, as we should see a kinematic edge (or endpoint) in the $m_{\ell\ell}$-distribution at the mass corresponding to the mass-splitting of the mother and daughter sparticles.



(a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-Low (b) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-Low-2

Figure 8.11: $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ for various signal hypotheses (lines) compared to the SM background (stacked, filled histograms) in the (a) SR-Low and (b) SR-Low-2 regions. The lower panels show the calculated Z-value as a function of lower cut, i.e., events to the right of a given cut in the upper panel are kept. The relative uncertainty on the background ($\Delta B/B$) used in the calculation is 30%. At least three signal events and at least one background event is required in order for a given cut/bin to get a non-zero Z-value. (The Z-values shown in the lower panel for bins to the left of the SR, i.e., $E_{\mathrm{T}}^{\mathrm{miss,sig}} < 6$, are superfluous, as they all get the same integrated number of signal and background events as the left-most bin in the SR.)

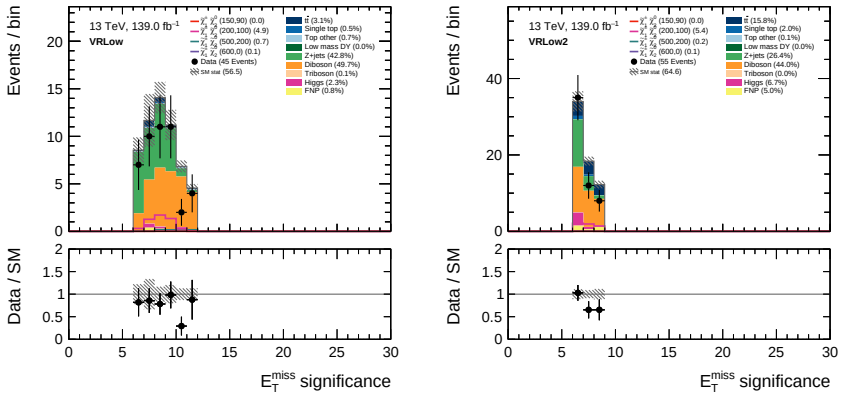
The signal region SR-OffShell has two bins, SR-OffShell-a and SR-OffShell-b, with $m_{\ell\ell} \in (12, 40)$ and $\in (40, 71)$ GeV, respectively. The upper cut at 71 GeV is chosen to avoid the Z-mass resonance at 91 GeV.

As the dominant backgrounds in the off-shell regions are diboson and $t\bar{t}$, the m_{T2}-cut is raised from 80 to 100 GeV to suppress any slightly-above-pole-mass Ws produced in $t\bar{t}$ processes. In order to boost the sensitivity of the otherwise very soft (low-p_{T}) final-state particles, we require a hard leading jet, with $p_{\mathrm{T}} > 100$ GeV, recoiling against the $E_{\mathrm{T}}^{\mathrm{miss}}$, with $\Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}}) > 2$. In the case of three or more signal jets in the event, the leading jet(s) are interpreted as likely coming from initial state radiation (ISR).

The distribution of $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-OffShell is shown in Figure 8.14a. The plot indicates that the region is close to expected exclusion sensitivity for the $(150, 90)$ GeV mass-point, with $\Delta m = 60$ GeV.

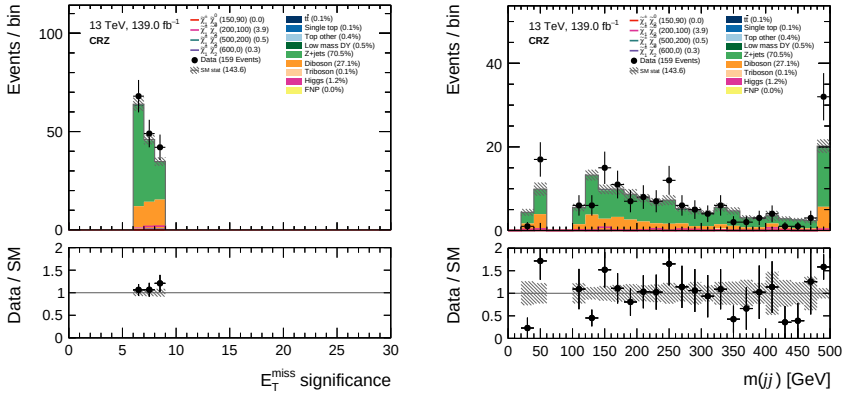
A control region, CR-DY, is defined in the lower m_{T2}- and $E_{\mathrm{T}}^{\mathrm{miss,sig}}$-sidebands in order to fit a normalization factor for the Z^*/γ^*-background (Drell-Yan). Figure 8.15 shows the distributions of $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and $m_{\ell\ell}$ in CR-DY.

The total background modeling is validated in VR-OffShell, which is a lower sideband to SR-OffShell in m_{T2}, and is plotted for $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in Figure 8.14b. Also here, the modeling looks decent, given the limited statistics.



(a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in VR-Low (b) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in VR-Low-2

Figure 8.12: $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in data (black dots) compared to the SM background (stacked, filled histograms) in the (a) VR-Low and (b) VR-Low-2 regions. The statistical uncertainty of the data points are indicated by black error bars, whereas the statistical uncertainty on the combined SM backgrounds are shown by hatched areas. The lower panels show the ratio of the number of data to background events for each bin in the upper panel.

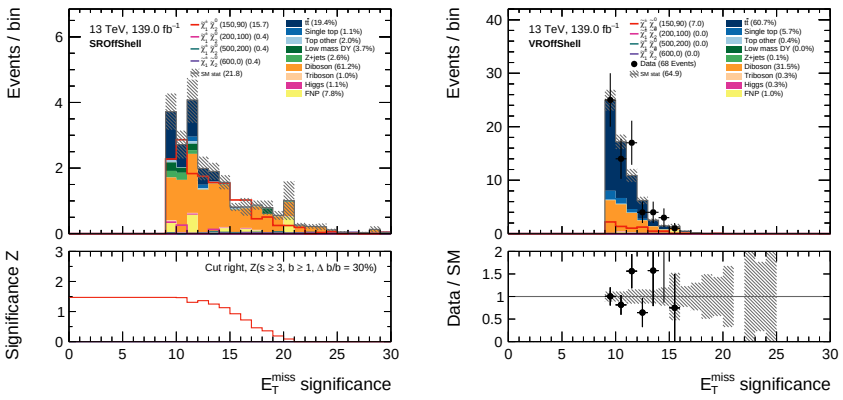


(a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in CR-Z (b) m_{jj} in CR-Z

Figure 8.13: (a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and (b) m_{jj} in data (black dots) compared to the SM background (stacked, filled histograms) in the CR-Z region. The statistical uncertainty of the data points are indicated by black error bars, whereas the statistical uncertainty on the combined SM backgrounds are shown by hatched areas. The lower panels show the ratio of the number of data to background events for each bin in the upper panel.

Table 8.4: Signal, validation and control regions (SRs, VR and CR, respectively) targeting off-shell scenarios. Common-OffShell denotes the cuts that are common to all the regions listed below.

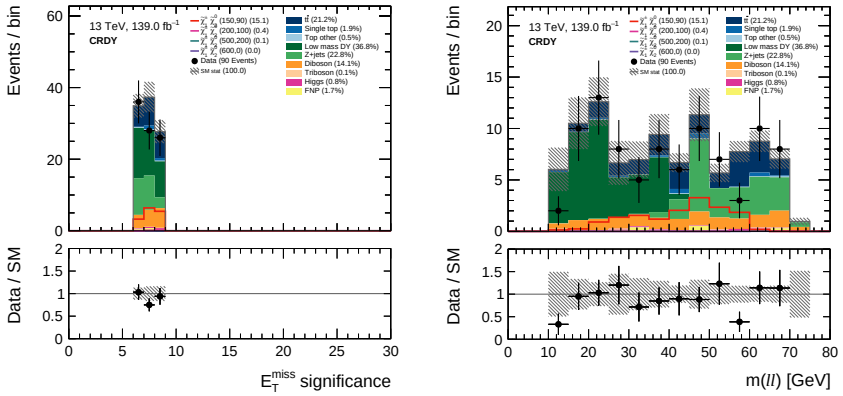
	Common-OffShell		
Leptons	2 SFOS, $p_{\mathrm{T}} > 25\,\mathrm{GeV}$		
$\boldsymbol{E_{\mathrm{T}}^{\mathrm{miss}}}$ **[GeV]**	> 100		
$\boldsymbol{n_{\mathrm{jets}}}$	≥ 2		
$\boldsymbol{n_{b\text{-tags}}}$	0		
	SR-OffShell	**VR-OffShell**	**CR-DY**
$\boldsymbol{m_{\mathrm{T2}}}$ **[GeV]**	> 100	$(80, 100)$	< 100
$\boldsymbol{p_{\mathrm{T}}(j_1)}$ **[GeV]**	> 100	> 100	–
$\boldsymbol{\Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}})}$	> 2	> 2	–
$\boldsymbol{E_{\mathrm{T}}^{\mathrm{miss,sig}}}$	> 9	> 9	$(6, 9)$
$\boldsymbol{m_{\ell\ell}}$ **[GeV]**	$(12, 40, 71)$	$(12, 71)$	$(12, 71)$



(a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-OffShell

(b) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in VR-OffShell

Figure 8.14: $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ for various signal hypotheses (lines), compared to the SM background (stacked, filled histograms) in the (a) SR-OffShell and (b) VR-OffShell regions. The values for the experimental data are also shown (black dots) in (b). In (a), the lower panel shows the calculated Z-value as a function of lower cut, i.e., events to the right of a given cut in the upper panel are kept. The relative uncertainty on the background ($\Delta B/B$) used in the calculation is 30%. At least three signal events and at least one background event is required in order for a given cut/bin to get a non-zero Z-value. (The Z-values shown in the lower panel for bins to the left of the SR, i.e., $E_{\mathrm{T}}^{\mathrm{miss,sig}} < 9$, are superfluous, as they all get the same integrated number of signal and background events as the left-most bin in the SR.) In (b), the statistical uncertainty of the data points are indicated by black error bars, whereas the statistical uncertainty on the combined SM backgrounds are shown by hatched areas. The lower panel shows the ratio of the number of data to background events for each bin in the upper panel.



(a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in CR-DY (b) $m_{\ell\ell}$ in CR-DY

Figure 8.15: (a) $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and (b) $m_{\ell\ell}$ in data (black dots) compared to the SM background (stacked, filled histograms) in the CR-DY region. The statistical uncertainty of the data points are indicated by black error bars, whereas the statistical uncertainty on the combined SM backgrounds are shown by hatched areas. The lower panels show the ratio of the number of data to background events for each bin in the upper panel.

8.4.5 Discovery regions

The signal regions described in the previous sections are optimized for sensitivity to the specific simplified signal models that motivate the search. In particular, single variables, such as $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and $m_{\ell\ell}$, are split in multiple bins to exploit the shape of the specific signal models. They are therefore meant for setting signal specific exclusion limits, if no significant excess is found in the data.

However, first we need to check whether the observed data are consistent with the background expectation or not. In that case, we want to look for any deviation in the total number of observed events compared to the expected SM background, without making any assumptions on the shape of the deviation. Therefore, we can only use single-bin regions, with a simple count of observed and expected events in the region.

We call these regions discovery regions, and they are based on the signal regions described earlier. The discovery regions are defined in Table 8.5.

8.5 Systematic uncertainties

In addition to the statistical uncertainties, related to the limited number of events in the observed data and the estimated backgrounds, we also have to account for other possible sources of uncertainties, related to the theoretical models being applied in the simulation of background and signal processes, and

Table 8.5: The model-independent discovery regions and which signal regions they are based on.

Discovery region	**Derived from**
DR-High	SR-High-8 (with $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ bins combined)
DR-Int	SR-Int (with $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ bins combined)
DR-Low	SR-Low (with $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ bins combined)
DR-OffShell	SR-OffShell (with $m_{\ell\ell}$ bins combined)

uncertainties in scales and resolutions of the experimental measurements and event reconstruction.

8.5.1 Theory uncertainties

8.5.1.1 Renormalization and factorization scales

The inclusive proton-proton cross-section into any final state, $pp \to X$, to nth order in perturbation theory, can be factorized [177] into a *perturbative QCD* (pQCD) part, $\hat{\sigma}^{(n)}$, dealing with the high-energy (short-range) partonic interactions between the quarks and/or gluons inside the protons, and a non-perturbative part, given by the *Parton Density Functions* (PDFs), describing the distribution of quarks and gluons as function of longitudinal proton momentum fraction, x, dealing with the lower energy (long-range) hadronic interactions between the protons:

$$\sigma^{(n)} = PDF(x_1, \mu_{\mathrm{F}}) \otimes PDF(x_2, \mu_{\mathrm{F}}) \otimes \hat{\sigma}^{(n)}(x_1, x_2, \mu_{\mathrm{R}}), \tag{8.5}$$

where

$$\hat{\sigma}^{(n)} = \hat{\sigma}^{(0)} + \alpha_s \hat{\sigma}^{(1)} + \alpha_s^2 \hat{\sigma}^{(2)} + \ldots + \alpha_s^n \hat{\sigma}^{(n)} + \mathcal{O}(\alpha_s^{n+1}). \tag{8.6}$$

The PDFs capture all low-energy (long-distance) radiation up to some factorization scale, μ_{F}, whereas the perturbative expansion in α_s, which is small for high-energy (short-distance) interactions, captures the contributions above μ_{F}. The hard scatter cross-section, $\hat{\sigma}^{(n)}$, is given in terms of $\alpha_s(\mu_{\mathrm{R}})$, which has been renormalized at some renormalization scale, μ_{R}.

The PDFs are given at a certain energy scale, Q, but can be translated to other energy scales by the DGLAP[1] equations [178, 179, 180]:

$$\frac{dPDF(x, Q^2)}{d \log Q^2} = P(\alpha_s, z) PDF(x/z, Q^2), \tag{8.7}$$

where

$$P(\alpha_s, z) = \alpha_s P^{(\mathrm{LO})} + \alpha_s^2 P^{(\mathrm{NLO})} + \ldots \tag{8.8}$$

[1] Dokshitzer–Gribov–Lipatov–Altarelli–Parisi

is a splitting function, expanded in a series of increasing number of quark and/or gluon splits, e.g., $q \to q + g$, where z is the momentum fraction of the initial state quark taken by the final state quark, and the gluon takes the remaining fraction, $1 - z$.

The strong coupling constant, α_s, is usually specified in terms of experimental measurements taken at $Q = m_Z$, where MC generators run the coupling to the energy scale of a given generated process by using the *Renormalization Group Equation* (RGE):

$$\frac{d\alpha_s(Q)}{d\log Q^2} = -\alpha_s^2(Q)[b_0 + b_1\alpha_s(Q) + \ldots], \tag{8.9}$$

where b_0 and b_1 are model-dependent constant coefficients. See, e.g., chapter 9 of Reference [181].

The cross-section, the PDFs and the running of α_s are all expanded in series of α_s, and the truncation of these series at nth order therefore neglects contributions of order α_s^{n+1}. Since α_s is a function of the energy scale, we can estimate the systematic uncertainties arising from these truncations by varying the renormalization and factorization scales, μ_{R} and μ_{F}, up and down by a factor of 2, both individually and together:

$$\{\mu_{\mathrm{R}}, \mu_{\mathrm{F}}\} \times \{0.5, 0.5\}, \{1, 0.5\}, \{0.5, 1\}, \{1, 1\}, \{2, 1\}, \{1, 2\}, \{2, 2\}. \tag{8.10}$$

The result of each of the variations above are evaluated in terms of the cross-section (or number of events) in each bin, i.e., CR, VR and SR, of the analysis. The uncertainty from energy scale dependence is taken as the envelope of all scale variations in each bin/region, i.e.

$$\delta^{\mu_{\mathrm{R}},\mu_{\mathrm{F}}}\sigma = \max_i[|\sigma(\mu_{\mathrm{R},i}, \mu_{\mathrm{F},i}) - \sigma(\mu_{\mathrm{R},0}, \mu_{\mathrm{F},0})|], \tag{8.11}$$

where i represents an index on the set of scale variations, and $\mu_{\mathrm{R},0}$ and $\mu_{\mathrm{F},0}$ represent the nominal renormalization and factorization scales, respectively.

8.5.1.2 Choice of parton distribution function

In addition to the missing higher order uncertainties in the DGLAP evolution, which are estimated by the scale variations described above, there are also uncertainties tied to how the PDFs are estimated from fits to experimental data.

There exist multiple sets of PDFs obtained by different collaborations [101]. They have obtained their estimates from fits to different experimental datasets, each with their own experimental uncertainties. There are also systematic uncertainties related to the functional form of these fits.

MC samples are generated with a certain choice of PDF, but typically also come with event weights that can be applied to the nominal (unweighted) events in order to obtain the outcome of using other (variations of) PDFs. The systematic uncertainty arising from different choices of PDFs is estimated by

the standard deviation from the average in the ensemble of the N_{mem} variations provided [101]:

$$\delta^{PDF}\sigma = \sqrt{\frac{1}{N_{\text{mem}}-1}\sum_{k=1}^{N_{\text{mem}}}\left(\sigma^{(k)} - \langle\sigma\rangle\right)^2}, \tag{8.12}$$

where the average over all the N_{mem} PDF variations is calculated by

$$\langle\sigma\rangle = \frac{1}{N_{\text{mem}}}\sum_{k=1}^{N_{\text{mem}}}\sigma^{(k)}. \tag{8.13}$$

8.5.1.3 Implementation of theory uncertainties

The theory uncertainties w.r.t. variations in (renormalization and factorization) scale and choice of PDF are calculated from so-called LHE3 (Les Houches event version 3) [182, 183] weights provided with the MC samples. There are separate LHE weights for every systematic up and down variation.

Systematic uncertainties in scale and PDF are estimated on a region-by-region basis for the three main sources of backgrounds: diboson, $t\bar{t}$ and $Z+\text{jets}$. The uncertainties are expressed as fractional uncertainties in terms of their difference from the nominal value in the corresponding region divided by the nominal value.

Since the low-mass DY process is a small background to our regions in general, we have used nearby $Z+\text{jets}$ uncertainties to estimate the size of the theory systematics for low-mass DY in the off-shell regions.

The smaller backgrounds, where the estimates are taken directly from MC simulations, are assigned global uncertainties on their total production cross-sections. All Higgs boson samples are assigned a conservative 10% uncertainty, based on the $t\bar{t}H$ cross-section uncertainty in Reference [184]. A 13% uncertainty is similarly taken from Reference [184] and applied to the Top Other samples, which consist predominantly of $t\bar{t}V$. The triboson (VVV) are assigned a 32% uncertainty based on comparisons between Sherpa and VBFNLO in Reference [185].

8.5.2 Experimental uncertainties

Comparisons are carried out between data measurements and corresponding predictions from simulation, and uncertainties are quantified to account for the discrepancies. In large, we have three kinds of experimental uncertainties: (i) discrepancies in kinematic variables, i.e., energy scale and resolution of the physics objects; (ii) discrepancies in efficiencies, related to the proportion of physics objects which are reconstructed, identified, isolated and triggered; and (iii) uncertainties tied to data-driven estimates, when MC simulations do not model the experimental data sufficiently well.

8.5.2.1 Uncertainties on kinematic variables

Kinematic uncertainties are related to the experimental measurements and the subsequent reconstruction of the energies and momenta of objects, where up and down variations of each source of uncertainty are implemented as separate versions of the dataset (separate ntuples/TTrees). This includes uncertainties related to the energy scale and resolution of, most prominently, jets, but also electrons/photons and muons, uncertainties related to the identification of these objects, and also uncertainties in the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft track terms, which are not matched to any reconstructed object.

Integrated luminosity The uncertainty on the integrated luminosity measurements for the full ATLAS Run 2 dataset, after all data quality requirements have been applied, amounts to 1.7% [49]. This uncertainty is included as a nuisance parameter in the likelihood fit for processes that are not normalized to data.

Jet energy scale (JES) and resolution (JER) The four-momenta of the reconstructed jets, both in data and MC, go through a series of calibration steps in order to better match the reconstructed four-momenta at particle-level in simulation. In addition, so-called *in situ* measurements of jets in both data and MC are made after the simulation-based calibrations, where remaining differences between data and MC are corrected by corresponding calibrations to jets in data only.

The modeling of jets in simulation, and the numerous *in situ* measurements and calibrations, all bring sources of statistical and systematic uncertainties. In total, 125 different uncertainty terms have been quantified for the full JES uncertainty [186], where several of the terms are uncertainties in bins of p_{T} and/or η. Since several of these terms are negligible to some analyses, several reduced uncertainty schemes have been derived to allow for simplified descriptions with a minimum loss of correlation information.

98 of these terms, or nuisance parameters, are p_{T}-dependent uncertainties from the absolute *in situ* analyses, and can be reduced by way of an eigenvector decomposition of the covariance matrix of these components. The resulting eigenvectors with the largest eigenvalues are kept as new, effective nuisance parameters, where the remaining terms are combined into a single, residual nuisance parameter.

We make use of the so-called *category reduction*, which combines the p_{T}-dependent *in situ* uncertainty components in terms of which of the following categories they belong to: detector, statistical, modeling or mixed. This reduces the *in situ* uncertainties to 15 effective nuisance parameters, and 30 JES nuisance parameters in total. The final 15 JES parameters are related to intercalibrations between η-regions, jet flavor composition and response, various contributions from pile-up and modeling of "punch-through" of jets beyond the hadronic calorimeter in MC.

The JER uncertainty, relative to the jet p_{T}, can for calorimeter-based resolution be parametrized by three terms:

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \frac{N}{p_{\mathrm{T}}} \oplus \frac{S}{\sqrt{p_{\mathrm{T}}}} \oplus C, \tag{8.14}$$

where N, S and C are parameters to be measured or fitted, and represent noise, stochastic and constant terms, respectively. The noise term consists of contributions that are independent of p_{T}-scale, such as electronic noise to the measured signal from the detector front-end electronics, and contributions from pile-up. The stochastic term contains contributions from statistical fluctuations, which scales as $1/\sqrt{p_{\mathrm{T}}}$. The constant term contains fluctuations which are constant fractions of the p_{T}, such as energy depositions in passive material, the starting point of the hadron showers and the non-uniformities of response across the calorimeter.

Similar eigenvalue decompositions are also carried out to reduce the number of JER uncertainty components. We make use of the so-called *SimpleJER* reduction, where the 34 original nuisance parameters are reduced to 8 effective nuisance parameters.

The p_{T} of the jets in MC are smeared to match the JER in data, and the smearing is applied differently for each systematic (nuisance parameter) variation according to the following expression:

$$\sigma^2_{\mathrm{smear},i} = (\sigma_{\mathrm{nom}} + |\sigma_{\mathrm{NP},i}|)^2 - \sigma^2_{\mathrm{nom}}, \tag{8.15}$$

where σ_{nom} is the nominal JER and $\sigma_{\mathrm{NP},i}$ represents the systematic variation by nuisance parameter NP_i. When the JER is smaller in data than in MC, the difference is accounted for as an additional nuisance parameter:

$$\sigma_{\mathrm{NP,\ data\text{-}MC}} = \sigma^{\mathrm{data}}_{\mathrm{nom}} - \sigma^{\mathrm{MC}}_{\mathrm{nom}}. \tag{8.16}$$

$\boldsymbol{E}^{\mathbf{miss}}_{\mathbf{T}}$ track soft term (TST) As the systematic variations applied to the reconstructed, hard objects in the events also get propagated to the reconstructed $E^{\mathrm{miss}}_{\mathrm{T}}$, the effect of systematic uncertainties from hard objects on $E^{\mathrm{miss}}_{\mathrm{T}}$ are taken care of simultaneously. However, the soft tracks not associated to reconstructed objects, fall outside the scope of the object-based systematic variations. Therefore, separate track soft term (TST) systematics need to be added.

Since

$$\boldsymbol{p}^{\mathrm{miss}}_{\mathrm{T}} = -(\boldsymbol{p}^{\mathrm{hard}}_{\mathrm{T}} + \boldsymbol{p}^{\mathrm{soft}}_{\mathrm{T}}) \tag{8.17}$$

an event with no $\boldsymbol{p}^{\mathrm{miss}}_{\mathrm{T}}$ should give

$$\boldsymbol{p}^{\mathrm{soft}}_{\mathrm{T}} = -\boldsymbol{p}^{\mathrm{hard}}_{\mathrm{T}}. \tag{8.18}$$

In practice, however,the measurements of these quantities will not be completely equal, due to detector resolution effects.

In order to fully cover these effects, three components are considered [175]: (i) *the parallel scale*, $p_{||}^{\text{soft}}$, i.e., the mean value of the parallel projection of $\boldsymbol{p}_{\text{T}}^{\text{soft}}$ along $\boldsymbol{p}_{\text{T}}^{\text{hard}}$; (ii) *the parallel resolution*, $\sigma_{||}$, i.e., the root-mean-square (RMS) of $p_{||}^{\text{soft}}$; and (iii) *the transverse resolution*, $\sigma_{\perp}$, i.e., the root-mean-square (RMS) of the perpendicular component of $\boldsymbol{p}_{\text{T}}^{\text{soft}}$ with respect to $\boldsymbol{p}_{\text{T}}^{\text{hard}}$. The transverse scale is not considered, as it has been shown to be consistent with zero in both data and MC.

These quantities are measured in bins of $p_{\text{T}}^{\text{hard}}$ both in data and multiple combinations of MC generators and parton shower models, and the maximal difference (envelope) between any of the MC models and data are taken as a measure of the corresponding systematic uncertainty.

The resolution uncertainties are applied to MC by smearing each event's $p_{||}^{\text{soft}}$ and $p_{\perp}^{\text{soft}}$ by Gaussians of width equal to $\sigma_{||}^{\text{soft}}$ and $\sigma_{\perp}^{\text{soft}}$, respectively. The scale variations on $p_{||}^{\text{soft}}$ are applied as $p_{||}^{\text{soft}} \pm \Delta_{\text{L}}$, where Δ_{L} represents the maximal difference between data and MC in the corresponding $p_{\text{T}}^{\text{hard}}$-bin.

Electrons and photons Electron and photon objects are reconstructed from clusters of energy depositions in the EM calorimeter, where an electron has a reconstructed track matched to the cluster, whereas a photon does not. Corrections to the electron and photon energies are added on top of the measurements from the cluster depositions, in order to take into account energy lost to material placed before the EM calorimeter, to the neighboring calorimeter cells of the reconstructed cluster and energy lost beyond the EM calorimeter [147, 187]. To this end, an algorithm trained on simulated $Z \to ee$ events are used to calibrate the absolute energy scale of both data and MC. In order to reduce remaining differences between data and MC, additional corrections are applied to data, such as intercalibrations of calorimeter layers, corrections to pile-up induced energy-shifts and improvements to the uniformity of the energy response.

Residual differences in energy scale between data and MC, in multiple bins of η indexed by i, are parametrized in terms of α_i as

$$E^{\text{data}} = E^{\text{MC}}(1 + \alpha_i), \tag{8.19}$$

and the differences in energy resolution, in terms of c_i, as

$$\left[\frac{\sigma(E)}{E}\right]^{\text{data}} = \left[\frac{\sigma(E)}{E}\right]^{\text{MC}} \oplus c_i, \tag{8.20}$$

where $\oplus$ represents a sum in quadrature. After extracting fitted values of α_i and c_i on the $Z \to ee$ invariant mass spectrum of data and MC, the energy scale in data is corrected to match MC by use of α_i, whereas MC is corrected to match the energy resolution of data by use of c_i.

Multiple systematic uncertainties [147, 187] are evaluated by how much they affect the values of α_i and c_i. This includes uncertainties tied to the method, the mass range, the event and object selection used to extract the parameters, binned in both E_{T} and η.

In total, more than 60 systematic variations are available related to the energy calibration. However, since the systematic uncertainties from electrons and photons are typically negligible in the SUSY searches presented here, we make use of a simplified correlation model, where only two nuisance parameters are considered, one for energy scale and one for energy resolution. For the simulated SUSY processes, an additional nuisance parameter modeling the uncertainty in the energy scale for the fast simulation of the calorimeter response is accounted for.

Muons The final muon objects, also referred to as *combined muons* (CB), are reconstructed using hits in both the inner detector (ID) and the muon spectrometer (MS). A muon momentum calibration is applied to CB muons in MC, in order for the $p_{\mathrm{T}}^{\mathrm{MC,ID}}$- and $p_{\mathrm{T}}^{\mathrm{MC,MS}}$-distributions of simulated muons to match the corresponding distributions of muons in data, $p_{\mathrm{T}}^{\mathrm{data,ID}}$ and $p_{\mathrm{T}}^{\mathrm{data,MS}}$, respectively.

The corrected MC momentum for CB muons, $p_{\mathrm{T}}^{\mathrm{cor,CB}}$, is obtained from the following combination of corrected ID- and MS-track momenta:

$$p_{\mathrm{T}}^{\mathrm{cor,CB}} = f \cdot p_{\mathrm{T}}^{\mathrm{cor,ID}} + (1-f) \cdot p_{\mathrm{T}}^{\mathrm{cor,MS}}, \tag{8.21}$$

where the relative fraction, f, is constrained to be equal to the corresponding fraction before the correction:

$$p_{\mathrm{T}}^{\mathrm{MC,CB}} = f \cdot p_{\mathrm{T}}^{\mathrm{MC,ID}} + (1-f) \cdot p_{\mathrm{T}}^{\mathrm{MC,MS}}. \tag{8.22}$$

The corrected momentum, $p_{\mathrm{T}}^{\mathrm{cor,CB}}$, is calculated as a function of the η and ϕ of the simulated muons, and includes correction to both the momentum scale and resolution. See Equation 5 of Reference [138] for the full expression.

The momentum scale correction consists of separate ID and MS terms that correct for inaccuracies in the description of the magnetic field integral and the dimension of the detectors in the direction perpendicular to the magnetic field. An additional term is added to account for inaccuracies in the MC modeling of energy loss in material between the interaction point and the MS.

In order to account for differences in the momentum resolution found in data and MC, the MC momenta are smeared using the following parametrization of the relative p_{T}-resolution:

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \frac{r_0}{p_{\mathrm{T}}} \oplus r_1 \oplus r_2 \cdot p_{\mathrm{T}}, \tag{8.23}$$

where $\oplus$ denotes a sum in quadrature. The parameter of the first term, r_0, accounts mainly for fluctuations of the energy loss in the traversed material. The second parameter, r_1, accounts for the contributions from multiple scattering, local magnetic field inhomogeneities and local, radial displacements of the hits. The third and final parameter, r_2, describes intrinsic resolution effects caused by residual misalignment of the MS. The first term is found to be negligible both for the ID and the MS.

Systematic variations (nuisance parameters) are provided for the momentum resolution of the ID and the MS separately, and for the momentum scale combined. Two additional systematic variations are also added to account for a charge-dependent sagitta[2] bias.

8.5.2.2 Uncertainties on efficiency corrections

Efficiency uncertainties are related to the scale factor corrections that account for differences in the reconstruction efficiencies of data and MC, where up and down variations of each source of uncertainty are implemented as weights to be applied to the nominal dataset. This includes uncertainties in the pileup reweighting procedure, the jet-to-vertex tagger, jet flavor tagging, trigger efficiencies, and the identification, isolation and reconstruction efficiencies of electrons and muons.

The *tag-and-probe* method is used for measuring efficiencies of quantities such as object reconstruction, identification, isolation and triggering. It is based on selecting a high-quality (tight selection) *tag* object along with a relatively unbiased (loose selection) *probe* object to be used for evaluation of the selection in question.

Take for example a measurement of the efficiency of a single-electron trigger. First, we select a high-quality electron as tag, which is matched to the single-electron trigger in question and passes stringent identification and isolation requirements. Then, we need to have a second electron in the event, to be the probe, with a looser set of requirements when it comes to identification as an electron and isolation. The two electrons additionally need to satisfy a collective requirement, such as having opposite electric charge and making up an invariant mass consistent with a resonance, e.g., Z or J/ψ. If these requirements hold, the probe should be comparable with the tag, and thereby also, in principle, have fired the same trigger. The trigger is most likely not 100% efficient, i.e., some of the probes will fail to fire the trigger in question for various reasons. The efficiency of the trigger can therefore be calculated by dividing the number of probes that fired the trigger by the number of tags, which are necessarily required to have fired the trigger:

$$\epsilon_{\text{trig}} = \frac{N_{\text{trig}}^{\text{probe}}}{N_{\text{trig}}^{\text{tag}}}. \tag{8.24}$$

Jet-vertex-tagger The *jet-vertex-tagger* (JVT) [188] and the *forward-JVT* (fJVT) [189] are algorithms designed to measure how likely it is that a jet belongs to the primary vertex (hard-scatter) of the event, or conversely most likely belongs to a pile-up interaction. The efficiencies of these discriminators are measured using tag-and-probe on Z + jets events, with the leading jet as the probe. Differences in the efficiencies in data and MC are corrected for by

[2]The sagitta of a curved particle track is a measure of how much the curve deviates from a straight line at its center, in terms of the orthogonal distance from the imagined straight line to the center of the arc.

applying scale factors to the MC events. Systematic uncertainties have been determined by way of one nuisance parameter for JVT and one for fJVT.

***b*-tagging** A (gradient) boosted decision tree, called MV2, is applied to all selected jets in order to identify jets from b-hadrons. This model is trained on simulated $t\bar{t}$ events with at least one reconstructed lepton, in order to target $t \to Wb$, where the W decays leptonically, while also keeping a sufficiently large fraction of c-jets. The b-tagging efficiency is measured [190] both in MC, making use of MC truth information before and after jet reconstruction, and in data, by performing a simultaneous fit to dileptonic $t\bar{t}$ regions enriched in bb, bq, qb and qq jets (quark-anti-quark pairs), where q represents a light flavor jet (u, d, s, c) and the jet pair is p_{T}-ordered. The bb-selection works as a signal region in which to measure the b- and light-flavor-jet efficiencies, and the other three regions work as control regions to constrain the light-flavor-jet contributions. The fraction of c-jets in the light-flavor-jet contribution is taken from MC.

The identification of b-hadrons therefore contains statistical uncertainties from the limited number of MC events used for the measurements, and from the fit of the b-, c- and light-flavor-jet contributions to the bb-enriched region in data.

Also systematic uncertainty variations from the kinematics of jets, leptons, pile-up, etc., are propagated through the analysis and quantified in terms of how each of them affect the b-tagging efficiency. There are also systematic contributions from physics modeling and theory uncertainties related to the MC simulations, and from fake and non-prompt lepton contributions in the data measurements.

An eigenvector decomposition of the total covariance matrix is performed in order to reduce the number of effective nuisance parameters, down to up and down variations on the b-, c- and light-flavor efficiencies, and on the extrapolation of the c- and light-flavor-jet contributions to the bb-enriched region.

Electrons The efficiency of the electron reconstruction, identification, isolation and charge identification are measured using tag-and-probe in $Z \to ee$ and $J/\psi \to ee$ events. Differences in the efficiencies measured in data and MC, in bins of p_{T} and η, are corrected for by applying scale factors to events in MC containing electrons. Systematic uncertainties are derived by varying different aspects of the method used to measure the efficiencies, such as event and object selection and background modeling. Since previous iterations of the analysis presented here have shown negligible impact from lepton systematic uncertainties, we only consider a simplified set of six nuisance parameters.

There are also two additional nuisance parameters covering differences in the dielectron trigger efficiencies between data and MC. These have been obtained by varying the quality requirements on the tag and probe muons, and the requirements on events to be considered for the background subtraction method used in the efficiency measurements [191].

Muons Variations of the tag-and-probe method are used to measure the efficiencies of the muon reconstruction, identification, isolation and vertex association, separately, in $Z \to \mu\mu$ and/or $J/\psi \to \mu\mu$ events. Differences in efficiencies between data and MC are corrected by applying scale factors to the MC events containing muons. Systematic uncertainties from the determination of these efficiencies are quantified by varying different aspects of the methods used to measure the efficiencies, such as event and object selection, uncertainties from background estimation, and so on. We consider eight nuisance parameters related to muon reconstruction, identification, isolation and vertex association in our analysis.

Additionally, there are two nuisance parameters quantifying uncertainties related to dimuon trigger matching efficiencies in data and MC. The uncertainties are derived by, e.g., considering two different pile-up selections, correlations between tag-and-probe muons and the effect of varying the Z-mass window by $\pm 5\,\text{GeV}$ [192].

8.5.3 Uncertainties in data-driven fake/non-prompt lepton estimates

The MC simulations do not model the fake and non-prompt (FNP) lepton contributions in data sufficiently accurate, and the so-called Matrix Method (MM) is used to extract these contributions from data instead. Using the tag-and-probe method, four control regions are defined to measure: (i) the rate of prompt leptons passing the baseline selection that also pass the signal selection, referred to as the *real efficiency*, and the rate of non-prompt leptons from (ii) decays of heavy-flavor jets, (iii) photon conversions into e^+e^- and (iv) decays of light-flavor jets or light-flavor jets incorrectly identified as (fake) electrons, passing the baseline selection that also pass the signal selection, collectively referred to as *fake rates*. Given that prompt leptons are well modeled by MC, and the difficulty of constructing a control regions sufficiently pure in light-flavor FNP leptons, the efficiencies/rates of these regions are taken directly from MC, using truth matching. The efficiencies/rates are measured separately for electrons and muons, and in bins of p_T and η. The contribution from each of the three sources of FNP leptons mentioned above to the so-called *fake weights*, are finally measured in signal-like regions for each of the following scenarios: off-shell, low-, intermediate- and high-Δm. Contamination from prompt/real leptons to each of the control regions are estimated by MC, and subtracted from the total to give the FNP leptons estimate in data.

Several kinds of uncertainties come along with the estimates described above. The MC scale factors, to correct for differences in reconstruction, identification and isolation efficiencies between data and MC, have associated uncertainties that are taken into account for the FNP estimates. The real lepton contamination in the control regions used to extract the fake rates, are subtracted from the region, and a 10% systematic uncertainty is assigned to the subtracted MC yields. For the fake weights, the relative fraction of each of the three FNP sources are varied up and down by one standard deviation, and if a given event enters multiple

control regions, the minimum and maximum (envelope) weights in these regions are used as uncertainty for the given event. The statistical uncertainties on the real efficiency and the fake rates are also propagated through to the FNP estimates. Finally, all sources of uncertainty are combined to give one nuisance parameter which is used in the likelihood fit.

8.5.4 Uncertainties implemented in HistFitter

HistFitter creates histograms for each of the up and down variations described above, either directly from each ntuple for the kinematic uncertainties, or by multiplying the nominal ntuple with the different sets of weights for the efficiency uncertainties. We then obtain estimates for how the number of events in each analysis bin/region changes with respect to every systematic variation, i.e., how large effect each of these variations have on our analysis.

8.6 Statistical analysis

8.6.1 The likelihood function

In HistFitter, the likelihood function is built from a product of Poisson factors, P, for each bin/region, i, considered in the fit, along with factors, C, constraining the nuisance parameters representing the systematic uncertainties:

$$L(\boldsymbol{n}, \boldsymbol{\theta}^0 | \mu_{\text{sig}}, \boldsymbol{\mu}_p, \boldsymbol{\theta}) = \mathcal{P}_{\text{SR}} \times \mathcal{P}_{\text{CR}} \times \mathcal{C}_{\text{syst}} \tag{8.25}$$

$$= \prod_{i \in \text{SR}} P(n_i | \lambda_i(\mu_{\text{sig}}, \boldsymbol{\mu}_p, \boldsymbol{\theta})) \tag{8.26}$$

$$\times \prod_{i \in \text{CR}} P(n_i | \lambda_i(\mu_{\text{sig}}, \boldsymbol{\mu}_p, \boldsymbol{\theta})) \tag{8.27}$$

$$\times C_{\text{syst}}(\boldsymbol{\theta}^0, \boldsymbol{\theta}). \tag{8.28}$$

The likelihood is a function of the vector, $\boldsymbol{n}$, of observed data yields (counts) in each CR and SR, n_i, considered in the fit, and of the vector, $\boldsymbol{\theta}^0$ consisting of auxiliary measurements of the nuisance parameters. This construct gives a measure of the likelihood for the observed data $(\boldsymbol{n}, \boldsymbol{\theta}^0)$ given the fit parameters of the model $(\mu_{\text{sig}}, \boldsymbol{\mu}_p, \boldsymbol{\theta})$, where $\lambda_i(\mu_{\text{sig}}, \boldsymbol{\mu}_p, \boldsymbol{\theta})$ is the expected event count in bin/region i. The event count λ_i is a function of μ_{sig}, which is a scale factor representing the signal strength; $\boldsymbol{\mu}_p$, which is a vector holding normalization factors for the most important background processes (p); and $\boldsymbol{\theta}$, which is a vector holding the location/size of the nuisance parameters.

The nuisance parameters, θ_j, are typically considered to be Gaussian distributed around their expectation values, θ_j^0, with standard deviations, $\pm\sigma_{\theta_j}$, represented by $\theta_j = \pm 1$ in the standardized Gaussian factors which constrain the corresponding nuisance parameters in the likelihood fit. In the case where all nuisance parameters are taken to be independent, C_{syst} is simply the product of Gaussian factors, G, for each systematic uncertainty, j, in the set of all

systematic uncertainties considered, S, having unit widths:

$$C_{\text{syst}}(\boldsymbol{\theta}^0, \boldsymbol{\theta}) = \prod_{j \in S} G(\theta_j^0 - \theta_j). \tag{8.29}$$

8.6.2 Maximum likelihood fit to data

8.6.2.1 The profile likelihood ratio

For the hypothesis tests, we follow the LHC standard of choosing the profile likelihood ratio to be our test statistic:

$$q_{\mu_{\text{sig}}} = -2 \ln \frac{L(\mu_{\text{sig}}, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}_{\text{sig}}, \hat{\boldsymbol{\theta}})} \tag{8.30}$$

In order to calculate the test statistic, we need to perform two separate maximum likelihood fits. To get the denominator, $L(\hat{\mu}_{\text{sig}}, \hat{\boldsymbol{\theta}})$, we allow all model parameters to float, in order to obtain the maximum likelihood estimates (MLEs) $\hat{\mu}_{\text{sig}}$ and $\hat{\boldsymbol{\theta}}$, given the data. To get the numerator, $L(\mu_{\text{sig}}, \hat{\hat{\boldsymbol{\theta}}})$, we keep the signal strength μ_{sig} fixed at a constant value depending on the type of fit, and let the remaining model parameters float to find the profiled MLEs of the nuisance parameters, $\hat{\hat{\boldsymbol{\theta}}}$, given the data and the specific value of μ_{sig}.

8.6.2.2 Test for discovery

The first hypothesis we would like to test, is whether the observed data is consistent with the test statistic distribution, given only contributions from SM background processes, $f(q_0|0)$, at some pre-determined confidence level (CL). This is called the background-only hypothesis. In this case the signal strength parameter, μ_{sig}, is equal to 0, which gives the following version of the test statistic:

$$q_0 = -2 \ln \frac{L(0, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}_{\text{sig}}, \hat{\boldsymbol{\theta}})}. \tag{8.31}$$

The compatibility of the observed data with the background-only hypothesis can be quantified in terms of the integral over $f(q_0|0)$ for more extreme (higher) values than the observed value, $q_{0,\text{obs}}$. I.e., the p-value of getting a test statistic at least as far away from the expectation value as what is observed, given only background contributions. This can be formulated mathematically as:

$$p_0 = \int_{q_{0,\text{obs}}}^{\infty} f(q_0|0)\, dq_0. \tag{8.32}$$

If $p_0 < 2.87 \times 10^{-7}$, which corresponds to at least a $5\,\sigma$ one-sided Gaussian deviation from the expectation, we discard the background-only hypothesis, as is conventional in high-energy physics experiments. Otherwise, we cannot claim

Table 8.6: Post-fit normalization factors from a background-only fit to data in the CRs and SRs, including all systematic uncertainties.

Norm factor	Post-fit value
μ_{Z}	1.2 ± 0.2
μ_{DY}	0.9 ± 0.2
μ_{VZ}	0.92 ± 0.08
μ_{tt}	0.99 ± 0.08

discovery of a new signal according to our standards, and we will instead proceed to set corresponding exclusion limits on the cross-section and/or mass of any new particle involved, given the observed data.

Parameters of the model-independent fit Each of the four control regions CR-VZ, CR-tt, CR-Z and CR-DY are designed to contain events of high purity in their respective background processes: diboson, $t\bar{t}$, $Z+\text{jets}$ and $Z^*/\gamma^*+\text{jets}$. Their objective is to constrain the floating normalization factors in simultaneous fits to data in all CRs, for each of these four processes: μ_{VZ}, μ_{tt}, μ_{Z} and μ_{DY}.

Figure 8.16 shows the post-fit results for all the floating parameters in a fit to data in all CRs and SRs, including the profiled uncertainties on these parameters. There are four categories of parameters: (i) background normalization factors, represented by the letter μ, (ii) uncertainty on the luminosity measurement, labeled "Lumi", (iii) statistical uncertainties in the fitted regions, represented by the letter γ, and (iv) the size of the theoretical and experimental systematic uncertainties in the fitted regions, represented by the letter α.

Table 8.6 shows the resulting normalization factors obtained by the fit. The post-fit values show that all four normalization factors are consistent with unity within uncertainties, which thereby indicate that no significant mis-modeling is found in the initial normalization of the dominant backgrounds.

Since the background normalization factors are consistent with one, and there are reasonably large numbers of events in the control regions, the size of the statistical uncertainties in the control regions are not significantly affected by the fit. The statistical uncertainties in the signal regions are not significantly altered either, but there are larger uncertainties associated with the corrections from the fit due to lower statistics in the SRs compared to the CRs. This is reflected in the sizes of the error bars of the SR γ factors in the pull plot. The most striking pull among the statistical uncertainties comes from the SR-Low-2 region. Large statistical uncertainty from low statistics and highly weighted $Z+\text{jets}$ events in this region, combined with the post-fit normalization factor of 1.2 for $Z+\text{jets}$, leads to a decrease in the relative statistical uncertainty in the region. Even though the relative statistical uncertainty is reduced by about 30% after the fit, the uncertainty in this reduction is still large enough to include unity in the $1\,\sigma$ interval.

The majority of the remaining systematics, associated with theoretical predictions and event kinematics, are also mostly unaffected by the fit, and are

consistent with their pre-fit estimates within one standard deviation. There are, however, some systematics that are more affected than others, most notably those that parametrize the jet energy resolution and scale (JER and JES, respectively), the e/γ energy scale, the electron charge-ID, the FNP-weights and the jet flavor composition.

Both the JES and JER have some parameters which are pulled to lower and others which are pulled to higher values. Sensitivity to the jet related parameters comes from the various jet selections applied to the control and signal regions. The e/γ energy scale and the electron charge-ID parameters are all pulled to lower values, which is also the case for the FNP-weight, whereas the jet flavor composition is pulled upwards. In order to get some clues to the causes of these pulls, it may be informative to look at their correlations with other parameters.

Correlations of the fitted parameters Figure 8.17 shows a correlation matrix between the most dominant systematic uncertainties in the fit. The α-parameters hardly show any correlation among themselves, which is by design. The obvious exceptions are the complete self-correlations shown on the diagonal.

Moderate anti-correlations can be seen between the normalization factors. This is a result of the same backgrounds contributing to multiple CRs, where the normalization factor of one process may decrease when another one increases in order to fit the data.

There are also moderate anti-correlations seen between the normalization factors and the statistical uncertainties in the CRs. Increased background contribution results in decreased relative statistical uncertainty in the CRs.

The fake and non-prompt weight parameter is negatively correlated with all the normalization factors, due to becoming less important relative to the total background when other background sources become more important.

The jet flavor composition parameter turns out to be positively correlated with the $Z+$jets normalization factor, where the value of the latter is largely controlled by the fit of the $Z+$jets background to data in CR-Z. This CR makes use of a b-jet veto in order to reject top backgrounds. Since the initial background estimates from MC in CR-Z underestimated the data in the region, the up-scaling of the $Z+$jets background driven by this region, may therefore also have inflated the effect of the heavy flavor jet contributions in the region.

A rather large number of moderate correlations is observed between the statistical uncertainty in the SR-Low-2 region with various kinematic parameters. This is most likely due to the large statistical uncertainty in the region, caused by low number of events and some highly weighted $Z+$jets events. These correlations are therefore not of great significance.

Numbers of events after the fit Figure 8.18 shows the post-fit background estimates compared to the observed data in all control, validation and signal regions, after a simultaneous fit to the CRs and SRs have been carried out. The control region yields show a successful fit of the backgrounds to the data, but with a small excess of observed events in CR-Z.

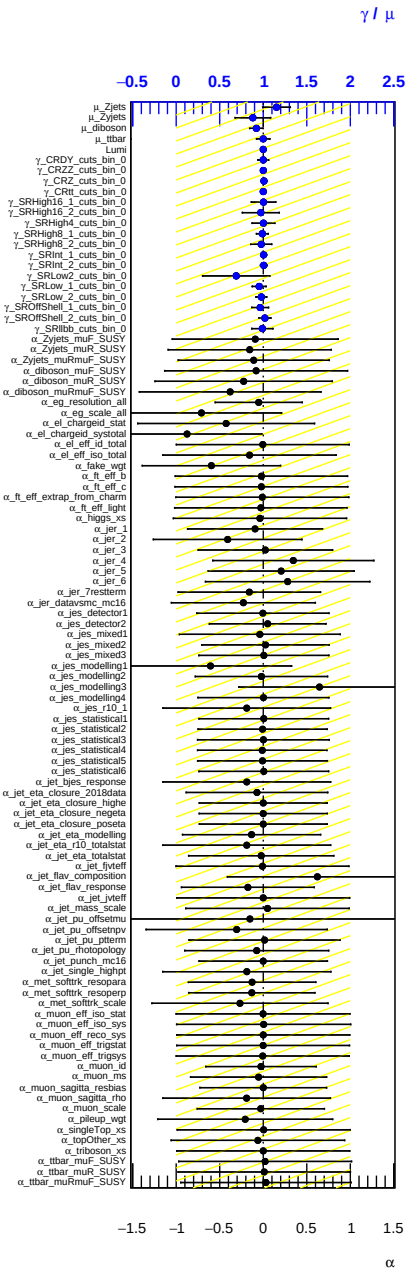


Figure 8.16: Post-fit pull plot of the normalization factors (μ), the size of statistical uncertainties (γ) and the theory, kinematic and efficiency parameters related to systematic uncertainties (α), with profiled uncertainties on the fitted parameters. The normalization factors and statistical uncertainties (shown in blue) are expressed relative to their pre-fit values (value of 1 means no deviation), whereas the systematic uncertainties (shown in black) are expressed as differences w.r.t. their nominal values/expectations (value of 0 means no deviation).

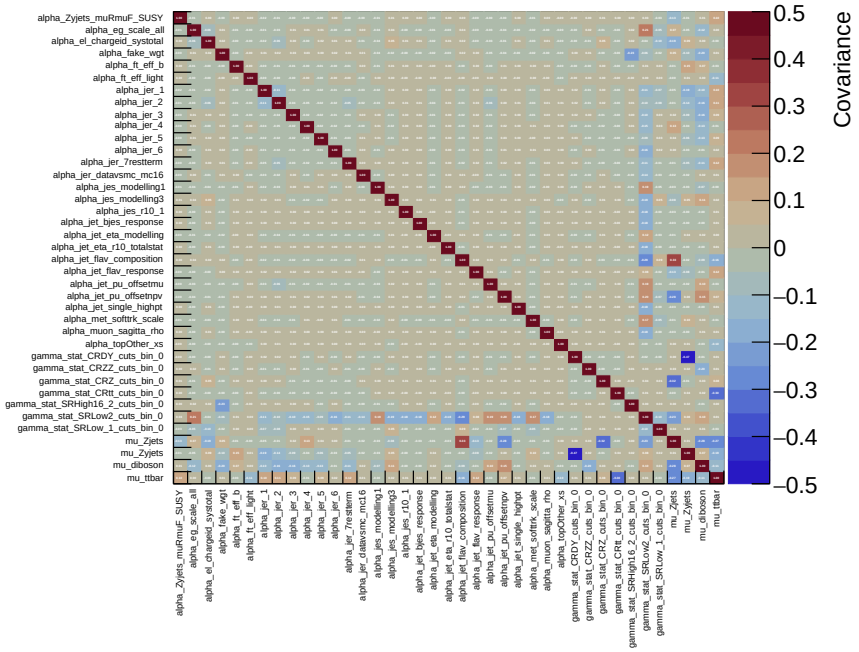


Figure 8.17: Post-fit correlations between the free parameters of the fit, including normalization factors, statistical and systematic uncertainties.

The validation regions, with the background normalization factors extrapolated from the fits to the CRs and SRs, generally show background predictions in good agreement with the data. VR-High, validating the background predictions near SR-High, has the largest gap between observation and prediction, with a slightly less than $2\,\sigma$ excess.

No significant excesses are observed in the signal regions. There are, however, a few moderately large downward fluctuations in the number of observed events compared to what was expected. Most noteworthy, no events are observed in the SR-High-8-a and -b regions, where 2.21 ± 0.23 and 1.57 ± 0.56 events were expected, respectively. These correspond to roughly $2\,\sigma$ deviations each.

More detailed summaries of the observed and expected number of events after the maximum likelihood fit are given for the CRs in Table 8.7, the VRs in Table 8.8 and 8.9, and the SRs in Table 8.10, 8.11 and 8.12.

Uncertainties on the numbers of events after the fit The dominant systematic uncertainties in each SR varies from region to region. For SR-High-4, which uses a single jet with mass consistent with the Z-mass, the jet mass scale is by far the most important uncertainty, and amounts to 30-40% of the expected

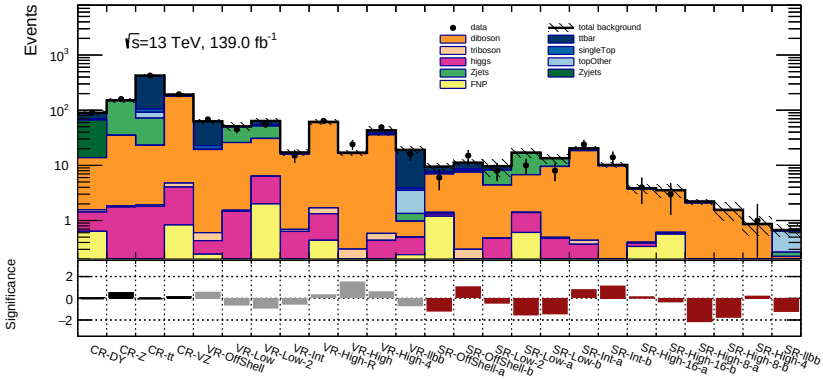


Figure 8.18: Post-fit pull plot of the observed number of events in all CRs, VRs and SRs compared to the expected number of background events in the regions.

Table 8.7: The number of observed (data) events compared with the expected number of background events in the regions used to control and constrain the background modeling in the signal regions. All statistical and systematic uncertainties are included.

	CR-ZZ	CR-tt	CR-Z	CR-DY
Observed events	194	424	159	90
Total exp. bkg. events	192.26 ± 12.54	424.23 ± 20.62	150.82 ± 12.40	90.00 ± 9.54
Diboson events	173.17 ± 14.28	21.13 ± 2.99	33.19 ± 3.65	12.18 ± 1.63
Top events	9.01 ± 1.10	352.52 ± 23.73	1.15 ± 0.28	23.75 ± 4.23
Z/γ^* events	$5.31^{+6.14}_{-5.31}$	48.67 ± 10.26	114.65 ± 13.42	52.52 ± 11.01
Other events	$4.76^{+1.45}_{-0.92}$	1.90 ± 0.93	1.83 ± 0.33	$1.56^{+1.07}_{-0.65}$

Table 8.8: The number of observed (data) events compared with the expected number of background events in the regions validating the background modeling near the high-Δm signal regions. All statistical and systematic uncertainties are included.

	VR-High-4	VR-High	VR-HighR
Observed events	49	24	64
Total exp. bkg. events	43.20 ± 8.23	17.03 ± 1.85	61.50 ± 5.67
Diboson events	35.06 ± 7.53	16.05 ± 1.83	57.72 ± 5.74
Top events	5.84 ± 1.02	0.62 ± 0.09	1.66 ± 0.24
Z/γ^* events	1.71 ± 1.15	0.06 ± 0.02	0.44 ± 0.17
Other events	0.59 ± 0.09	0.30 ± 0.08	$1.69^{+0.69}_{-0.46}$

Table 8.9: The number of observed (data) events compared with the expected number of background events in the regions validating the background modeling near the intermediate, low and off-shell-Δm signal regions. All statistical and systematic uncertainties are included.

	VR-Int	VR-Low	VR-Low-2	VR-OffShell
Observed events	15	45	55	68
Total exp. bkg. events	17.15 ± 2.04	51.22 ± 9.02	63.79 ± 7.33	63.00 ± 6.06
Diboson events	15.17 ± 1.98	24.35 ± 2.93	24.33 ± 3.04	18.71 ± 2.22
Top events	0.92 ± 0.40	1.96 ± 0.79	12.32 ± 2.06	43.69 ± 5.29
Z/γ^* events	$0.37^{+0.51}_{-0.37}$	23.48 ± 8.70	20.81 ± 5.71	0.00 ± 0.00
Other events	$0.70^{+0.16}_{-0.09}$	$1.53^{+0.32}_{-0.24}$	6.34 ± 1.73	$0.60^{+0.39}_{-0.25}$

Table 8.10: The number of observed (data) events compared with the expected number of background events in the high-Δm signal regions. All statistical and systematic uncertainties are included.

	SR-High-4	SR-High-8-a	SR-High-8-b	SR-High-16-a	SR-High-16-b
Observed events	1	0	0	4	3
Total exp. bkg. events	0.87 ± 0.34	2.21 ± 0.23	1.57 ± 0.56	3.85 ± 0.73	3.54 ± 0.93
Diboson events	0.82 ± 0.32	2.04 ± 0.23	1.48 ± 0.56	3.29 ± 0.61	2.88 ± 0.72
Top events	0.03 ± 0.01	0.12 ± 0.03	0.03 ± 0.02	0.14 ± 0.05	0.40 ± 0.02
Z/γ^* events	0.00 ± 0.02	0.02 ± 0.02	0.02 ± 0.02	0.02 ± 0.02	0.02 ± 0.01
Other events	0.02 ± 0.00	0.02 ± 0.00	0.01 ± 0.00	$0.40^{+0.39}_{-0.34}$	$0.02^{+0.69}_{-0.56}$

Table 8.11: The number of observed (data) events compared with the expected number of background events in the intermediate- and low-Δm signal regions. All statistical and systematic uncertainties are included.

	SR-Int-a	SR-Int-b	SR-Low-a	SR-Low-b	SR-Low-2
Observed events	24	14	10	8	8
Total exp. bkg. events	20.38 ± 1.69	10.20 ± 1.11	17.06 ± 3.45	13.48 ± 2.49	9.68 ± 3.35
Diboson events	17.96 ± 1.71	9.52 ± 1.04	5.33 ± 1.65	9.04 ± 1.20	3.91 ± 2.31
Top events	1.57 ± 0.42	0.25 ± 0.05	0.03 ± 0.01	0.12 ± 0.03	0.03 ± 0.02
Z/γ^* events	0.41 ± 0.32	$0.25^{+0.31}_{-0.25}$	9.96 ± 2.84	3.78 ± 2.02	3.79 ± 3.41
Other events	0.44 ± 0.05	0.19 ± 0.04	1.41 ± 0.35	0.06 ± 0.00	0.48 ± 0.25

Table 8.12: The number of observed (data) events compared with the expected number of background events in the off-shell signal regions. All statistical and systematic uncertainties are included.

	SR-OffShell-a	SR-OffShell-b
Observed events	6	15
Total exp. bkg. events	9.50 ± 1.44	11.21 ± 1.32
Diboson events	5.55 ± 0.81	7.20 ± 0.81
Top events	1.99 ± 0.57	3.125 ± 0.93
Z/γ^* events	$0.56^{+1.27}_{-0.56}$	0.60 ± 0.25
Other events	1.41 ± 0.41	$0.30^{+0.11}_{-0.07}$

background yield in the region.

In SR-High-8-a, i.e., with $\Delta R_{jj} < 0.8$ and $18 < E_{\mathrm{T}}^{\mathrm{miss,sig}} < 21$, the two dominant uncertainties are from the diboson normalization factor and the statistical uncertainty from the MC samples, corresponding to approximately 8 and 7% of the background expectation, respectively. The neighboring bin, SR-High-8-b, with $\Delta R_{jj} < 0.8$ and $E_{\mathrm{T}}^{\mathrm{miss,sig}} > 21$, on the other hand, is dominated by uncertainties in the determination of the e/γ and jet energy scales (JES) of approximately 20% each, and the statistical uncertainty from the MC samples of about 10-15%.

SR-High-16-a, with $\Delta R_{jj} < 1.6$ and $18 < E_{\mathrm{T}}^{\mathrm{miss,sig}} < 21$, the most important uncertainties come from the MC statistics of about 15% and from the FNP weights of about 10%. The diboson normalization factor also corresponds to an uncertainty roughly 8% of the background expectation in the region. SR-High-16-b, with $E_{\mathrm{T}}^{\mathrm{miss,sig}} > 21$, is dominated by the same sources of uncertainty, but here the MC statistics and the FNP weights amount to roughly 20% each, with the diboson normalization factor staying the same at 8%.

Moving on to the intermediate mass-splitting regions with more statistics, the dominant uncertainty in both SR-Int-a with $12 < E_{\mathrm{T}}^{\mathrm{miss,sig}} < 15$, and SR-Int-b with $15 < E_{\mathrm{T}}^{\mathrm{miss,sig}} < 18$, is the diboson normalization factor of about 8% in both regions.

Whereas the high and intermediate mass-splitting SRs are dominated by one source of background, namely diboson, the low mass-splitting SRs get sizeable contributions from both diboson and $Z+$jets. This fact also leads to higher numbers of significant uncertainty contributions of comparable sizes. In SR-Low-a, with $6 < E_{\mathrm{T}}^{\mathrm{miss,sig}} < 9$, the most dominant sources of uncertainty are jet flavor composition and electron charge ID of roughly 10-15% each. There are also sizeable contributions from the jet energy resolution (JER), MC statistics and the $Z+$jets normalization factor of slightly under 10% each. The uncertainties in the neighbouring bin, SR-Low-b, with $9 < E_{\mathrm{T}}^{\mathrm{miss,sig}} < 12$, are somewhat smaller, where six of the ten largest uncertainties come from JER, however none of them contribute more than 10% of the expected background in the region, and most of them are closer to 5%.

The final low mass-splitting SR, SR-Low-2, which is the only region with $m_{\mathrm{T2}} < 80\,\mathrm{GeV}$, the dominant uncertainty, by far, is the MC statistics, which amounts to about 30% of the total background expectation. There are, however, also smaller contributions from JER, especially from the $Z+$jets background, of around 10% in size.

Finally, the off-shell SRs have relatively modest uncertainty contributions, mostly from JER and MC statistics, and none of them larger than 10% of the expected number of background events in the region.

8.6.2.3 Test for exclusion

Since no significant excesses are observed on top of the background expectations, we formulate a new hypothesis test: is a given signal strength, μ_{sig}, consistent with the test statistic distribution, $f(q_{\mu_{\mathrm{sig}}}|\mu_{\mathrm{sig}})$, given the observed value, $q_{\mu_{\mathrm{sig}},\mathrm{obs}}$,

at some pre-determined CL? Similar to the discovery test, we quantify the compatibility between the test statistic distribution, given background plus the specific signal strength, and the observed value of the test statistic, in terms of a p-value, $p_{\mu_{\text{sig}}}$, of obtaining at least as extreme (higher) test statistic values than the one observed. I.e., the expected fraction of the distribution that is more background-like (non-signal-like) than the value observed. This can be expressed as:

$$p_{\mu_{\text{sig}}} = \int_{q_{\mu_{\text{sig}},\text{obs}}}^{\infty} f(q_{\mu_{\text{sig}}}|\mu_{\text{sig}})\, dq_{\mu_{\text{sig}}}. \tag{8.33}$$

The HEP standard is to set exclusion limits at 95% CL, which means that $p_{\mu_{\text{sig}}} < 0.05$, or equivalently, at least $1.64\,\sigma$ one-sided Gaussian deviations from the expectation, are excluded, given the observed data. We use the slightly more conservative requirement that $\text{CL}_s < 0.05$, where the CLs-value is defined as

$$\text{CL}_s \equiv \frac{p_{\mu_{\text{sig}}}}{1 - p_0}. \tag{8.34}$$

A range of signal strengths are scanned to find the specific μ_{sig} that gives $\text{CL}_s = 0.05$. This becomes the upper limit on the signal strength. We can then exclude all larger signal strengths, for which $\text{CL}_s < 0.05$.

The expected exclusion limits from the conventional SRs, with the full Run 2 dataset, are illustrated with dashed lines in Figure 8.19. The yellow band indicates the uncertainty in the exclusion limits w.r.t. the up- and down variations in the statistical and systematic uncertainties. For a massless LSP, this corresponds to the expectation of excluding $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ masses below about 750 GeV, whereas we expect to exclude LSP masses below about 10 GeV for a $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ mass of 100 GeV. In terms of mass-splittings, this means that the exclusion sensitivity ranges from $\Delta m \approx 750$ GeV down to the start of the off-shell regime at $\Delta m \approx 90$ GeV. In the end, I did not include the off-shell region in my exclusion fits. However, this region is covered by the final ATLAS publication, presented in the next section (8.7).

The results presented in this section have been produced by me for this thesis. They are derived using the common analysis framework and ntuples produced by the working group I have been a part of, but they do not fully represent the final product of the ATLAS publication. The differences w.r.t. the ATLAS publication are discussed in the following section.

8.7 Published analysis

At the time of delivery of the thesis, the complete ATLAS Run 2 analysis has been submitted to the European Physical Journal C, and has been published as a pre-print in Reference [193].

As mentioned in the previous section, there are some differences in the analysis of the publication w.r.t. what I have presented so far in this chapter. Probably the most notable difference is the addition of exclusion limits for both

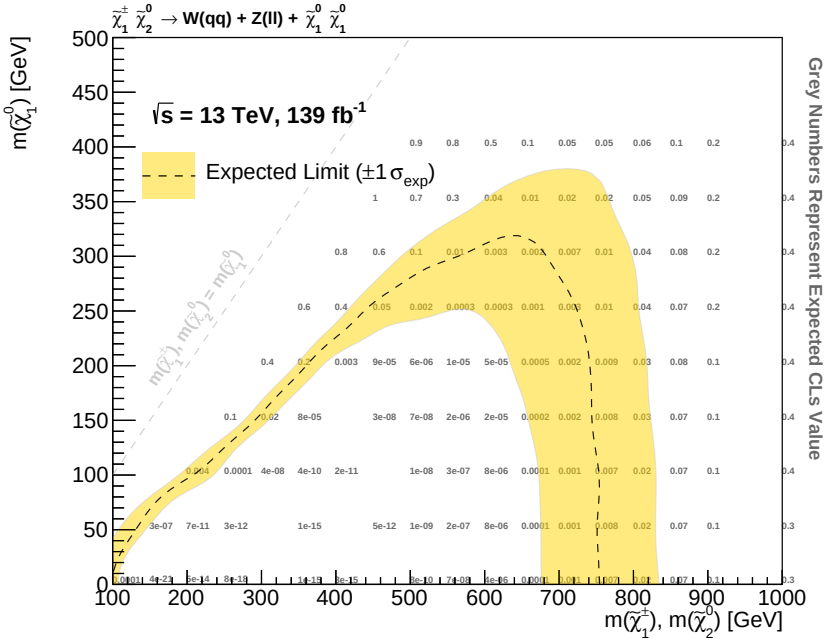


Figure 8.19: Expected 95% CL exclusion contour with the full Run 2 dataset. Signal mass points in the so-called off-shell region, where $\Delta m = m_{\tilde{\chi}_1^\pm,\tilde{\chi}_2^0} - m_{\tilde{\chi}_1^0} < m_Z \approx 91\,\text{GeV}$, have not been considered in the calculation of the exclusion contour.

model-independent and model-dependent signals. Another important difference is the addition of more systematic uncertainties, such as uncertainties to account for differences between the nominal and alternative MC samples for the most important background processes, and experimental uncertainties in the modeling of jet masses. The latter is an especially important addition for the boosted high-Δm SR, where we assume that the two jets from the decay of the W boson have effectively merged into a single reconstructed jet, with $\Delta R < 0.4$, and this jet has a mass consistent with the mass of the W boson. I have previously referred to this region as SR-High-4, whereas the publication has renamed this region to SR-1J-High-EWK. Apart from the different name, the region definitions are otherwise identical.

The publication also modified the treatment of the renormalization and factorization scale uncertainties and the PDF uncertainties somewhat w.r.t. what I have described in Section 8.5.1.1 and 8.5.1.2. Instead of taking the envelope of the various combinations of μ_R and μ_F variations as a single nuisance parameter, three nuisance parameters where added, one for μ_R, one for μ_F and one for the correlation between the two, $\mu_\text{R}\mu_\text{F}$. For the PDF uncertainties, instead of using the sample mean and variance of the ensemble of PDF variations,

the publication makes use of an alternative prescription which uses quantiles of reweighted samples.

Since the publication also presents search regions for strong production of sparticles with the same final-state, the search regions that target electroweak production of sparticles get an "EWK"-suffix appended to the names I have referred to in this chapter. The region definitions are otherwise identical to what I have presented in Table 8.1, 8.2, 8.3 and 8.4.

The publication also includes an additional electroweak SR, called SR-$\ell\ell bb$-EWK, which targets a gauge-mediated SUSY breaking (GMSB) model with pair-production of higgsino NLSPs, where the higgsinos decay via Higgs or Z bosons into gravitino LSPs.

The expected and observed event yields in the electroweak SRs are presented in Table 8.13 and 8.14.

Table 8.13: Breakdown of expected and observed yields in the electroweak search High and $\ell\ell bb$ signal regions after a simultaneous fit to the SRs and CRs. All statistical and systematic uncertainties are included. The table and caption are taken from Reference [193].

	SR-High_16a-EWK	SR-High_8a-EWK	SR-1J-High-EWK	SR-$\ell\ell bb$-EWK
Observed events	4	0	1	0
Total exp. bkg. events	3.9 ± 0.7	2.00 ± 0.23	0.85 ± 0.34	0.58 ± 0.20
Diboson events	3.2 ± 0.6	1.86 ± 0.22	0.80 ± 0.31	0.13 ± 0.03
Top events	$0.00^{+0.01}_{-0.00}$	0.0 ± 0.0	$0.03^{+0.04}_{-0.03}$	$0.05^{+0.08}_{-0.05}$
Z + jets events	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Other events	0.7 ± 0.4	0.15 ± 0.07	$0.02^{+0.04}_{-0.02}$	0.39 ± 0.16

	SR-High_16b-EWK	SR-High_8b-EWK
Observed events	3	0
Total exp. bkg. events	3.4 ± 0.9	2.00 ± 0.33
Diboson events	2.5 ± 0.6	1.94 ± 0.33
Top events	0.0 ± 0.0	0.0 ± 0.0
Z + jets events	0.0 ± 0.0	0.0 ± 0.0
Other events	0.9 ± 0.7	0.06 ± 0.04

Distributions of the expected background and signal yields along with the observed number of events in the most discriminating signal-vs-background variables, for five of the SRs, are shown in Figure 8.20. The observed data seems to fit well with the background expectations, with some notable deviations, in particular for the SR-High_8-EWK and SR-$\ell\ell bb$-EWK, where no events are observed.

The yields in the electroweak CRs, VRs and SRs after a simultaneous background-only fit to the CRs and SRs are summarized in their single-bin form in Figure 8.21. A measure of the significance of the deviation between observation and background expectation in each region is also shown in the lower panel of the plot. This figure corresponds to the one shown in Figure 8.18, but where Figure 8.21 has the additional systematic uncertainties outlined above included

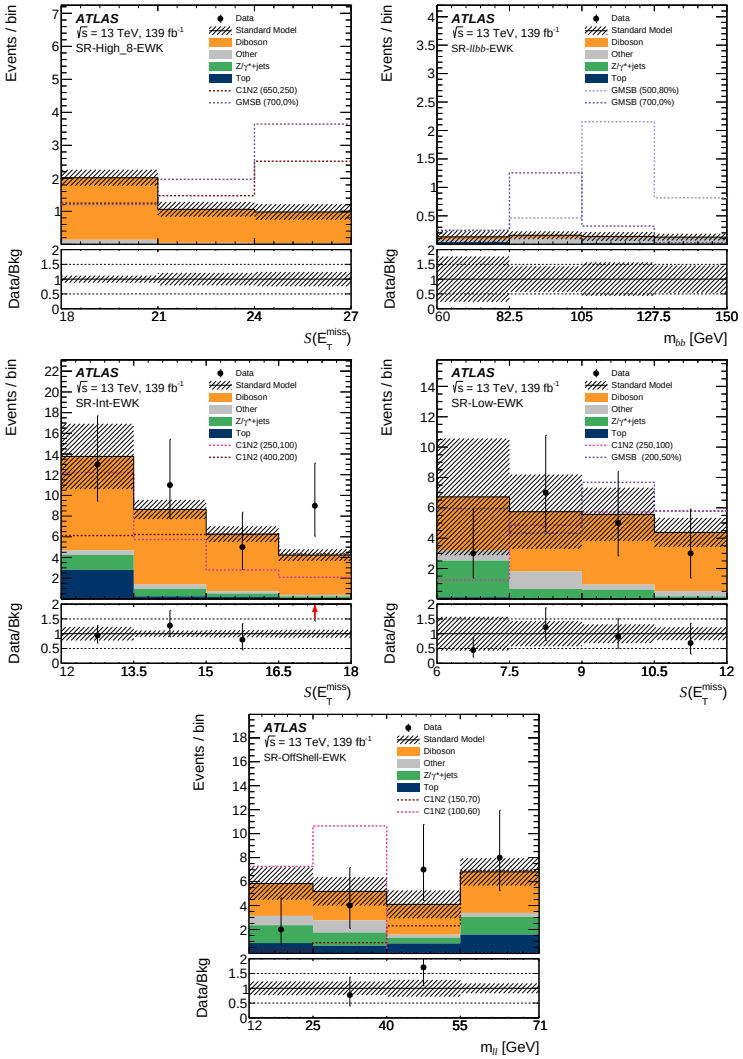


Figure 8.20: Observed and expected distributions in five EWK search regions after a simultaneous fit to the SRs and CRs. In the top row, left-to-right, are $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-High_8-EWK and m_{bb} in SR-$\ell\ell bb$-EWK. In the middle row, left-to-right, are $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-Int-EWK and $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ in SR-Low-EWK. In the bottom row is $m_{\ell\ell}$ in SR-OffShell-EWK. Overlaid are example C1N2 and GMSB signal models, where the numbers in the brackets indicate the masses, in GeV, of the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ or the mass of the $\tilde{\chi}_1^{\pm}$ and branching ratio to the Higgs boson respectively. All statistical and systematic uncertainties are included in the hatched bands. The last bin includes the overflow. The figures and caption are taken from Reference [193].

Table 8.14: Breakdown of expected and observed yields in the electroweak search Int, Low, and OffShell signal regions after a simultaneous fit to the SRs and CRs. All statistical and systematic uncertainties are included. The table and caption are taken from Reference [193].

	SR-Int_a-EWK	SR-Low_a-EWK	SR-Low-2-EWK	SR-OffShell_a-EWK
Observed events	24	10	8	6
Total exp. bkg. events	22.8 ± 3.5	12.8 ± 3.4	9 ± 4	9.2 ± 1.7
Diboson events	16.5 ± 1.7	7.3 ± 1.3	4.0 ± 2.1	4.9 ± 1.3
Top events	4 ± 4	$0.06^{+0.14}_{-0.06}$	$1.0^{+1.2}_{-1.0}$	1.4 ± 0.7
Z + jets events	2.1 ± 0.7	3.7 ± 3.3	4 ± 4	1.2 ± 1.2
Other events	0.44 ± 0.13	1.7 ± 0.4	0.58 ± 0.3	1.6 ± 0.4

	SR-Int_b-EWK	SR-Low_b-EWK	SR-OffShell_b-EWK
Observed events	14	8	15
Total exp. bkg. events	10.1 ± 1.0	10.5 ± 2.5	12.5 ± 1.9
Diboson events	9.2 ± 1.0	8.6 ± 1.2	6.1 ± 1.5
Top events	0.22 ± 0.13	0.0 ± 0.0	2.8 ± 1.4
Z + jets events	0.51 ± 0.31	$1.3^{+2.2}_{-1.3}$	3.1 ± 1.4
Other events	0.19 ± 0.08	0.70 ± 0.11	0.54 ± 0.24

as nuisance parameters in the fit. The additional uncertainties modifies the post-fit results somewhat, but not significantly and do not alter the conclusions.

The results after performing signal-plus-background fits to data in the the CRs and a (model-independent) single-bin discovery region (DR), separately for each DR in turn, are shown in Table 8.15. The DRs are defined in Table 8.5. There are no significant excesses observed above the expected SM background. In fact, only one of the DRs have an excess of observed events, which is DR-Int-EWK, with 35 ± 4 background events expected and 38 events observed in the data. This corresponds to a $0.8\,\sigma$ excess, which is not considered significant. The remaining four DRs all have less events observed than expected, with DR-High-EWK having the largest deficit, where 3.9 ± 0.5 background events were expected and none events were observed in the data. This corresponds to a -2.8 σ deficit. Table 8.15, however, caps p-values at 0.5, i.e., significance (Z) values of $0\,\sigma$.

The expected and observed exclusion contours obtained from simultaneous model-dependent signal-plus-background fits to the CRs and the SRs are shown in Figure 8.22. The contours for the $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ model is shown on the left and for the GMSB model on the right. The expected exclusion contour shown previously in Figure 8.19 is consistent with the left plot in Figure 8.22, apart from the additional expected exclusion in the off-shell region (the dent in the contour at the lower left) of the latter, which was not considered in the former. Additionally, the observed exclusion contour is added as a red line in Figure 8.22. Due to deficits in the off-shell and, in particular, in the high-Δm region (the rightmost part of the contour), the observed exclusion limits are stronger than expected. We expected to exclude $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ masses up to 750 GeV, whereas the observed

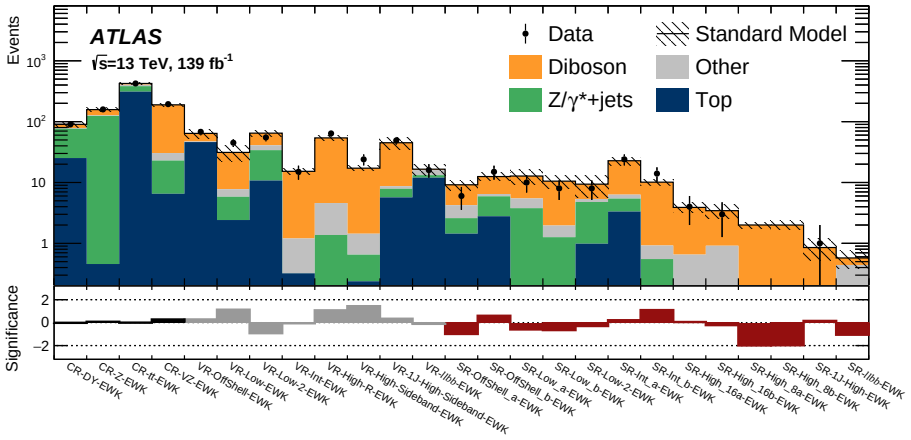


Figure 8.21: The observed and expected yields in the control regions (left), validation regions (middle), and signal regions (right) of the EWK search after a simultaneous fit to the SRs and CRs. The hatched band includes the statistical and systematic uncertainties of the background prediction in each region. The significance of the difference between the observed data and the expected yield in each region is shown in the lower panel using the profile likelihood method of Reference [194]; the colours black, grey, and red separate the CRs, VRs, and SRs, respectively. For the cases where the expected yield is larger than the data, a negative significance is shown. The figure and caption are taken from Reference [193].

limit extended all the way up to 820 GeV. In terms of $\tilde{\chi}_1^0$ masses, we expected to exclude masses up to 300 GeV, but ended up excluding $\tilde{\chi}_1^0$ masses up to 380 GeV. These limits extend significantly beyond the previous limits shown in gray, obtained with 36.1 fb^{-1}, as presented in Chapter 7.

The addition of an SR targeting small enough mass-splittings between $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$ that the Z boson needs to be produced off-shell, extends the exclusion contour into a region of phase space that the two lepton searches of ATLAS have not covered before. In fact, the exclusion contour goes as low as to exclude two mass-points with $\Delta m = 10$ GeV, namely the $m_{\tilde{\chi}_1^\pm/\tilde{\chi}_2^0,\tilde{\chi}_1^0} = (100, 90)$ and (110, 100) GeV mass-points.

The exclusion contours for the GMSB model, shown on the right in Figure 8.22, shows the $\tilde{\chi}_1^0$ (NLSP) mass on the x-axis and the branching fraction (B or BF) of $\tilde{\chi}_1^0$ to a Higgs (h) boson and a gravitino on the y-axis (the alternative is $\tilde{\chi}_1^0$ to a Z boson and a gravitino). $\tilde{\chi}_1^0$ pair-production with 100% BF to Z, i.e., 0% BF to h, is excluded up to $\tilde{\chi}_1^0$ masses of 890 GeV, whereas we reach exclusion all the way up to about 94% for $\tilde{\chi}_1^0$ to h for an $\tilde{\chi}_1^0$ mass of 450 GeV.

The CMS collaboration has also published exclusion limits with their full Run 2 dataset. The expected and observed exclusion limits with 137 fb^{-1} of integrated luminosity are shown in Figure 8.23. They expected to exclude $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$-masses up

Table 8.15: Model-independent upper limits on the observed visible cross-section in the five electroweak search discovery regions, derived using pseudo-experiments. Left to right: background-only model post-fit total expected background, with the combined statistical and systematic uncertainties; observed data; 95% CL upper limits on the visible cross-section ($\langle A\epsilon\sigma\rangle_{\text{obs}}^{95}$) and on the number of signal events (S_{obs}^{95}). The sixth column (S_{exp}^{95}) shows the expected 95% CL upper limit on the number of signal events, given the expected number (and $\pm 1\sigma$ excursions of the expectation) of background events. The last two columns indicate the confidence level of the background-only hypothesis (CL_b) and discovery p-value with the corresponding Gaussian significance ($Z(s=0)$). CL_b provides a measure of compatibility of the observed data with the signal strength hypothesis at the 95% CL limit relative to fluctuations of the background, and $p(s=0)$ measures compatibility of the observed data with the background-only hypothesis relative to fluctuations of the background. The p-value is capped at 0.5. The table and caption are taken from Reference [193].

Signal Region	Total Bkg.	Data	$\langle A\epsilon\sigma\rangle_{\text{obs}}^{95}$ [fb]	S_{obs}^{95}	S_{exp}^{95}	CL_b	$p(s=0)$ (Z)
DR-OffShell-EWK	22.1 ± 2.7	21	0.10	14.3	$12.3^{+4.7}_{-3.1}$	0.68	0.50 (0.0)
DR-Low-EWK	22 ± 4	18	0.08	10.8	$15.3^{+5.7}_{-4.0}$	0.09	0.50 (0.0)
DR-Int-EWK	35 ± 4	38	0.15	20.9	$17.5^{+5.9}_{-3.9}$	0.73	0.23 (0.8)
DR-High-EWK	3.9 ± 0.5	0	0.02	3.0	$5.6^{+2.2}_{-1.5}$	0.00	0.50 (0.0)
DR-$\ell\ell bb$-EWK	0.51 ± 0.20	0	0.02	3.0	$3.0^{+1.3}_{-0.0}$	0.19	0.50 (0.0)

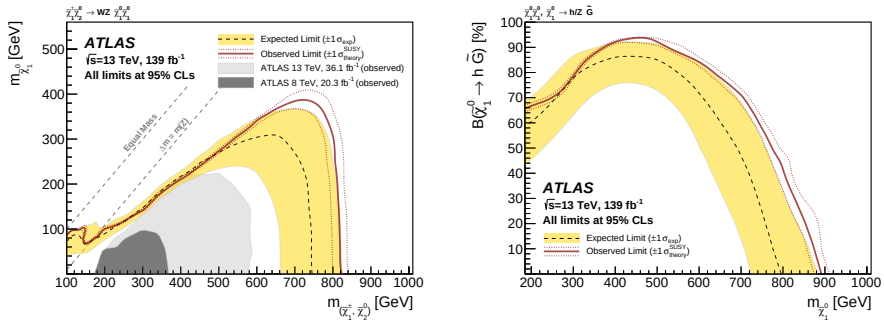


Figure 8.22: Expected and observed exclusion contours from the EWK analysis for the C1N2 model (left) and GMSB model (right). The dashed line indicates the expected limits at 95% CL and the surrounding band shows the 1σ variation of the expected limit as a consequence of the uncertainties in the background prediction and experimental uncertainties of the signal ($\pm 1\sigma_{\text{exp}}$). The red dotted lines surrounding the observed limit contours indicate the variation resulting from changing the signal cross-section within its uncertainty ($\pm 1\sigma_{\text{theory}}^{\text{SUSY}}$). The grey-shaded areas indicate observed limits on these models from the two-lepton channels in Reference [195] and Reference [196]. The figures and caption are taken from Reference [193].

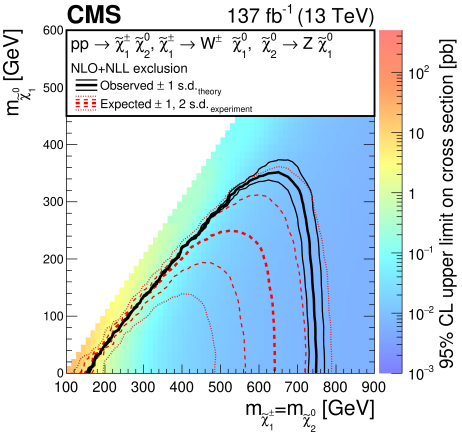


Figure 8.23: Expected and observed exclusion limits from the CMS collaboration with 137 fb^{-1} of integrated luminosity, in the 2ℓ+jets channel. The figure is taken from Reference [197].

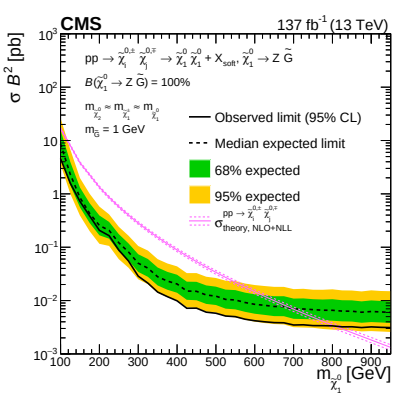
to 640 GeV and $\tilde{\chi}_1^0$-masses up to 250 GeV. The observed limits ended up being stronger, excluding $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$-masses up to 750 GeV and $\tilde{\chi}_1^0$-masses up to 350 GeV.

Figure 8.24 shows exclusion limits obtained by CMS for the GMSB model with 100% BF for $\tilde{\chi}_1^0$ to a Z boson and a gravitino in (a) and with 50% BF for $\tilde{\chi}_1^0$ to a Z boson or a Higgs boson and a gravitino in (b). From Figure 8.24a we can see that the observed exclusion limit (black line) surpasses (excludes) the signal cross-section times 100%BF to Z (magenta line) for all $\tilde{\chi}_1^0$ masses below 800 GeV, whereas Figure 8.24a shows that all $\tilde{\chi}_1^0$ masses below 520 GeV are excluded for cross-section times 50% BF for $\tilde{\chi}_1^0$ to a Z or a H.

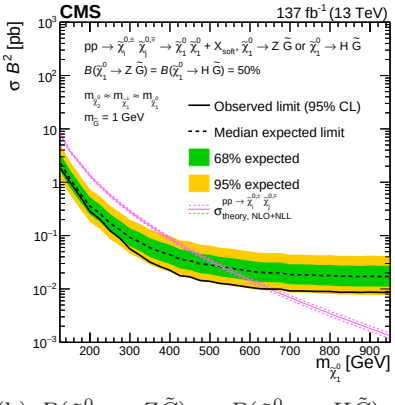
8.8 Summary

My sensitivity studies show that the expected exclusion limits in the plane of $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$- and $\tilde{\chi}_1^0$-masses, extracted from events with two leptons, jets and $E_{\mathrm{T}}^{\mathrm{miss}}$, are significantly improved using the complete Run 2 dataset with 139 fb^{-1}, as demonstrated in Figure 8.19, compared to the previous iteration of the Run 2 analysis using 36.1 fb^{-1}, shown in Figure 7.8.

In the high $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$-mass region, i.e., for large mass-splittings, the expected exclusion contour moves from 530 GeV with 36.1 fb^{-1} to 750 GeV with 139 fb^{-1}. The highest expected exclusion reach for the $\tilde{\chi}_1^0$-mass moves from 200 GeV at $m_{\tilde{\chi}_1^\pm,\tilde{\chi}_2^0} \approx 500$ GeV with 36.1 fb^{-1}, to about 300 GeV at $m_{\tilde{\chi}_1^\pm,\tilde{\chi}_2^0} \approx 650$ GeV with 139 fb^{-1}. The latter corresponds to the reach of the intermediate mass-splitting region. In the low mass-splitting region, we reach exclusions close to mass-splittings similar in size to the mass of the Z-boson, i.e., $\Delta m \approx 90$ GeV. My exclusion fits do not include the off-shell mass-points, i.e., with $\Delta m < 90$ GeV,



(a) $B(\tilde{\chi}_1^0 \to Z\tilde{G}) = 100\%$



(b) $B(\tilde{\chi}_1^0 \to Z\tilde{G}) = B(\tilde{\chi}_1^0 \to H\tilde{G}) = 50\%$

Figure 8.24: CMS exclusion limits at 95% CL for electroweak pair-production cross-section (σ) for $\tilde{\chi}_1^0$ $\tilde{\chi}_1^0$ times (a) 100% branching fraction (B) for $\tilde{\chi}_1^0$ to a Z boson and a gravitino, or (b) 50% branching fraction to a Z boson or a Higgs boson and a gravitino, as a function of $\tilde{\chi}_1^0$ mass. The figures are taken from Reference [197].

and therefore the expected exclusion contour does not represent our actual exclusion potential in the off-shell region.

The complete ATLAS publication shows that also the observed exclusion limits are significantly improved w.r.t. the analysis with 36.1 fb^{-1} of data. Due to no observed events in the high-Δm SRs, the exclusion limits become even stronger than expected, and reaches all the way up to $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ masses of 820 GeV, compared to the expectation of 750 GeV. The observed exclusion limit for $\tilde{\chi}_1^0$ masses reaches up to 380 GeV, compared to the expectation of 300 GeV. In addition, the ATLAS publication also incorporates exclusion fits for off-shell mass points, which extends the exclusion contour into new regions of phase space compared to previous ATLAS two lepton searches. Mass-splittings between the NLSPs and LSPs as low as 10 GeV are excluded by the off-shell SR.

In the following, and final, analysis chapter, I present the multivariate analysis studies I have performed using the XGBoost algorithm, to investigate the possibility of gaining expected sensitivity to the same signal model targeted in this chapter, with the integrated luminosity of the complete Run 2 dataset.

Chapter 9

Multivariate signal region optimization

The conventional approach to signal region optimization is to manually look for or create variables that best separate the behavior of signal and background events. Then, sequential cuts are applied to these, one after the other, where the cut on each variable is typically chosen based on the behavior in that variable alone. Therefore, we may end up cutting away events that look background-like in one variable, but may look signal-like in another. Such univariate cuts are often referred to as rectangular cuts, and may make a sub-optimal separation/decision-boundary between signal and background in the multidimensional space of event variables.

A more sophisticated approach to this problem is to consider multiple variables simultaneously, sometimes called *multivariate analysis* (MVA). From a cut-based analysis point of view, one may think of it as correlating the cuts on multiple variables, so that a cut on one variable effectively becomes a function of the other variables we choose to include. We can, e.g., be more lenient on the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, i.e., allow for lower values of $E_{\mathrm{T}}^{\mathrm{miss}}$, given that other variables of the event show signal-like behavior.

Many methods exist for performing MVA. In high-energy physics (HEP), the application of *boosted decision trees* (BDTs) to optimize and define signal regions have been around for some time. BDT analyses have in several cases shown enhanced sensitivity, compared to traditional cut-based analyses, to challenging signal scenarios. Notable examples are the first evidence of single top quark production by the D0 experiment [198] in 2008, and the subsequent observations of the same process by the D0 [199] and CDF [200] experiments the year after, all making use of BDT discriminators.

More recently, *deep neural networks* have also been demonstrated in searches for exotic particles, where, e.g., Reference [201] found that deep neural nets were able to learn high-level representations from low-level kinematic input variables, which in some cases performed better than shallow methods (BDT and shallow neural net) did by using hand-crafted, high-level input features calculated from physics knowledge.

In what follows, I will start by introducing the XGBoost algorithm (Section 9.1), which is used to fit the models presented in this chapter. The concept of binary classification is then presented (Section 9.2) in the context of separating signal from background, before explaining the typical machine learning methodology of training, validating and testing the models on separate samples of the dataset (Section 9.3). Then, I move on to the actual data analysis, by first discussing the preselection of events that go into the training of the models (Section

9.4), before presenting the training and test performance of the models that target the low, intermediate and high mass-splitting signal scenarios (Sections 9.5, 9.6 and 9.7, respectively), on a few selected benchmark signal mass points. Finally, the performance of the XGBoost models are summarized in terms of their expected exclusion sensitivity in the full grid of signal mass hypotheses (Section 9.8), and the chapter is concluded by a summary of the feasibility study with future prospects (Section 9.9).

9.1 The XGBoost algorithm

The XGBoost [202] algorithm, for training classification and regression trees (CARTs), quickly became one of the most popular and powerful machine learning (ML) methods on the Kaggle website [203] for ML competitions, after it was used by several of the highest scoring contestants in the Higgs Boson Machine Learning Challenge (HiggsML) [204], hosted by the ATLAS collaboration on the same website. The developers of XGBoost were awarded the *HEP meets ML Award* [205] for creating a model that was considered to be the most useful for the ATLAS experiment, in terms of accuracy, simplicity, performance and robustness. In recent years, it has become one of the most widely used and highest performing ML methods on *structured* (tabular) data, which are data that fit a predefined model and are therefore relatively straightforward to read in and analyze. In addition, the data points, such as particle collision events, can often (to a good approximation) be taken to be *independent and identically distributed* (i.i.d.).

For so-called *unstructured* data, such as images, audio or text, where the data points (pixels, audio samples or words) are correlated in space and/or time, deep neural networks such as convolutional neural networks (CNNs) for images and recurrent neural networks (RNNs) for audio or text sequences, have special network architectures and perception mechanisms that capture the nature of these data better than decision trees. Such methods have still found very effective applications in HEP. For example, by treating the energy depositions in the calorimeter cells as if they were pixels of an image, one can use CNNs to tell apart shower shapes of jets initiated by quarks and gluons [206]. Another example is using an RNN to read in the sequence of p_{T}-ordered jets in each event, to accommodate the fact that the number of jets, and thereby the length of the sequence, varies from event to event. The RNN maps these sequences to a fixed number of output values, which further can be fed to a regular feedforwad neural net for classification, that takes a fixed number of inputs. This method is in use to tag jets which originate from a hadronic decay of a tau lepton [207].

One of the strengths of deep learning is that these models can learn high-level, non-linear representations from low-level inputs. For particle experiments, this means that we in principle should not need to calculate high-level, multi-particle quantities, such as invariant masses, angular separations or $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ of an event, because the deep neural net should be able to learn similar representations directly from the low-level detector measurements of single particles, given

enough training examples and appropriate architecture and parameter values. Just because deep neural nets have the ability to learn high-level features, does not mean that we do not benefit from adding high-level features as inputs. That may speed up the learning significantly, as the network does not need to learn those representations by itself. They may also be very useful in preselecting relevant events for the training.

Because BDTs cannot learn high-level features from low-level features, it is more important to supply them with informative and discriminatory variables than for neural nets. Since we already have a solid theoretical understanding and description of the mechanics of the particle interactions, with mathematical models to calculate high-level variables which can separate signal from background well, the added complexity and challenges of tuning and designing a well-functioning deep neural net may in many cases not be worth the struggle, nor significantly increase the sensitivity compared to the simpler BDTs. The BDTs are also more transparent and easier to interpret than the black-box model of neural nets.

As much as I am fascinated by the abilities of deep neural networks, the XGBoost model appeared to be a more promising starting point for MVA exploration. In the following I will try to give a short introduction to decision trees in general and the XGBoost algorithm in particular.

9.1.1 Decision trees and adaptive boosting

A decision tree can be represented graphically like a flowchart with a tree structure, which holds a set of tests for the input to be evaluated on. An example is shown in Figure 9.1. Every internal node in the tree corresponds to a binary (yes/no) test, and the input will continue along one of two branches out from the test node, determined by whether the input passes the test or not. An arbitrary number of tests, with accompanying branch splits, can be added to the tree. The outcome of the final test in the tree will end up in a terminal node, called a leaf, that assigns the input to a real value for a regression task, or a certain class probability or label for a classification task.

A single decision tree tends to overfit on the training set, and is therefore not able to generalize well to new, unseen data. In machine learning jargon this is referred to as producing a large *variance* (overfitting). By restricting the decision tree to be shallow, i.e., restricting the depth (number of decisions) that can be included in a single tree, we can reduce the variance. However, the decision tree would then typically be less precise in its predictions, which in machine learning jargon is referred to as giving large *bias* (underfitting). This is known as the bias-variance trade-off.

The concept of boosting, most commonly applied in terms of the *adaptive boosting* algorithm, known as AdaBoost (the default model for ROOT's TMVA [208] BDTs), makes use of an ensemble of decision trees, where events (data points) that were previously misclassified, or were assigned predictions different from their true values, will be given larger weight in the training/growing of subsequent trees, in proportion to the size of their errors. This will in turn lead

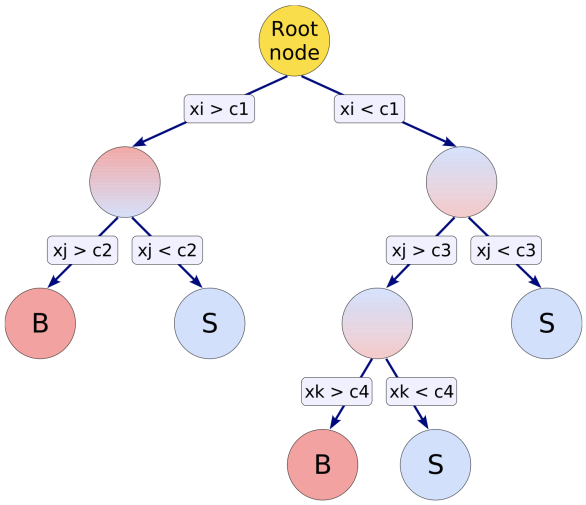


Figure 9.1: A visual representation of a decision tree for binary classification, separating signal, S, from background, B. The variables of an event enter the tree in the root node. The first test in this tree is whether the value of the variable `xi` in the given event is less or greater than the cut/threshold value `c1`. If `xi` is greater than `c1`, the event will follow the right path, otherwise it will go to the left. Next, the event's value for the variable `xj` will be evaluated against the value `c2` if `xi` $<$ `c1` or against `c3` if `xi` $>$ `c1`. In practice, `xj` may represent a different variable on each side of the tree, and a variable may enter at mulitple levels of the tree, e.g., evaluate whether the variable is greater than a certain value, but smaller than another. The event continues down the branches of the tree until it reaches a terminal node, called a leaf, which labels the event as either signal- or background-like. In practice, each leaf node is assigned a value between 0 and 1, where 0 represents background and 1 signal. The value (weight) of each leaf is tuned during the training process. All events which end up in a given leaf node are assigned the value (weight) of that node. When multiple trees are used in the training and evaluation, each event is passed down all the trees and the value it gets from each tree is added together to a total score (total weight) for that event. One could then choose to label all events with a total score of less than 0.5 to be classified as background, and all event with a score greater than 0.5 are classified as signal. In the work of this thesis, I consider this threshold to be a tunable cut value that defines the signal region, i.e., events which end up with a total score greater than the cut value is considered highly signal-like and falls in the signal region, whereas the rest of the events are considered less signal-like and are rejected from the statistical analysis. The figure is taken from Reference [208].

to a greater loss/cost of getting these events wrong again, and effectively makes the optimizer focus more on getting these hard cases right. By also restricting each decision tree to be a weak learner, typically only 2 or 3 levels deep, each tree will only be a little better than random guessing. But combined with other trees (typically of the order of a hundred), which are forced to be different by the boosting mechanism, the final combined "forest" of trees typically end up being a much stronger classifier than any single tree – both with less variance than a single strong learner and less bias than a single weak learner.

9.1.2 Gradient boosting

A more recent concept is that of *gradient boosting*, which makes use of a loss function that can be minimized by gradient descent, similar to how artificial neural networks are trained. The loss function is parametrized in terms of the so-called weights (trainable parameters) of the model, and the gradient of the loss function is then calculated w.r.t. small variations in these weights.

For neural networks, there is one weight, w_{ij}, associated with every input, x_i, to a specific node in the network, here denoted by an index j. The corresponding output of this node, o_j, can be calculated by

$$o_j = \varphi\left(\sum_i w_{ij} x_i\right) = \varphi\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}\right), \qquad \text{(neural network node)} \qquad (9.1)$$

where φ is some non-linear activation function. During training, these weights are varied in order to up- or down-weight the contribution from each input in a way that minimizes the loss. These weights are numbers that can be varied continuously throughout the training process.

For decision trees, however, the corresponding weights are the score values associated with the different end nodes, or leaves, of each tree. That is, the value given to every event ending up in a certain leaf after having traversed the tree. A decision tree is grown by scanning all the variables in the input data, searching for the most powerful cut to distinguish signal from background. Then, a split is made on this variable that gives two child nodes or leaves, one for signal-like events and one for background-like events. Then, given the decision on the most powerful variable, the algorithm continues to scan the variables again for the possibility of making additional splits on the newly grown leaves. The process continues either until the maximum allowed depth of the tree has been reached, or there is no more progress to be made within the constraints set for the learning process. This means that each decision tree can learn a different structure, and possibly also have a different number of leaves, thereby potentially giving a different number of weights (trainable parameters) in each tree.

We can think of a decision tree as a function, f, that maps the variable (feature) vector of an event, $\boldsymbol{x}$, to a certain leaf in the tree, j. The event's path through the decision tree is represented by the function $q(\boldsymbol{x})$, and the associated output score/value of that leaf is represented by the weight $w_{q(\boldsymbol{x})} \to w_j$. The

decision tree's mapping from input to a leaf output can then be expressed as

$$f(\boldsymbol{x}) = \boldsymbol{w}_{q(\boldsymbol{x})}, \quad q : \mathbb{R}^m \to T, \quad \boldsymbol{w} \in \mathbb{R}^T, \tag{9.2}$$

where m is the number of input variables (dimensions) of the event, and T is the number of leaves in the given tree [202].

The predicted output value of event i, $\hat{y}_i$, is then calculated by summing up the weights given to this event in all the trees in the ensemble:

$$\hat{y}_i = \phi(\boldsymbol{x}_i) = \sum_{k=1}^{K} f_k(\boldsymbol{x}_i) = \sum_{k=1}^{K} \boldsymbol{w}_{k,q(\boldsymbol{x}_i)}, \tag{9.3}$$

where K is the number of additive functions/trees in the ensemble, and each new weight attempts to get the predicted value $\hat{y}_i$ closer to the true target value, y_i.

Whereas neural networks minimize the loss function w.r.t. a fixed number of weights, by performing gradient descent on (ideally) a convex loss function, gradient boosted decision trees perform so-called *functional gradient descent* on a convex set of functions, that is, the loss function is a convex functional of the tree functions.

In order to optimize a classifier by gradient descent, we need to define a differentiable loss function which is suitable for the task at hand. For binary classification, the so-called *logarithmic loss* or *binary cross-entropy* is typically used:

$$L(y_i, \hat{y}_i) = -y_i \ln \hat{y}_i - (1 - y_i) \ln(1 - \hat{y}_i), \tag{9.4}$$

where y_i denotes the true binary class label of an event with index i, 0 for background and 1 for signal, while $\hat{y}$ denotes the value predicted by the model, which can take on values in the interval [0, 1].

Since the logarithm of a value between 0 and 1 is negative, the value of the loss function will always be positive, or 0 if $\hat{y} = 1$, since $\ln 1 = 0$. Also, since the true class label, y, is always either 0 or 1, only one of the two terms in the loss function will survive in any given case.

E.g., if the given event is a signal event, the true value is $y_i = 1$. Let us say that the model predicts a value $\hat{y}_i = 0.5$. The corresponding loss is then

$$L(1, 0.5) = -1 \cdot \ln 0.5 - (1 - 1) \cdot \ln(1 - 0.5) \tag{9.5}$$

$$\approx -1 \cdot (-0.69) = 0.69. \tag{9.6}$$

If instead the predicted value is $\hat{y}_i = 0.9$, the corresponding loss is

$$L(1, 0.9) = -1 \cdot \ln 0.9 - (1 - 1) \cdot \ln(1 - 0.9) \tag{9.7}$$

$$\approx -1 \cdot (-0.11) = 0.11, \tag{9.8}$$

which is lower than for the first prediction. The optimizer of the algorithm will therefore use the gradient of the loss function w.r.t. the weights to take the predicted values closer to their true values, and thereby reduce the loss.

The full loss function at the tth iteration (after t trees have been added to the ensemble), summing over all the n input events, can be expressed as

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{(t)}) \tag{9.9}$$

$$= \sum_{i=1}^{n} L(y_i, \hat{y}_i^{(t-1)} + f_t(\boldsymbol{x}_i)), \tag{9.10}$$

where f_t represents the tth decision tree, which is added as a correction to the prediction after iteration $t-1$, $\hat{y}_i^{(t-1)}$.

The trick is to use a *Taylor expansion* of the loss function $L(y_i, \phi(\boldsymbol{x}_i))$ to second order (the loss function therefore needs to be twice differentiable) in terms of small corrections $f(\boldsymbol{x}_i)$ around the predicted loss value $\hat{y}_i^{(t-1)}$ after $t-1$ iterations, and apply this correction in the tth iteration to get:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[L(y_i, \hat{y}_i^{(t-1)}) + \left.\frac{\partial L(y_i, \phi(\boldsymbol{x}_i))}{\partial \phi(\boldsymbol{x}_i)}\right|_{\phi(\boldsymbol{x}_i)=\hat{y}_i^{(t-1)}} f_t(\boldsymbol{x}_i) \right. \tag{9.11}$$

$$\left. + \frac{1}{2} \left.\frac{\partial^2 L(y_i, \phi(\boldsymbol{x}_i))}{\partial \phi(\boldsymbol{x}_i)^2}\right|_{\phi(\boldsymbol{x}_i)=\hat{y}_i^{(t-1)}} f_t^2(\boldsymbol{x}_i) \right] \tag{9.12}$$

$$\equiv \sum_{i=1}^{n} \left[L(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(\boldsymbol{x}_i) + \frac{1}{2} h_i f_t^2(\boldsymbol{x}_i) \right], \tag{9.13}$$

where g_i denotes the first order gradient of the loss function for event i, and h_i denotes the corresponding second order gradient, or hessian, for this particular event.

Expressing the loss function in terms of the weights instead, gives:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[L(y_i, \hat{y}_i^{(t-1)}) + g_i w_{t,q(\boldsymbol{x}_i)} + \frac{1}{2} h_i w_{t,q(\boldsymbol{x}_i)}^2 \right]. \tag{9.14}$$

Since the term $L(y_i, \hat{y}_i^{(t-1)})$ is constant and does not depend on the trees/weights, we can drop it for the expression to be minimized, which results in:

$$\tilde{\mathcal{L}}^{(t)} \approx \sum_{i=1}^{n} \left[g_i w_{t,q(\boldsymbol{x}_i)} + \frac{1}{2} h_i w_{t,q(\boldsymbol{x}_i)}^2 \right]. \tag{9.15}$$

Instead of expressing the total loss as an explicit sum over all events, we can sum over all the leaves of the tree, with an implicit sum over all events that fall in each leaf. In order to write this compactly, we define the instance (event) set $I_j = \{i | q(\boldsymbol{x}_i) = j\}$ that holds all events which are mapped to the leaf with index j. We can then express the total loss as a sum over the contributions from each

leaf as:

$$\tilde{\mathcal{L}}^{(t)} \approx \sum_{j=1}^{T} \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i \right) w_j^2 \right] \tag{9.16}$$

$$\equiv \sum_{j=1}^{T} \left[G_j w_j + \frac{1}{2} H_j w_j^2 \right], \tag{9.17}$$

where the sum of gradients and hessians in leaf j in the last expression has been replaced with G_j and H_j, respectively, and the tree index t on the weights now has been dropped for simplicity.

For a fixed tree structure $q(\boldsymbol{x}_i)$, we can calculate the optimal weight value for each leaf by differentiating $\tilde{\mathcal{L}}^{(t)}$ w.r.t. the weights w_j, and set $\partial \tilde{\mathcal{L}}^{(t)} / \partial w_j = 0$ to find an expression for w_j that minimizes the loss:

$$\frac{\partial \tilde{\mathcal{L}}^{(t)}}{\partial w_j} = G_j + H_j w_j \stackrel{!}{=} 0, \tag{9.18}$$

where the optimal weight value, w_j^*, is

$$w_j^* = -\frac{G_j}{H_j}. \tag{9.19}$$

The optimal weight therefore chooses a value that moves in the opposite direction to that of the gradient of the loss. E.g., if we vary the weight slightly upwards in the positive direction, and the gradient of the loss is also positive, we choose to go in the opposite direction (given by the negative sign in front), where the loss decreases. This resembles the standard, stochastic gradient descent method, where we would move down the slope of the gradient averaged over the events.

Here, however, we also use second order gradient information, which effectively adjusts the size of the weight update (step length) by dividing by the summed hessian values of each event in the leaf. This is in fact the same as Newton's method of finding the minimum of a function, and is in the context of decision trees also known as *Newton boosting*. By also using curvature information in the weight update, we will generally achieve a faster convergence to a local or global minimum, compared to only using the gradient, by taking a more direct path toward the minimum.

9.1.3 Regularized gradient boosting

One of the key developments of XGBoost compared with most traditional tree boosting algorithms, is that it adds more options for regularizing the tree structure and the corresponding weights during the learning process. Regularization is commonly used in the training of neural networks, but, except for simple pruning techniques, has not been applied as extensively to the training of decision trees

until more recently, with XGBoost being among the first and few to implement this.

Regularization means to add constraints to the learning algorithm or to add a penalty to the loss function for learning very complex models or very large weights. It thereby aims to control what and how much the model can learn, and to avoid overfitting to the training set.

There are two basic algorithmic hyperparameters that constrain how much the model can learn: `n_estimators`, which corresponds to the number of trees that will be trained, i.e., the number of boosting iterations; and `max_depth` (default: 6), which corresponds to the maximum number of levels (successive number of tests) that a tree can grow before it stops.

In addition to the above constraints on the learning algorithm, XGBoost can add three regularization parameters to the loss function that handle tree structure: `alpha` (α), which adds a penalty on the absolute value (L1 norm) of the weights, $|w_j|$, also known as L1 regularization; `lambda` (λ), which adds a penalty on the square (L2 norm) of the weights, w_j^2, also known as L2 regularization; and `gamma` (γ), which adds a penalty on the number of leaves (terminal nodes) of the tree, T. The corresponding regularization terms to be added to the loss function are then:

$$\Omega(f_t) = \alpha|w_j| + \frac{1}{2}\lambda w_j^2 + \gamma T, \tag{9.20}$$

where α (default: 0), λ (default: 1) and γ (default: 0) are hyperparameters[1] to be set prior to the training.

We then add the regularization terms to the loss function, where the combination of loss and regularization is called a *cost* or *objective function*:

$$\tilde{\mathcal{L}}^{(t)} \equiv \sum_{j=1}^{T}\left[G_j w_j + \frac{1}{2}H_j w_j^2 + \alpha|w_j| + \frac{1}{2}\lambda w_j^2\right] + \gamma T \tag{9.21}$$

$$= \sum_{j=1}^{T}\left[G_j w_j + \frac{1}{2}(H_j + \lambda) w_j^2 + \alpha|w_j|\right] + \gamma T. \tag{9.22}$$

The optimal weight, given the regularization constraints, then becomes:

$$w_j^* = \begin{cases} -\frac{G_j+\alpha}{H_j+\lambda} & \text{if } G_j < -\alpha, \\ -\frac{G_j-\alpha}{H_j+\lambda} & \text{if } G_j > \alpha, \\ 0 & \text{else.} \end{cases} \tag{9.23}$$

[1]Hyperparameters are fixed during the training and define constraints of the model. Decision tree algorithms typically perform well with the default values of the hyperparameters, but one can often obtain significant gains in performance of the model if the hyperparameters are tailored to the specific task at hand. A common approach is to train the model multiple times with different combinations of hyperparameters, and select the combination that resulted in the best model performance (e.g., lowest loss or highest accuracy).

Equivalently, we can express the optimal weight as:

$$w_j^* = -\frac{T_\alpha(G_j)}{H_j + \lambda}, \tag{9.24}$$

where

$$T_\alpha(G_j) = \begin{cases} G_j + \alpha & \text{if } G_j < -\alpha, \\ G_j - \alpha & \text{if } G_j > \alpha, \\ 0 & \text{else.} \end{cases} \tag{9.25}$$

If we plug the optimal weight back into the cost function, we get an expression for how much the cost is reduced by adding the new tree, also known as the *structure score*:

$$\tilde{\mathcal{L}}^{(t)} = -\frac{1}{2}\sum_{j=1}^{T} \frac{T_\alpha^2(G_j)}{H_j + \lambda} + \gamma T. \tag{9.26}$$

We can then evaluate whether a new split would be beneficial or not by adding the structure scores of the new left and right nodes, L and R, respectively, and subtract off the structure score we get if we do not make the split. The corresponding *gain*, that is, the negative of the difference in the cost reduction, is then equal to:

$$\text{gain} = \frac{1}{2}\left[\frac{T_\alpha^2(G_\text{L})}{H_\text{L} + \lambda} + \frac{T_\alpha^2(G_\text{R})}{H_\text{R} + \lambda} - \frac{T_\alpha^2(G_\text{L+R})}{H_\text{L+R} + \lambda}\right] - \gamma. \tag{9.27}$$

The γ parameter therefore effectively sets a lower limit on how much loss reduction is required in order to allow a split to be made.

9.1.4 Shrinkage

Another parameter that can be used to avoid overfitting is the so-called shrinkage, or `learning_rate` parameter, also called `eta` (η, default: 0.3). The shrinkage parameter is multiplied with the weights after boosting, i.e., after the tree structure and weights have been settled. This effectively reduces the size of each update, so that each tree has less effect on the prediction. This results in a more controlled path, that is less likely to be affected by random fluctuations in the training set, but on the other hand leads to slower learning, and thereby the need for more trees to converge to a local or global minimum.

9.1.5 Subsampling of rows and columns

In order to promote diversity in the different trees of the ensemble, XBGoost offers the option to randomly select a subsample of the rows (events) and columns (variables) for the training of any given tree.

The act of subsampling rows, or events, with replacement, is known as *bagging* (bootstrap aggregating) [209], and has been widely used in different

areas of statistical modeling. Each tree will therefore be trained on a different subset of the events in the training set, where the size of the subsample is a hyperparameter to be determined before the training starts. In XGBoost, this parameter is called `subsample` and is given as the fraction of the events in the full dataset to be sampled (default: 1, i.e., 100% of the events).

The concept of randomly subsampling columns, or features/variables, is a more recent implementation in training of decision trees. It has become very popular through the *Random Forests* [210] algorithm, which implements subsampling of both rows and columns. In XGBoost, there are three hyperparameters for subsampling of columns: `colsample_bytree`, `colsample_bylevel` and `colsample_bynode` (default: 1, for all three), which corresponds to fractions of the columns to be subsampled for each tree, for each level in the tree and for each node in the tree, respectively. These column subsampling methods can be combined, and will then work in a cumulative manner. I.e., only columns that pass the subsampling by tree are passed to the subsampling by level, and only the columns that pass the subsampling by level are passed to the subsampling by node.

9.1.6 Pruning

The default way of constructing a tree is to first build the tree level by level until the maximum depth has been reached, and then to prune the leaves of the tree if necessary.

The first pruning criterium, in fact, enters already in the build process, or the initial growing of the tree. This is yet another hyperparameter, called `min_child_weight` (default: 1), and corresponds to the minimum required sum of weights in a leaf. If a split is made that results in one of the leaves not satisfying the minimum required sum of weights, the two leaves from this split are pruned, and the split that produced them will therefore be removed.

After the build step, the algorithm will iterate through all the nodes in the tree and check if the gain of each split is greater than the minimum required split loss set by the `gamma` (γ) parameter, also known as `min_split_loss` (default: 0). If the loss reduction does not meet this requirement, the split along with its leaves will be pruned away.

9.2 Binary classification

We formulate the problem in terms of binary classification, where we have events coming from two classes: (i) background, assigned the class label 0 (also referred to as the negative class, N), and (ii) signal, assigned the class label 1 (also referred to as the positive class, P). The output of the model is a real value between 0 and 1, and the objective for the algorithm is to optimize the weights of the model to estimate a number that is as close to the true class label of each event as possible. One can interpret this as estimating some sort of pseudo-probability that a given event is a signal event, as a function of the variables of

the event and the weights of the model. This is in principle the same as logistic regression.

The simplest way to assign a predicted class label to a given event is to say that if the output of the model is less than 0.5, then we assign the event to the background class (0), otherwise, we assign it to the signal class (1). With a decision boundary in place between the two classes, we can count the number of events that the model gets right and wrong in each class. Figure 9.2 shows an illustration of the MVA output distributions of two classes (the positive class shown in red and the negative in blue colors) and how the decision boundary (shown in green) separates which data points (events) get assigned to each class. Given a chosen decision boundary, we can count the number of true and false positives and negatives (TP, FP, TN and FN, respectively), and calculate the corresponding true and false positive and negative rates ($TPR = TP/P$, $FPR = FP/P$, $TNR = TN/N$ and $FNR = FN/N$, respectively). Since $FPR = 1 - TPR$ and $FNR = 1 - TNR$, we have an overcomplete scenario with four figures of merit for a two dimensional problem. Therefore, the performance of the MVA model can be completely specified by any pair of the aforementioned metrics. One possible choice is to plot a ROC-curve in terms of the *true positive rate* (TPR), i.e., the *signal efficiency*,

$$\text{signal efficiency} = \frac{TP}{P} = \frac{TP}{TP + FN}, \tag{9.28}$$

and the *true negative rate* (TNR), i.e., the *background efficiency*,

$$\text{background efficiency} = \frac{TN}{N} = \frac{TN}{TN + FP}. \tag{9.29}$$

We are more interested in getting the relatively few expected signal events right, than getting the many background events right. That is, we are willing to sacrifice background efficiency for as high *signal efficiency*, also called *recall*, as possible:

$$\text{recall} = \text{signal efficiency}. \tag{9.30}$$

We also want to maximize the ratio of signal to background events that are assigned to the signal class (the signal region), i.e., we value high true *signal purity* in the signal region, also called *precision*, more than high true background purity[2] in the background class:

$$\text{precision} = \frac{TP}{TP + FP}. \tag{9.31}$$

The concepts of precision and recall are illustrated in Figure 9.3.

These principles can be used to find where to put the decision boundary, or cut in the output distribution, to decide which events get assigned to the signal

[2]Background purity $= TN/(TN+FN)$, where TN denotes true negative events (background correctly classified as background) and FN denotes false negative events (signal falsely classified as background).

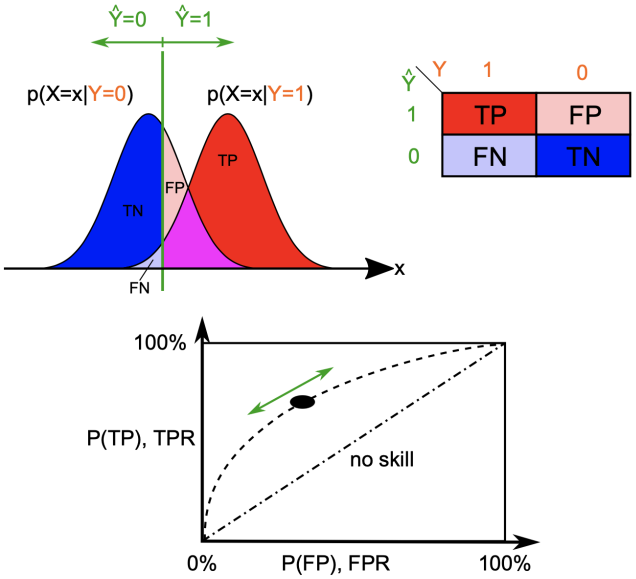


Figure 9.2: The upper left plot shows an example of two overlapping MVA output distributions (bell curves), where the left-most distribution consists of data points (events) from the negative class, N, with true value $Y = 0$, and the right-most distribution consists of data points (events) from the positive class, P, with true value $Y = 1$. The MVA model calculates a predicted output value, $\hat{Y}$, for each data point, based on the feature (variable) values of the data point. The decision boundary, shown in green, is used to assign a class label to each data point, depending on which side of the boundary the data point falls. The decision boundary is a free parameter for us to determine, depending on the performance of the model with the given hyperparameters and dataset, and which data points are most important for us to get right. Given the decision boundary, there are four numbers that quantify the accuracy of the classifier: true positives (TP) and true negatives (TN), which are the number of data points that are correctly classified, and false positives (FP) and false negatives (FN), which are the number of data points that are wrongly classified (fall on the wrong side of the decision boundary). The upper right plot shows a so-called confusion matrix, where a balanced and well-performing binary classifier should have most of the data-points on the TP-TN-diagonal. The bottom plot shows a Receiver Operating Characteristic (ROC) curve, with the false positive rate ($FPR = FP/P = FP/(TP + FP)$) on the x-axis and the true positive rate ($TPR = TP/P = TP/(TP + FP)$) on the y-axis. The curve is traced out by scanning decision boundaries in the entire range of the MVA output, and for each decision boundary calculate the corresponding TPR and FPR. A ROC curve following the diagonal indicates an MVA model with no binary classification skill, as it corresponds to random guessing (right and wrong at an equal rate). From Reference [211] (the original figure file is licensed under Creative Commons Attribution-Share Alike 3.0 Unported).

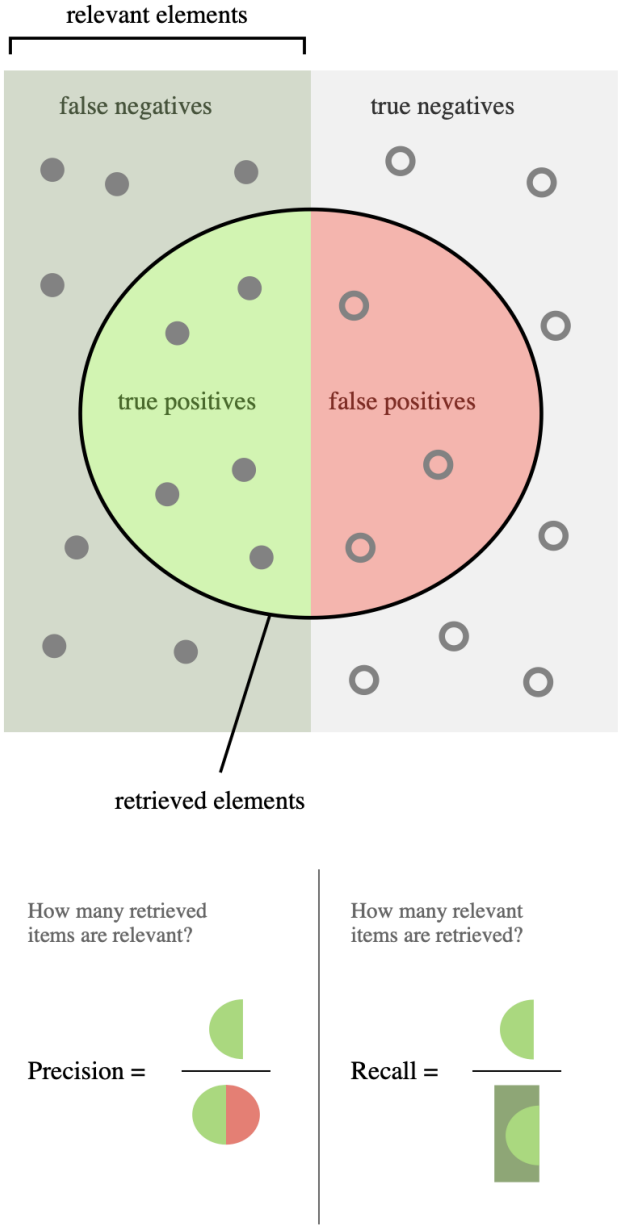


Figure 9.3: Illustrations of precision and recall in binary classification. These quantities are relevant when one of the two classes is considered more important than the other. From Reference [212] (the original figure file is licensed under Creative Commons Attribution-Share Alike 4.0 International).

region. Typically, we obtain the highest signal sensitivity by putting the decision boundary at relatively high output values, e.g., above 0.9. The optimal cut is found by performing a significance scan over multiple cut-values in the range between 0.5 and 1..

9.3 Train, validate, test

Since we have a lot more simulated background than signal events available, we encounter the issue of *imbalanced datasets*. To avoid that the model effectively neglects the signal compared to the background events during training, I make use of so-called *class weights* or `scale_pos_weight`, which is the name of the parameter in the XGBoost library. By setting `scale_pos_weight` equal to the ratio of background to signal events in the training set, the signal (positive class) events are scaled up by this factor, so that the sum of weights for each class is equal. Even though the decision trees are trained on a lot more background than signal events, the signal events count a lot more, so the cost of getting the signal events wrong is much higher, and we thereby effectively achieve a balanced training scenario.

In order to take event characteristics into account, the signal grid is divided into three regions according to the size of the mass-splitting, in exactly the same way as is done for the conventional analysis in Chapter 8. Hence, we are training separate models for the three mass-splitting scenarios: high, intermediate and low mass-splittings.

It is common practice to divide the full dataset into three exclusive portions for use in machine learning: (i) a *training set*, for weight-optimization on multiple sets of hyperparameters, (ii) a *validation set*, for evaluating the performance of the weights learned during training, which is re-used for every new set of hyperparameters, and finally, (iii) an independent *test set*, only used once, for evaluating the performance of the model with the hyperparameters that showed the best performance on the validation set. Typically, the training set is chosen to be larger than the validation and test sets, to get as high statistics as possible for training.

Since the model optimizes its weights on the training set, it will be biased toward that portion of the data, i.e., possibly overestimating its classification performance on those events, and we can therefore not use these events for estimating the expected signal and background contributions in the signal region. It is crucial to have an unbiased estimate of the expected background in the signal region, which will be used to compare the experimentally observed data counts with and to determine whether we see an excess or not. Together with the expected background, the expected signal yields are also important for determining exclusion limits, if no excess is observed. We are then left with finding a compromise between sufficient training statistics for learning the optimal weights of the model, and having sufficient test statistics to be used for the final signal region estimates. Since we are effectively losing luminosity (or cross-section) to the training set, we need to account for the lost background

Table 9.1: Event and object selection for the MVA.

Variable	Preselection XGBoost
SFOS leptons ($p_{\mathrm{T}} > 25\,\mathrm{GeV}$)	2
central light jets ($p_{\mathrm{T}} > 30\,\mathrm{GeV}$)	2
b-jets ($p_{\mathrm{T}} > 20\,\mathrm{GeV}$, 77% b-tag eff.)	0
$m_{\ell\ell}$ [GeV]	$\in (71, 111)$
m_{jj} [GeV]	$\in (60, 100)$

and signal contributions when comparing with the integrated luminosity of the observed data. By multiplying by the appropriate correction (scale) factor, we are also blowing up the statistical uncertainties of the MC events correspondingly, and thereby decreasing our sensitivity to the signal.

An alternative to the strategy described above, is to train two models instead of one (for each mass-splitting region), where one of the models are only trained and validated on events with *odd run numbers*, and tested on events with *even run numbers*, and vice versa for the other model. Since the test events from both odd and even run numbered events are scored on a model they have not been trained on, we can combine the test set output from the two models in the end, and thereby avoid losing any statistics for the final signal region estimates. This is the approach that has been chosen for the following studies.

9.4 Preselection

In order to select events which are relevant for the signal scenario we are looking for, I apply a set of event and object level requirements that the events must satisfy in order to be considered for the MVA. These cuts are listed in Table 9.1. They are intended to be exclusive enough so that the training will not have to process lots of events that are not really relevant for the search, but still inclusive enough to have sufficient training statistics to learn the general features that distinguish the signal from the backgrounds.

After the preselection cuts have been applied, we are left with about 4×10^6 background events and about 6×10^4, 9×10^4 and 4×10^4 signal events in the low, intermediate and high mass-splitting regions, respectively. This means that we have roughly 100 times more background events compared to signal events in each SR. In addition, these datasets are split, roughly in half, according to whether the run numbers are even or odd, where one half will be used for training and the other half for testing the performance of the model after training. This leaves me with about 3×10^4, 4.5×10^4 and 2×10^4 signal events for training each model in the respective mass-splitting regions, and roughly 2×10^6 background events which can be re-used for all three regions.

In order to monitor the performance progress during the training, a validation set is sampled from the training set before the training starts. In the following studies, 10% of the training set is set aside for validating the loss on events not used in the actual model training, after each new tree has been added. Therefore,

in practice, only 90% of the events in the training set are actually used for the model optimization.

9.5 Low mass-splitting

The low mass-splitting scenario, characterized by $\Delta m(\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0, \tilde{\chi}_1^0) \sim m_Z$, and thereby final states with soft (low-p_{T}) objects, is here taken to consist of signal models with $\Delta m < 200\,\mathrm{GeV}$. Due to the low final state momenta, and thereby also low $E_{\mathrm{T}}^{\mathrm{miss}}$, we cannot rely on the typical SUSY-signatures, like large $E_{\mathrm{T}}^{\mathrm{miss}}$, to the same extent anymore. The signal events will blend more and more in with the SM background processes as the $E_{\mathrm{T}}^{\mathrm{miss}}$ value decreases. Therefore we need to refine the selection and bring in other aspects of the event kinematics, correspondingly.

We will here try to design a multivariate analysis, using the XGBoost algorithm, to learn a more sophisticated selection model that can make up for some of the lost sensitivity of the conventional analysis approach compared to the signal scenarios with more $E_{\mathrm{T}}^{\mathrm{miss}}$. In the training process, the models will hopefully learn good values for a number of parameters, here referred to as weights, that define the behavior or selection of the classification model. However, there are also a number of parameters that are not being learned by the training routine, i.e., they are constant during the training, and that have to be set by us. They are referred to as hyperparameters, and typically pertain to the architecture of the model and define the scope of what can be learned. The hyperparameters chosen for the classification model that targets the low mass-splitting region, is given in Table 9.2, and the choices of values are discussed below. As the table header indicates, the same hyperparameter values were reused also for the intermediate and large mass-splitting models.

9.5.1 Hyperparameters

The `max_depth` of the decision trees, i.e., the maximum number of levels that can be used in a given tree, is reduced from XGBoost's default value of 6, to 5. This restricts the maximum number of decisions in each tree, which reduces the possibility of overfitting, i.e., learning statistical "noise" in the given training set, which does not reflect the general behavior of the underlying distributions.

The number of trees, or `estimators`, to train, where each consecutive tree tries to correct the mistakes done by the previously trained trees, is here set to a large enough number so that an early stopping mechanism kicks in before the specified number of trees is reached. A value of 10,000 is used in the following studies. What this means is that the loss on a validation set, not used in the training, is calculated after every new tree has been trained, and if the validation loss has not decreased in a given number of steps/updates, the training is stopped in order to avoid further overfitting. The `early_stopping_rounds` parameter is here chosen to be 20.

The `learning_rate` controls how much the weights can be updated compared to the size of the gradient of the loss function at a given position on the loss-

manifold. A more conservative value of 0.001 has been chosen here, compared with the XGBoost default value of 0.3. This slows down the learning, thereby requiring more steps to reach the optimal solution, but with the aim of achieving a more well-behaved progression that avoids overfitting or divergence.

The `gamma` parameter, also know as `min_split_loss`, is a measure of how much the loss has to decrease in order to allow another split on a node in the tree to be made. The default value is 0, so the chosen value of 1000 is a much more conservative one, that aims to reduce overfitting by making sure that the trees only learn sizable differences between signal and background, and avoid picking up dataset specific noise that does not reflect the underlying distributions.

The `min_child_weight` parameter is a threshold on the sum of instance weights needed in the child of a node after a new split has been made. This makes sure that there is a minimum contribution to both children in a split, in order for the split to be made. The default value of 1 has been used here.

`max_delta_step` sets an upper limit on the value of a leaf, in order to restrict the allowed importance of a single decision, to promote contributions from multiple features. The default value of 0 is chosen here, which turns off this constraint.

`subsample` represents the fraction of events in the available training set that is used for the training of a given tree. This parameter is here set to 0.5, which means that half of the events in the training set are randomly sampled before each tree is trained. This choice exposes each tree to a different collection of events (bagging), which aim to learn different aspects of the dataset and widen the focus of what is being learned. The default value is 1, which means that all events in the training set would be considered in the training of every tree.

`colsample` is similar to `subsample`, but introduces a random sampling of the columns of the dataset, that is, which features or variables are being considered in the training of a given tree. This parameter is chosen to be 0.5, which then means that each tree only sees half of the variables, and each tree gets a different, randomly chosen, set of variables to learn from. This also makes the learning more versatile, and reduces overfitting. The XGBoost default value of `colsample` is 1.

Two regularization terms that are commonly added to loss/cost functions for neural network training, are so-called L1- and L2-regularization terms, which add penalties to the L1 (absolute value)- and L2 (squared value)-norms of the weights that are being learned, respectively. L1-regularization encourages sparse representations, where less important weights are set to 0, leaving a reduced number of features that effectively contribute to the decisions. L2-regularization, on the other hand, also called weight-decay, penalizes large weights more than L1-terms, because the penalty is proportional to the square of the weights, and therefore encourages low-value weights, in order to restrict the learning and to avoid overfitting. XGBoost refers to the coefficient in front of the L1-norm of the weights as `alpha`, and, similarly, the coefficient in front of the L2-norm term as `lambda`. Here, L1-regularization has been turned off (`alpha` is set to 0), which is the XGBoost default value. The L2-regularization, however, is chosen to be

Table 9.2: Hyperparameter values used in the training of the XGBoost models of this work.

Hyperparameter	XGBoost-low, -int and -high
`max_depth`	5
`estimators`	10,000
`learning_rate`	0.001
`early_stopping_rounds`	20
`gamma`	1000
`min_child_weight`	1
`max_delta_step`	0
`subsample`	0.5
`colsample_by_tree`	0.5
`alpha` (L1)	0
`lambda` (L2)	1000

active with the coefficient `lambda` set to 1000, which puts a significant penalty on learning large leaf weights.

9.5.2 Learning curves

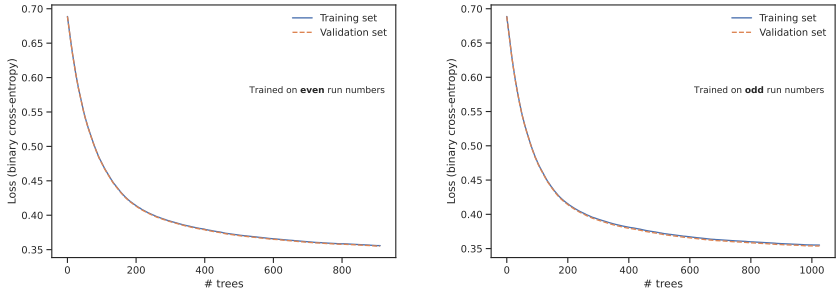
Figure 9.4 shows the evolution of the model performance during the training phase, as a function of the number of trees that have been added to the ensemble, which is calculated after each subsequent tree has been trained and added to the model. The performance is here given in terms of the total cost (loss plus regularization terms) on the training set (blue) and the total loss (no regularization terms) on the validation set (orange). The model trained on even numbered events is shown on the left, and the model trained on odd numbered events is shown on the right.

We can see how the performance of the models quickly progresses with the first 20 trees, or so, before the rates of progress start to flatten out. At some point, the validation loss, a measure of the generalization error, reaches a minimum, at which that particular model, with the given hyperparameters, is not able to do any better. The training loss, however, will just keep on improving as the model is being overfitted to the training set, which in turn will result in increased validation loss. Therefore, we would like to stop the training around the point where the validation loss has its minimum.

In this work, the training is stopped when the validation loss has not improved, i.e., been reduced, in 20 consecutive trees. From Figure 9.4, we can see that this happens after roughly 1000 trees have been added to each of the models.

9.5.3 Example of a learned decision tree

The first tree learned by the model trained on even run numbers is visualized as a graph in Figure 9.5. The root node, i.e., the top node of the tree graph, represents the best split found by the algorithm on any of the available features. In Figure 9.5, this turns out to be $E_{\mathrm{T}}^{\mathrm{miss,sig}} < 3.21$, where events fulfilling this requirement are found to be more background-like, and events with $E_{\mathrm{T}}^{\mathrm{miss,sig}} >$



(a) Model trained on even run numbers (b) Model trained on odd run numbers

Figure 9.4: Learning curves of the low-Δm models, which show the performance progressions (decrease in loss) as functions of the number of trees added to the models.

3.21 to be more signal-like. After the first split has been determined, the algorithm moves on to scan for the best split in each of the child nodes. For events with $E_{\mathrm{T}}^{\mathrm{miss,sig}} < 3.21$ (left side of the tree), the split that best separates background and signal is found to be $H_{\mathrm{T},30} < 154\,\mathrm{GeV}$, where events fulfilling this requirement are more background-like, and events with $H_{\mathrm{T}} > 154\,\mathrm{GeV}$ are more signal-like ($H_{\mathrm{T},30}$ represents the scalar sum of the p_{T} of all hadronic jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$). On the right hand side of the tree, however, for events with $E_{\mathrm{T}}^{\mathrm{miss,sig}} > 3.21$, the best split is found using $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ yet again, where $E_{\mathrm{T}}^{\mathrm{miss,sig}} < 5.09$ is more background-like and $E_{\mathrm{T}}^{\mathrm{miss,sig}} > 5.09$ more signal-like.

The tree is five levels deep, which corresponds to the hyperparameter value set for the training, `max_depth` $= 5$. This means that there are at most five decisions (node splits) along any branch down the tree, where the lowermost decision nodes have a pair of terminal leaf nodes, each holding the weight assigned to all events falling into that leaf node.

9.5.4 Training vs. test performance

After the training of a model has completed, all events in the training and the test set for that model are passed through all the trees of that model, and each event given a score in the range from 0 to 1, according to how signal-like the model deems it to be. Figure 9.6 shows the score distribution of the backgrounds in the training set using filled, stacked histogram bins, and the scores of the test set backgrounds overlaid on the same plot in dashed lines. The backgrounds are compared to the full set of signal samples in the low-Δm region, overlaid using a solid red line for the training set and a dashed red line for the test set events. Figure 9.6a shows the results of training events with even run numbers and test events with odd run numbers, whereas Figure 9.6b has training events with odd and test events with even run numbers.

If the training procedure has been successful in fitting a model to the training

set, which generalizes well, i.e., without significantly overfitting, the test scores should lie more or less on top of the training scores. On the other hand, if the training has failed in picking up characteristics of the underlying event distributions, i.e., showing significant overfitting to the training set, the distributions of the test scores will be skewed away from their true values.

The two models show decent skill on average, as the background distributions peak at low XGBoost values, i.e., close to their true class values of 0. The same goes for the signal distributions, which peak at high XGBoost scores, i.e., close to their true class values of 1. However, both the background and signal distributions are very wide, which means that the models struggle to recognize the true origin of relatively large proportions of the events. This can to a great extent be attributed to the background-like characteristics of the low-Δm signal samples.

Only minor deviations are observed between training and test scores in the two models in Figure 9.6, as both the background and signal test-to-training ratios, with statistical uncertainties, are close to unity, considering that no systematic uncertainties have been taken into account. The uncertainty bands are most likely underestimated compared to the true uncertainties in the modeling, which should include model-dependent uncertainties, such as variations w.r.t. hyperparameter values, in addition to uncertainties related to the simulation and reconstruction of the datasets.

The most prominent deviations are observed in the test-to-training ratio of signal roughly below XGBoost scores of 0.2, in particular in Figure 9.6a. These events correspond to signal events which have been characterized as highly background-like, and where the statistics is relatively low. It is, however, not critical to the analysis of this work that the modeling in the background-dominated region is sub-optimal, as long as it does not significantly affect the modeling in the signal-dominated region. The goal here is to find the (class) decision boundary, i.e., signal region, in the XGBoost score which optimizes the signal significance. In consequence, the events below the decision boundary, i.e., outside of the signal region, are not considered in the final results. The optimal signal region is typically found at XGBoost scores greater than 0.95, where it is important to have sufficient confidence in the modeling. One way of getting confidence in the modeling is to verify that the test results are consistent with the training results. Test vs. training consistency indicates a model which is unbiased w.r.t. the training set.

9.5.5 Feature importance

Figure 9.7 shows rankings of the top 20 most important features in the two models, given in terms of the total gain (see Equation (9.27)) obtained by making splits in the respective features. There are two features that stand out as particularly important in both of the trained models, namely $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}$. This is not surprising, as they are the most powerful variables in the conventional cut-and-count analyses as well, for separating signal from background. They obtain between two and three times more total gain compared to the next three

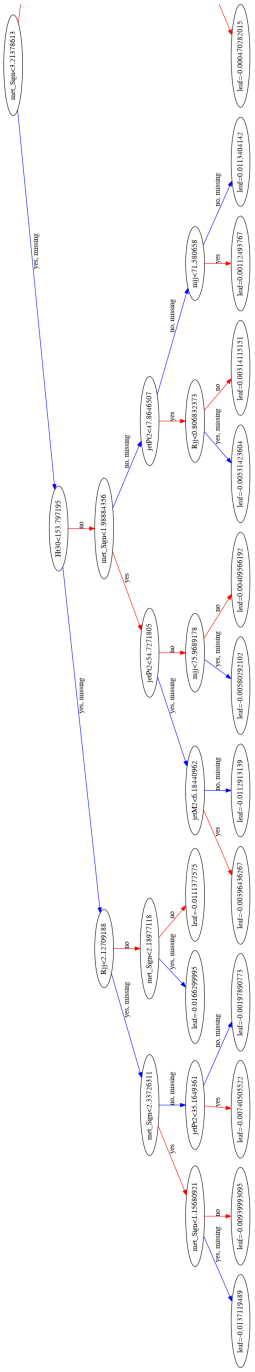


Figure 9.5: Left half of the first decision tree trained on even run numbers, using low-Δm signal samples. The flow of the chart follows the arrows downward from the root node of the tree, at the top right.

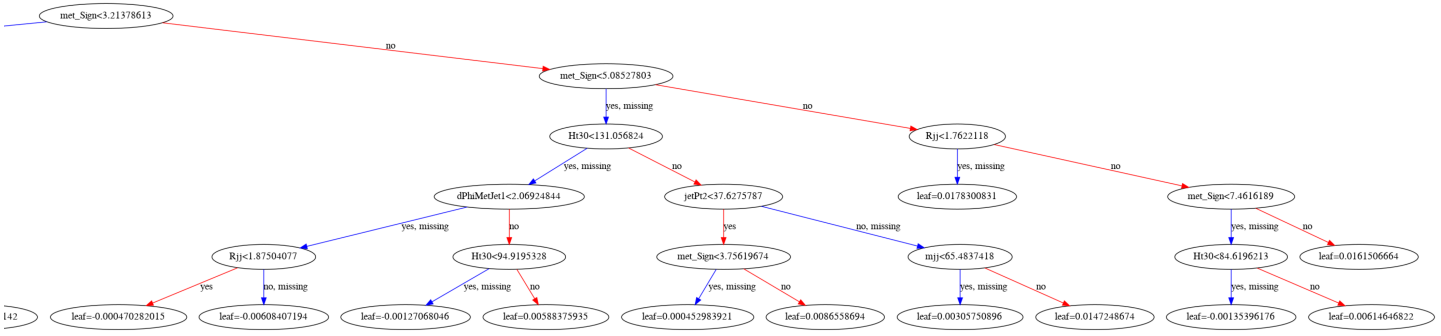
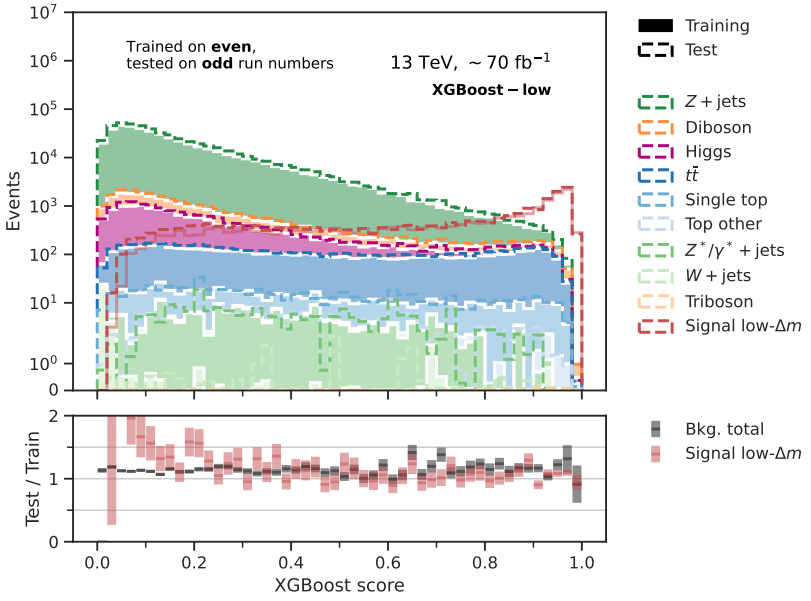
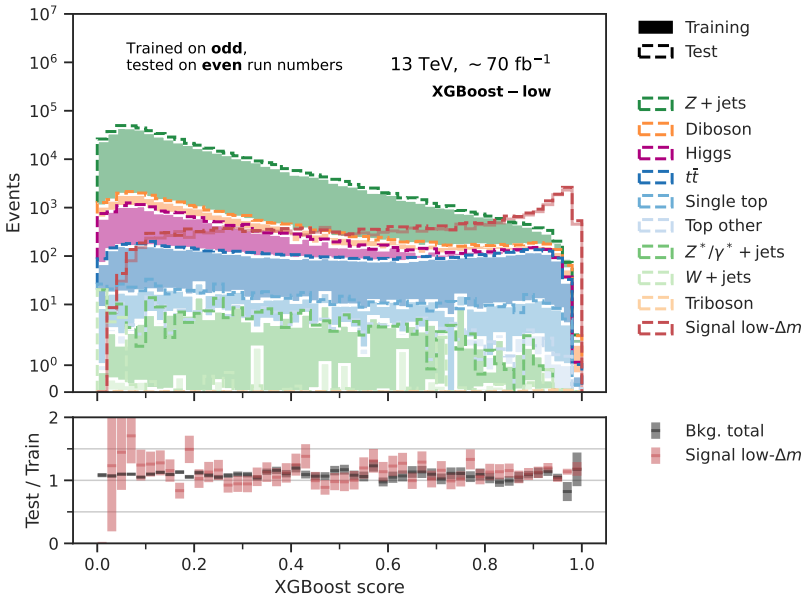


Figure 9.5: Right half of the first decision tree trained on even run numbers, using low-Δm signal samples. The flow of the chart follows the arrows downward from the root node of the tree, at the top left.



(a) Model trained on even and tested on odd run numbers



(b) Model trained on odd and tested on even run numbers

Figure 9.6: XGBoost output scores on the training (filled histograms/lines) and test (dashed lines) sets, for all backgrounds and the low-Δm signal samples. The bottom subplots show the bin-wise test-to-training ratio of the sum of background contributions in black and the sum of low-Δm signal samples in red. The uncertainty bands represent statistical uncertainties only.

features on the list, namely the angular separation between the two p_{T}-leading jets (assumed to be the decay products of the intermediate W-boson in the signal model), $\Delta R(j, j)$; the stransverse mass of the two leptons and the $E_{\mathrm{T}}^{\mathrm{miss}}$, $m_{\mathrm{T2}}(\ell, \ell; 0)$; and the hadronic p_{T} (scalar sum of the jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$), H_{T}.

Both the model trained on even run numbers and the model trained on odd run numbers agree on the importance-ordering of the two leading features, with $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ on top (16-17% gain) followed by regular $E_{\mathrm{T}}^{\mathrm{miss}}$ (14-15% gain). For the three following features on the list, however, the ordering is different between the models, although the features show similar importance in terms of total gain in the two cases, where all three have gains of roughly 7% each.

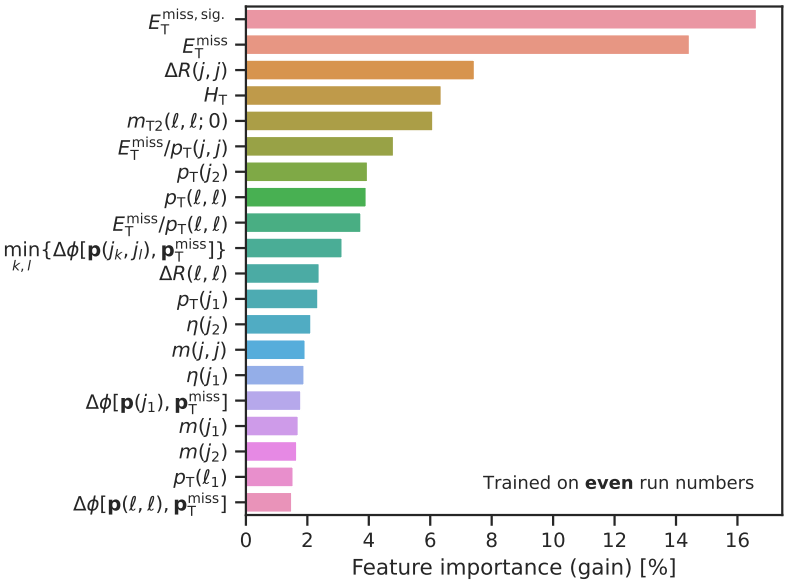
The remaining features on the top 20 list are also quite consistent between the two models, both in terms of the size of the gain and the ordering. This gives additional confidence in the consistency and stability of the models, and indicates that they are not significantly overfitting to statistical fluctuations in their respective training sets. They both seem to learn very similar representations, independently. This may also indicate that the models are, on the one hand, not significantly limited by low training set statistics, and on the other hand, are properly constrained by the hyperparameters and regularization terms, so that they converge toward similar model representations.

9.5.6 Binary classification performance

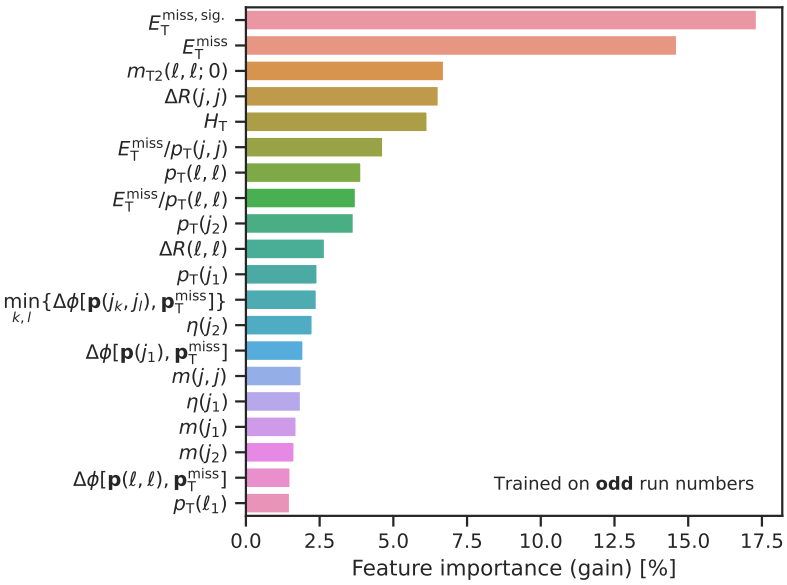
9.5.6.1 Background efficiency vs. signal efficiency

Figure 9.8 illustrates the performance of the two low-Δm models in terms of Receiver Operating Characteristic (ROC) curves, here expressed through the background efficiency (true negative rate) as a function of the signal efficiency (true positive rate). The ROC curves of the training sets are shown as solid blue lines, and the corresponding test set curves are shown as dashed blue lines. The line of no discrimination, i.e., the ROC curve of a model with no skill in separating signal from background, is shown as a dotted gray line. Each point along the ROC curve represents the background and signal efficiency at a particular decision boundary on the axis of the XGBoost score.

A common way to summarize the overall performance of the model, is to scan decision boundaries in the entire range of possible XGBoost scores, i.e., [0, 1], and to integrate the performance (background efficiency as a function of signal efficiency) of the corresponding decision boundaries. This is called the *Area Under the ROC curve* (AUC). The AUCs of the training and test set ROC curves are given in the plot legends. The AUC of a model with no discriminating power is equal to 0.5, whereas the AUC of a model with perfect discriminating power is equal to 1. The AUC of the two models on the training sets are calculated to be 0.918, with slightly lower test set AUCs of 0.910 and 0.913 for the models trained on even run numbers and on odd run numbers, respectively. These are decent results with significantly better skill than random guessing, but still with sizeable overlaps of the signal and background distributions. The small deviation

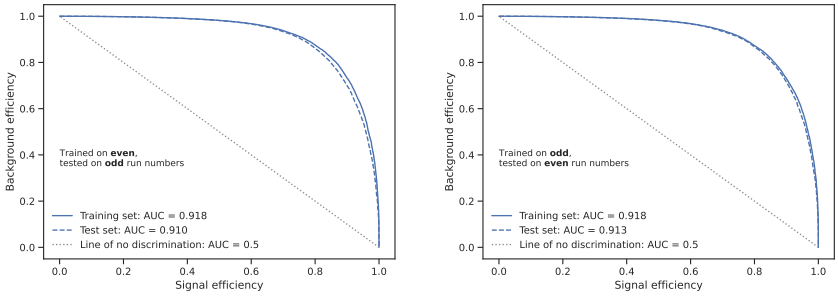


(a) Model trained on even run numbers



(b) Model trained on odd run numbers

Figure 9.7: Top 20 most important features in the low-Δm models, in terms of the total gain obtained in the full models (all added trees included).



(a) Model trained on even and tested on odd run numbers

(b) Model trained on odd and tested on even run numbers

Figure 9.8: ROC curves for models targeting low-Δm signal: background efficiency vs. signal efficiency.

between training and test set AUCs indicate that no significant overfitting to the training set is observed.

9.5.6.2 Background rejection vs. signal efficiency

Figure 9.9 gives another interpretation of the model performance, now in terms of the background rejection as a function of the signal efficiency for the two models. The background rejection is defined as:

$$\text{background rejection} = \frac{1}{1 - \text{background efficiency}} \tag{9.32}$$

$$= \frac{1}{1 - TN/(TN + FP)} \tag{9.33}$$

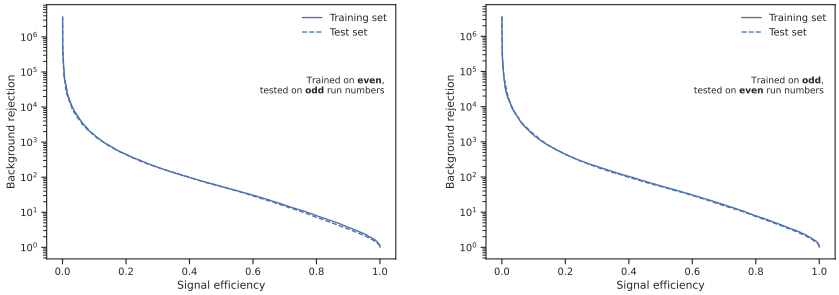
$$= \frac{TN + FP}{FP} = \frac{1}{FPR}, \tag{9.34}$$

where FPR is the false positive rate, i.e., the rate at which background events are mis-classified as signal, and thereby enters the signal region.

For example, given a decision boundary where the background efficiency is equal to 0.9, i.e., 90% or 9 out of 10 background events are correctly classified as background (true negatives, TN). This means that $1 - \text{background efficiency} = 1 - 0.9 = 0.1 = 1/10$, i.e., 1 out of 10 background events are mis-classified as signal (false positives, FP). Then, for every 10 background events there are 1 background event that ends up in the signal region, whereas the remaining 9 are correctly rejected, i.e., do not end up in the signal region. The corresponding background rejection therefore becomes

$$\text{background rejection (bkg. eff. } = 0.9)) = \frac{TN + FP}{FP} \tag{9.35}$$

$$= \frac{9 + 1}{1} = 10, \tag{9.36}$$



(a) Model trained on even and tested on odd run numbers

(b) Model trained on odd and tested on even run numbers

Figure 9.9: ROC curves for models targeting low-Δm signal: background rejection vs. signal efficiency, where background rejection is equal to $1/(1 -$ background efficiency), and represents the number of correctly rejected background events for every background event that ends up in the signal region (plus one).

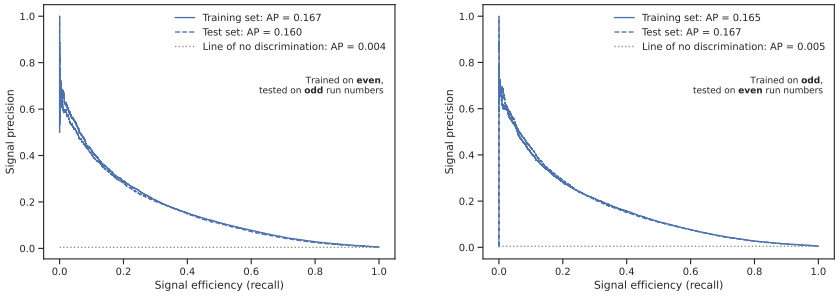
i.e., the total number of background events (negatives, $TN + FP$) is 10 times greater than the number of false positives (FP). Or equivalently, for every background event that ends up in the signal region, 9 are correctly rejected.

In Figure 9.9, we can observe that for, e.g., a signal efficiency of 0.5, the XGBoost-low models obtain a background rejection of roughly 100.

9.5.6.3 Signal precision vs. signal efficiency

A third way of representing the performance of the binary classifiers is by way of precision-recall (PR) curves, which are shown for the two models in Figure 9.10. Where the previous ROC curves showed the efficiency of background classification on one axis and signal classification on the other axis, the PR curve shows two different aspects of signal classification accuracy, namely the precision and recall of signal events in the signal region. Recall measures the fraction of signal events which end up in the signal region, whereas precision measures the fraction of events in the signal regions which are actually signal.

Where the signal and background efficiencies (true positive and negative rates) are only concerned with one class distribution each (either the positive (signal) or the negative (background) class distribution), precision mixes characteristics from both class distributions, where it measures the fraction of the positive (signal) and negative (background) class out of the total number of events in the signal region. Precision is therefore sensitive to class imbalance in the datasets, unlike the signal and background efficiencies. Consequently, the line of no discrimination in a precision-recall plot can be drawn at a constant precision value equal to the ratio of the number of events in the positive class to the number of events in the negative class. Since I have roughly 100 times more background



(a) Model trained on even and tested on odd run numbers

(b) Model trained on odd and tested on even run numbers

Figure 9.10: Precision vs. recall curves for models targeting low-Δm signal.

than signal events in the training and test sets, a model with no classification skill would uniformly distribute background and signal events with XGBoost scores in the interval between 0 and 1. The signal region would therefore also contain the same uniform distribution in its interval, where there would be 100 times more background than signal events, and consequently give a signal precision score of $1/100 = 0.01$.

E.g., for a signal efficiency, i.e., recall, of 0.5, the XGBoost-low models obtain a signal precision value of about 0.1.

We can also in a PR plot calculate one number that summarizes the performance of a model, similarly to the AUC for ROC plots. It is calculated in the same way as AUC, as the area under the PR curve, but is usually referred to as *average precision* (AP). The line of no discrimination corresponds to AP values of 0.004 and 0.005 for the model trained on even run numbers and on odd run numbers, respectively. The two numbers are slightly different because the ratio of signal-to-background events for even and odd run numbers are slightly different.

For the model trained on even run numbers, the AP is calculated to be 0.167 on the training set and 0.160 on the test set. For the other model, trained on odd run numbers, the AP is found to be 0.165 on the training set and 0.167 on the test set. The average performance on the training and test sets are thereby found to be similar also in terms of precision and recall.

9.5.7 Data vs. background

Figure 9.11 shows data compared to expected background and a benchmark signal from MC, in the XGBoost score interval $[0, 0.9]$, which is completely dominated by background. Since this analysis is only carried out as a feasibility study, the data in the interval $[0.9, 1]$, where the signal becomes non-negligible, is blinded, to avoid observation-bias for anyone who may carry out a similar analysis in the future. The plots show only results from the test sets of the

two XGBoost models. Figure 9.11a shows the results of events with odd run numbers only and Figure 9.11b the results of events with even run numbers only, whereas Figure 9.11c shows the combination of both models, as a sum of the two histograms.

Only one of the low-Δm signal samples is shown in the plots, namely the benchmark signal with $(m(\tilde{\chi}_1^{\pm}, \tilde{\chi}_2^0), m(\tilde{\chi}_1^0)) = (200, 100)$ GeV.

The lower panel of each plot shows the ratio of the number of data to the sum of background events per histogram bin. The statistical uncertainty of the sum of backgrounds in each bin is highlighted by gray bars. The ratio of data to background is roughly within $\pm 10\%$ of unity, which indicate relatively good agreement between the data count and the expected background in the unblinded region. The statistical uncertainties of the total background do not cover the deviation between data and background fully. However, they are conservative estimates of the full uncertainties, which would include systematic uncertainties from the datasets and the MVA procedure.

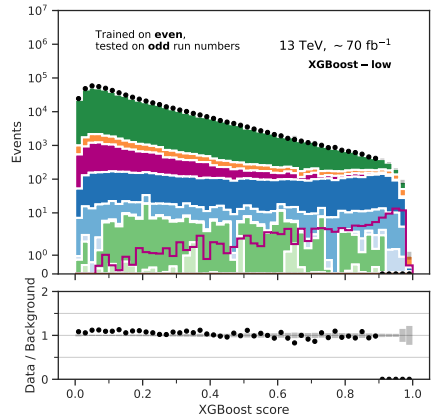
9.5.8 Benchmark signal vs. background

Finally, in Figure 9.12, we compare the test scores of the backgrounds to the test scores of the low-Δm signal benchmark, $(200, 100)$ GeV, in the highly signal-like region, $[0.9, 1]$. The lower-panel of the plots now show the Z_N significance of the low-Δm benchmark, as a function of the lower-cut on the XGBoost score. The test set scores of the model trained on even run numbers are shown in Figure 9.12a and the test scores of the model trained on odd run numbers are shown in Figure 9.12b, with the bin-wise sum of the two preceding histograms shown in Figure 9.12c.

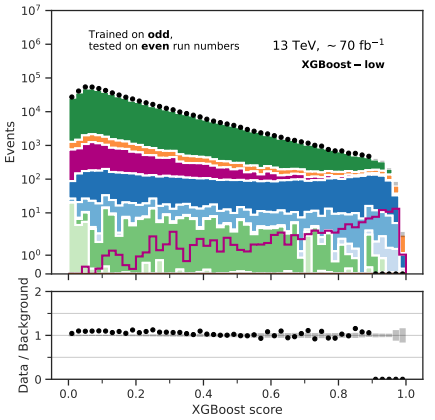
Since the test sets of the two models are mutually exclusive, we can add their output scores to form a common distribution. By scanning multiple decision values/boundaries on the score axis, we can find which one maximizes the signal significance. The corresponding score value then defines the lower cut of the MVA SR for the low-mass-splitting region.

Looking at the lower panel of the combined score distribution in Figure 9.12c, the significance increases gradually from $Z_N \approx 0$, at a lower-cut on the XGBoost score of 0.92, up to $Z_N \approx 0.8$, at a lower-cut of 0.97 on the XGBoost score. At least one background event and three signal events are required to calculate the Z_N-value. Otherwise, the Z_N-value is set to 0, in order to avoid regions with unsatisfyingly low background statistics and/or regions where the signal count is too low to exclude the signal at 95% C.L. This is what happens at the lower-cut on the XGBoost score at 0.98, where Z_N suddenly drops to 0.

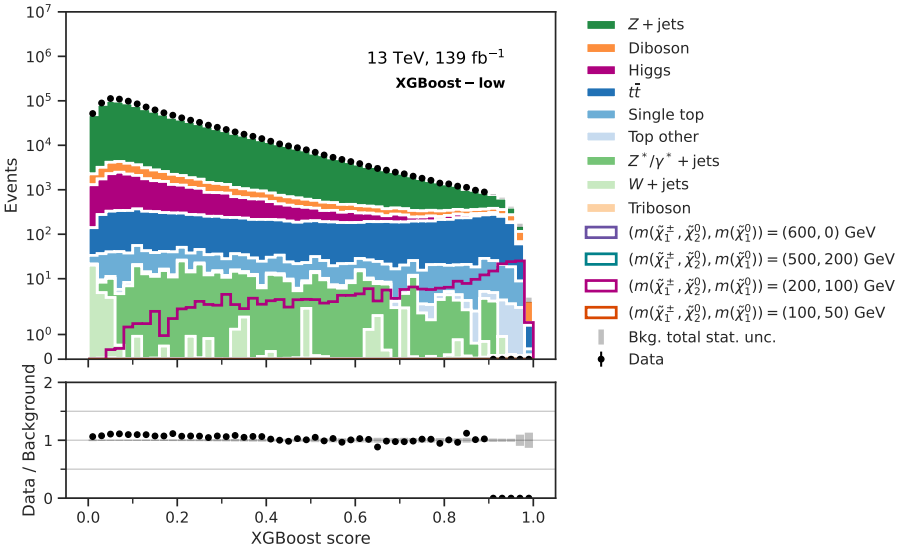
The optimal lower-cut is thereby found to be XGBoost score > 0.97 for the low-Δm signal benchmark, which I use to define the XGBoost-low SR. The Z_N-value for the low-mass benchmark signal point of (200, 100) GeV is 0.71, which is lower than 1.64. Therefore, we do not expect to exclude this mass hypothesis at 95% C.L. However, we still make use of the region to evaluate the signal significance of this benchmark and the other low mass-splitting signal mass points.



(a) Model trained on even and tested on odd run numbers.

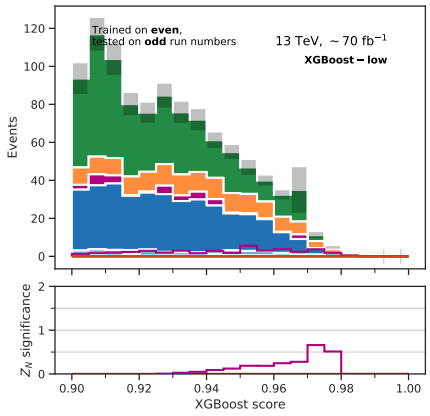


(b) Model trained on odd and tested on even run numbers.

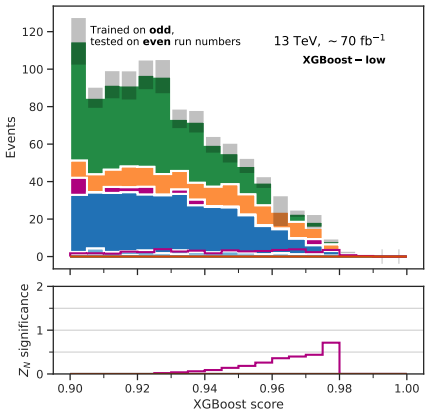


(c) Combination of both models.

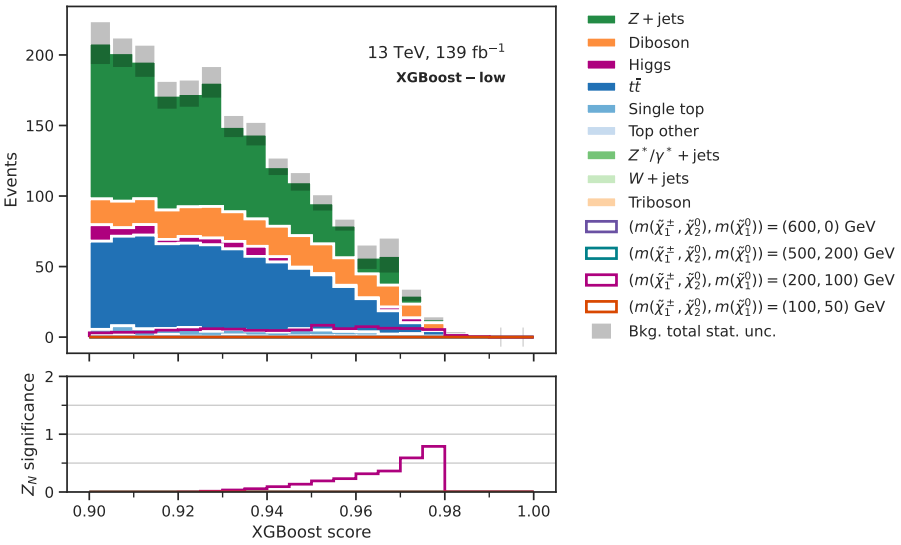
Figure 9.11: Data vs. expected background in the background-dominated XGBoost score region of $[0, 0.9]$, using the XGBoost-low model. The data points in the region with XGBoost score > 0.9 are not shown ("blinded").



(a) Model trained on even and tested on odd run numbers.

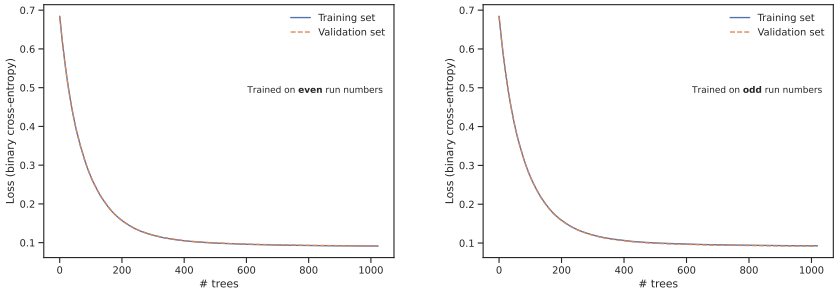


(b) Model trained on odd and tested on even run numbers.



(c) Combination (bin-wise sum) of the test outputs of both models.

Figure 9.12: Low-Δm benchmark signal sample vs. expected background in the highly signal-like XGBoost score region of $[0.9, 1]$, using the XGBoost-low model.



(a) Model trained on even run numbers (b) Model trained on odd run numbers

Figure 9.13: Learning curves of the intermediate-Δm models, which show the performance progressions (decrease in loss) as functions of the number of trees added to the models.

9.6 Intermediate mass-splitting

The same procedure is followed for the intermediate mass-slitting case, but now training and testing on signal samples with mass-differences in the range between 200 and 450 GeV. The same set of hyperparameter values used to train the XGBoost-low models are also used for the training of the intermediate-Δm models, which are given in Table 9.2.

9.6.1 Learning curves

Figure 9.13 shows the evolution of the cost (loss + regularization terms) for the training set and the loss only for the validation set, as a function of the number of decision trees that have been added to the ensemble model, for odd run number events on the left and even run number events on the right. Whereas the value of the loss function flattened out around 0.35 for XGBoost-low, we achieve loss values down to 0.1 for the XGBoost-int models. These are large improvements in performance, and suggests that the trained models are able to separate the signal and background distributions a lot better in the intermediate mass-splitting case than for more compressed signals, as expected.

This choice of hyperparameters shows no significant, visible sign of overtraining of the models in the plots of the learning curves, since the training and validation set losses almost lie on top of each other.

9.6.1.1 Training vs. test performance

This fact is also supported by the plots in Figure 9.14, which show comparisons of the output distributions for the training and test sets, where the dashed lines of the test set basically lie on top of the filled/solid line histograms of the training set. Both the signal and background distributions are shifted more toward their true class values then they were for the low-mass-splitting samples.

Due to the fact that the vast majority of signal events are assigned very high XGBoost scores, i.e., are recognized as highly signal-like, the statistical uncertainties in the ratio plots in the lower panels are small in the bins of XGBoost scores greater than roughly 0.8-0.9. The remaining signal events, which are assigned less signal-like XGBoost scores, lower than roughly 0.8-0.9, fall in bins with low statistics, of the order of ten events per bin. This results in relatively large statistical uncertainties on the signal yield in 80-90% of the full score interval. However, since we are really only interested in the events which fall into the highly signal-like region, this is a more desirable outcome than the corresponding distributions of the XGBoost-low models.

9.6.2 Feature importance

Figure 9.15 shows plots of the top 20 most important variables/features in the two XGBoost-int models. There are three variables that stand out in both models. For intermediate mass-splitting signals $E_{\mathrm{T}}^{\mathrm{miss}}$ turns out to be the most important feature, closely followed by $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and m_{T2}. They are responsible for roughly 20% of the total gain each, and 60% of the total gain combined. For comparison, the three top features in the XGBoost-low models account for about 38% of the total gain. The intermediate mass-splitting models therefore obtain better separation from fewer variables compared to the low-mass-splitting models.

The next features on the list, ranked as 4th and 5th, are ratios of $E_{\mathrm{T}}^{\mathrm{miss}}$ to the p_{T} of the two jets and two leptons, presumably coming from the W and Z bosons, respectively. Ranked 6th and 7th are the opening angles of the two jets, assumed to come from the W, and the p_{T} sum of all hadronic jets.

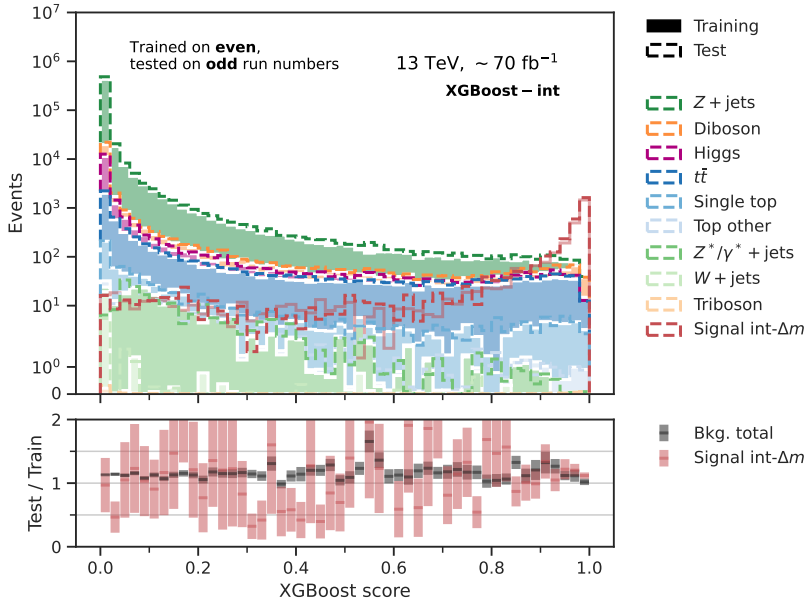
9.6.3 Binary classification performance

9.6.3.1 Background efficiency vs. signal efficiency

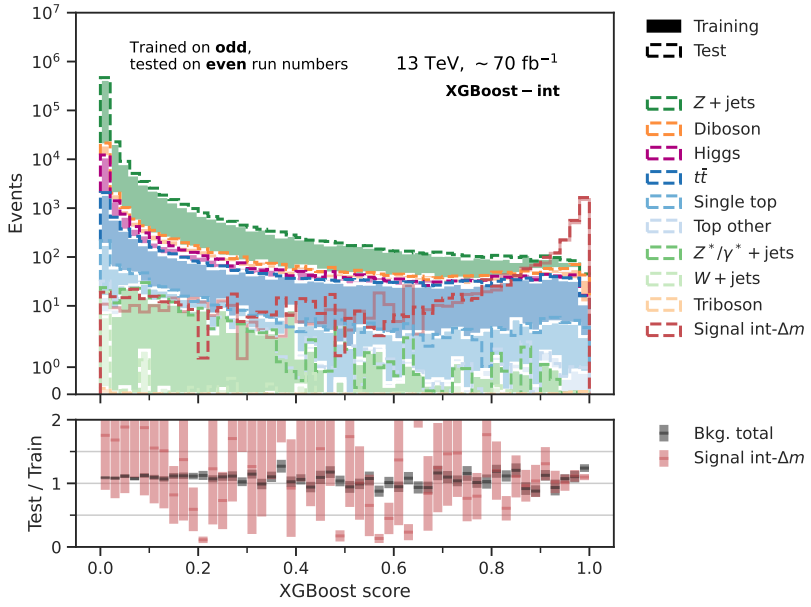
The ROC curves in Figure 9.16 are close to being perfect, with just below 100% signal and background efficiencies for almost any choice of decision boundary. This yields AUC scores of around 0.994 for both models, which are very close to the maximum value of one. However, this gives a somewhat misleading picture of the signal sensitivity, as it does not take the class-imbalance into account, i.e., the fact that there are roughly 100 times more background events than signal events.

9.6.3.2 Background rejection vs. signal efficiency

Figure 9.17 shows the ROC curves in terms of the background rejection as a function of signal efficiency. The good separation of the two class distributions here result in a high background rejection factor for a given signal efficiency. E.g., for a signal efficiency of 0.5, we will correctly classify and reject approximately 1000 background events as background for every background event we misclassify as signal.

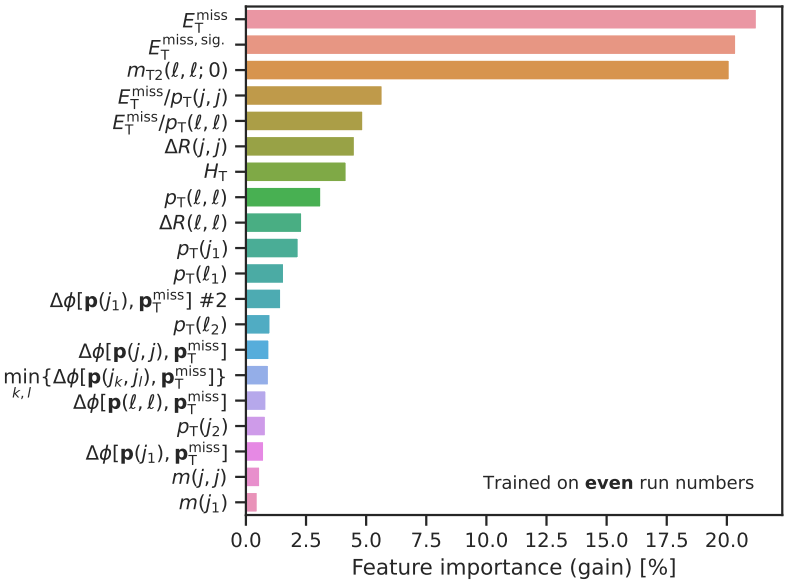


(a) Model trained on even and tested on odd run numbers

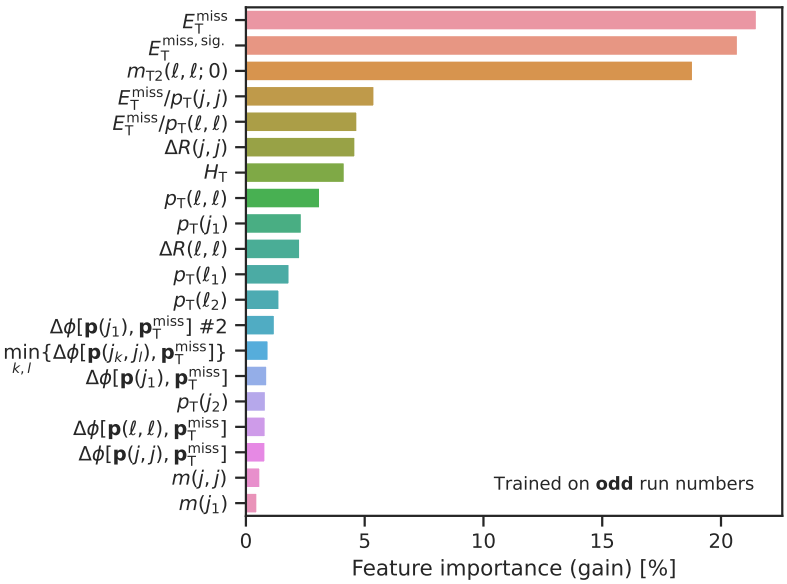


(b) Model trained on odd and tested on even run numbers

Figure 9.14: XGBoost output scores on the training (filled histograms/lines) and test (dashed lines) sets, for all backgrounds and the intermediate-Δm signal samples. The bottom subplots show the bin-wise test-to-training ratio of the sum of background contributions in black and the sum of low-Δm signal samples in red. The uncertainty bands represent statistical uncertainties only.

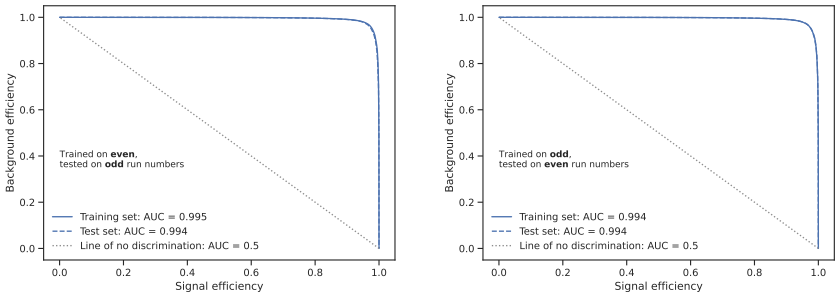


(a) Model trained on even and tested on odd run numbers



(b) Model trained on odd and tested on even run numbers

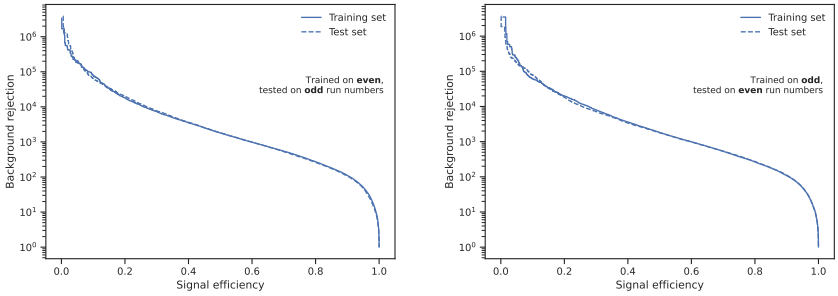
Figure 9.15: Top 20 most important features in the intermediate-Δm models, in terms of the total gain obtained in the full models (all added trees included).



(a) Model trained on even and tested on odd run numbers

(b) Model trained on odd and tested on even run numbers

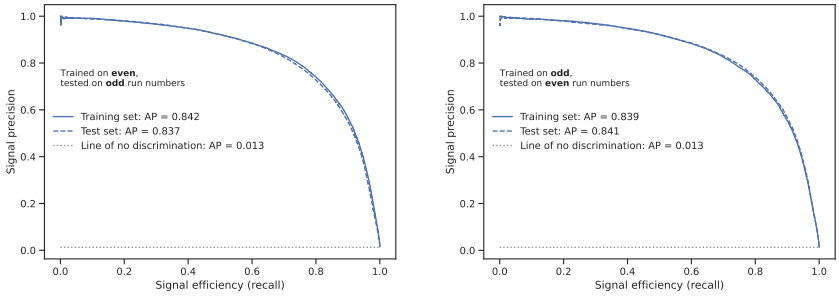
Figure 9.16: ROC curves for models targeting intermediate-Δm signal: background efficiency vs. signal efficiency.



(a) Model trained on even and tested on odd run numbers

(b) Model trained on odd and tested on even run numbers

Figure 9.17: ROC curves for models targeting intermediate-Δm signal: background rejection vs. signal efficiency, where background rejection is equal to $1/(1 - \text{background efficiency})$, and represents the number of correctly rejected background events for every background event that ends up in the signal region (plus one).



(a) Model trained on even and tested on odd run numbers

(b) Model trained on odd and tested on even run numbers

Figure 9.18: Precision vs. recall curves for models targeting intermediate-Δm signal.

9.6.3.3 Signal precision vs. signal efficiency

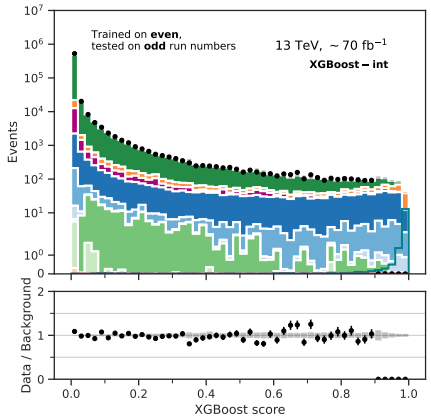
The precision-recall curves shown in Figure 9.18 give a more representative view of the signal sensitivity, compared to the standard ROC curves, because the number of events from both classes are compared in the signal region. For a signal efficiency, i.e., recall, of 0.5, the XGBoost-int models obtain a signal precision value of about 0.9, compared with 0.1 obtained by the XGBoost-low models. The XGBoost-int models obtain average precisions of about 0.84, which is a more modest result compared to the AUC of 0.99, but also a lot better than the average precisions of 0.16 and 0.17 obtained by the XGBoost-low models.

9.6.4 Data vs. background

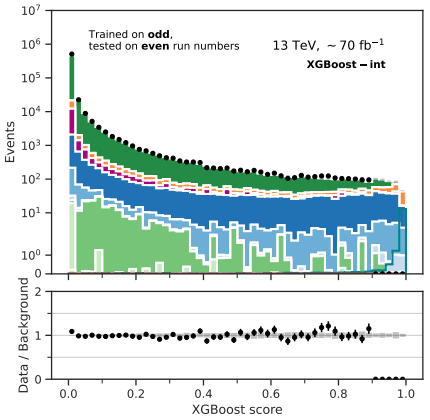
The data events are compared to the sum of simulated background events in Figure 9.19, for the highly background-dominated region of XGBoost score in the interval $[0, 0.9]$. The agreement looks convincing, especially in Figure 9.19c, which is the sum of the two histograms above, and thereby has roughly double the integrated luminosity of each of the summands. The fact that the agreement improves with more data adds confidence in the modeling. The total background uncertainties are statistical only, and therefore underestimates the full uncertainties, which should include systematic uncertainties from various sources.

9.6.5 Benchmark signal vs. background

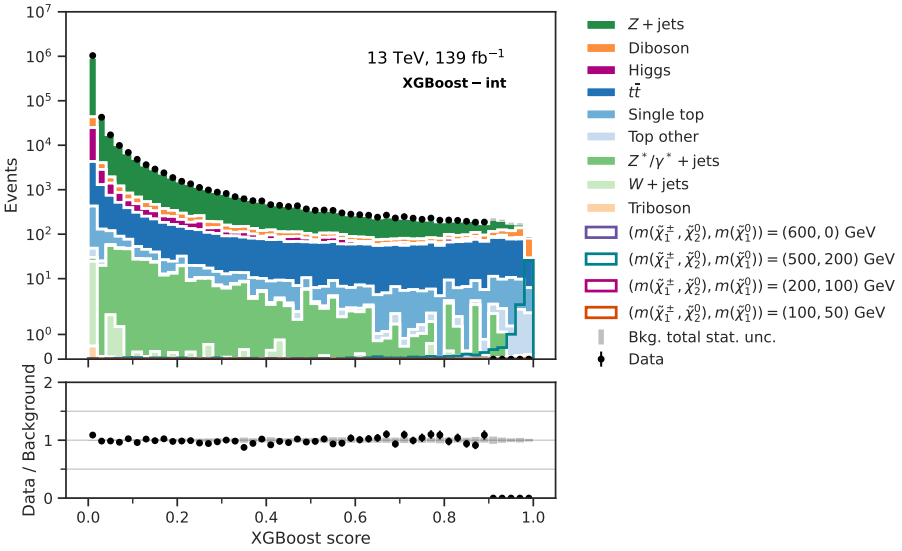
Figure 9.20 compares the intermediate-Δm benchmark signal, $(500, 200)$ GeV, to the total background in the highly signal-like region, with XGBoost score > 0.9. Only events from the test sets of the two XGBoost-int models are considered here. In the lower panels, the signal significance is shown as a function of signal region (lower) cuts.



(a) Model trained on even and tested on odd run numbers



(b) Model trained on odd and tested on even run numbers



(c) Combination of both models

Figure 9.19: Data vs. expected background in the background-dominated XGBoost score region of $[0, 0.9]$, using the XGBoost-int models.

The output from the two XGBoost-int models are combined in Figure 9.20c, which correspond to the full integrated luminosity of Run 2. Both the distributions of signal events, in the upper panel, and the signal sensitivity, in the lower panel, peak in the rightmost bin, which correspond to a signal region with XGBoost score > 0.995. The Z_N significance for the $(500, 200)$ GeV mass-point in this SR is 4.30. Since $Z_N > 1.64$, this SR has sufficient expected sensitivity to be able to exclude the $(500, 200)$ GeV signal at 95% CL, if no excess is observed in data.

9.7 High mass-splitting

The final set of MVA models, targeting high mass-splitting scenarios, are trained and tested on signal mass points with mass-differences greater than or equal to 450 GeV. The same set of hyperparameter values used to train the XGBoost-low and -int models are also used for the training of the high-Δm models, which are given in Table 9.2.

9.7.1 Learning curves

The learning curves for the XGBoost-high models are shown in Figure 9.21. The loss flattens out at around 0.05, which is significantly lower than for XGBoost-low at around 0.35, and even lower than XGBoost-int with losses slightly above 0.1. This means that an even greater accuracy is obtained for XGBoost-high than for any of the others. This is again expected, as this scenario deals with the phase space where the event topologies and kinematics of the signal differ the most from the background.

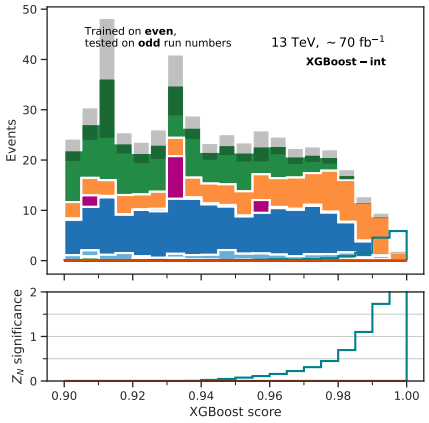
As the training and validation losses end up similar in values, there are no significant indication of overtraining.

9.7.2 Training vs. test performance

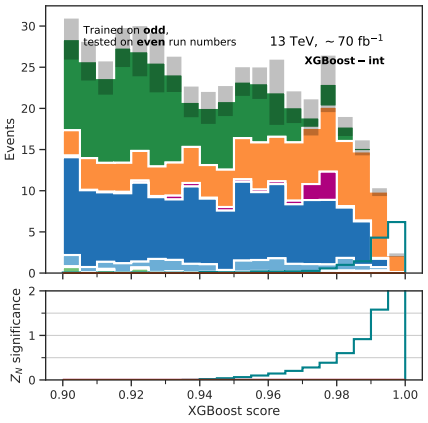
This is backed up by the plots of Figure 9.22, which show comparisons between the output scores of the training and the test sets. No large deviations between the solid lines of the training sets and the dashed lines of the test sets are observed. The distributions are now even more shifted towards their true class values compared to XGBoost-int, which means even better separation of the distributions in general.

The vast majority of the signal events are assigned XGBoost scores higher than 0.98 (the rightmost bin in Figure 9.22), both in the training and test sets. This means that most high-Δm signal events are recognized as highly signal-like.

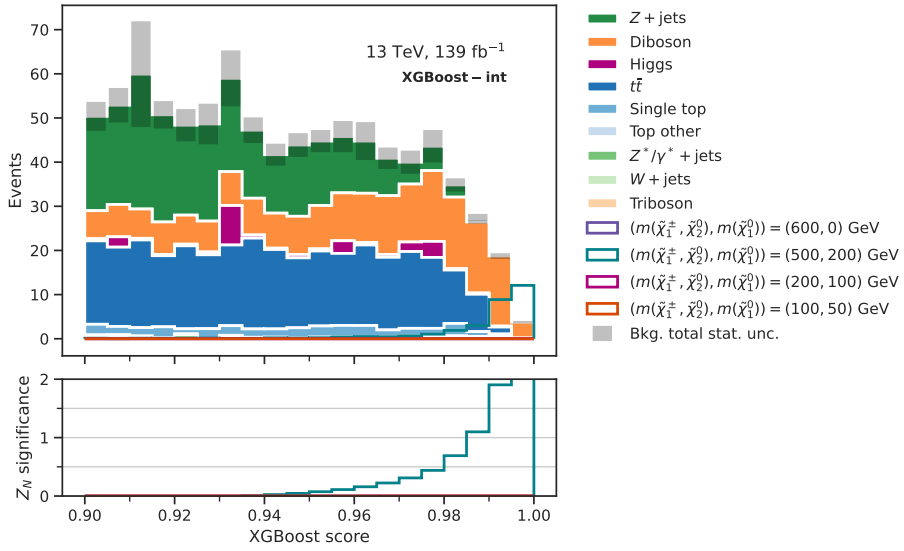
The XGBoost-high models show similar results for background, where the vast majority of background events are recognized as highly background-like, with XGBoost scores lower than 0.02 (the leftmost bin).



(a) Model trained on even and tested on odd run numbers



(b) Model trained on odd and tested on even run numbers



(c) Combination of both models

Figure 9.20: Intermediate-Δm benchmark signal sample vs. expected background in the highly signal-like XGBoost score region of $[0.9, 1]$, using the XGBoost-int model. The data points in the region with XGBoost score > 0.9 are not shown ("blinded")

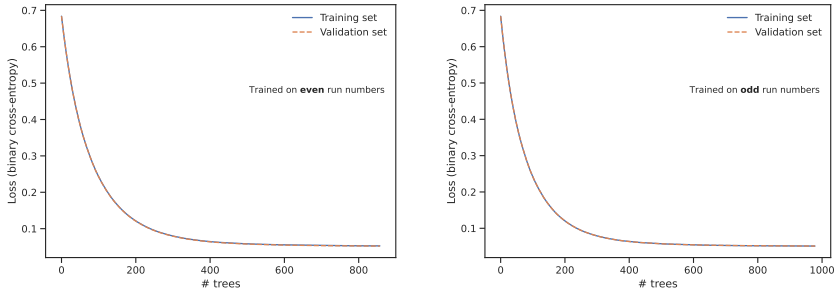


Figure 9.21: Learning curves of the high-Δm models, which show the performance progressions (decrease in loss) as functions of the number of trees added to the models.

9.7.3 Feature importance

As for XGBoost-int, the top 20 feature importance plot of the two XGBoost-high models, shown in Figure 9.23, have $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,sig}}$ and m_{T2} as the three outstanding features. They are also of similar combined importance, accounting for roughly 60% of the total gain, as was the case for XGBoost-int.

The next features on the list show similar distributions in terms of the size of the gains, decreasing from roughly 5% gain from the 4th most important feature and downwards. However, the order of features are slightly different. Where both of the XGBoost-int models had the ratios of $E_{\mathrm{T}}^{\mathrm{miss}}$ to the p_{T}s of W and Z ranked higher than the p_{T} of W and Z alone, both the XGBoost-high models rank the raw p_{T} slightly higher than the ratios. The opening angle of the jets coming from the W are also valued slightly lower. These differences in gain are, however, relatively small, and may not be of noteworthy significance.

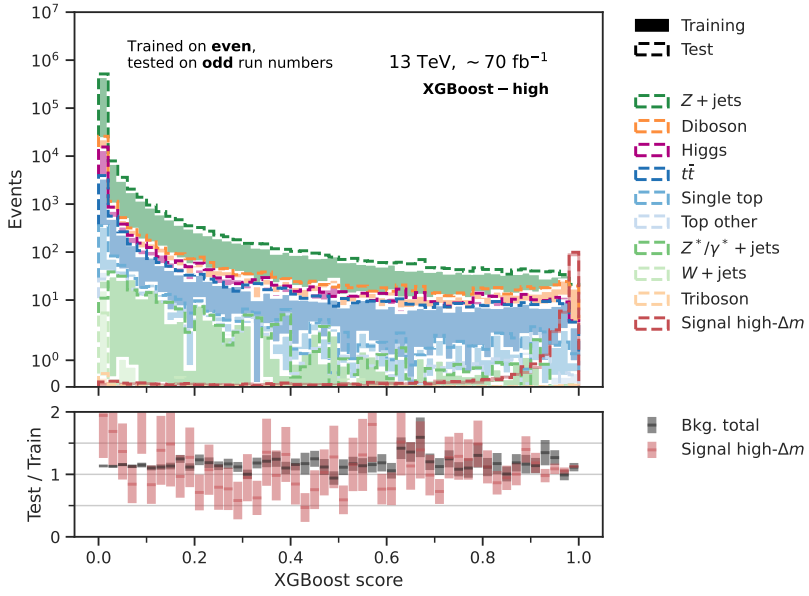
9.7.4 Binary classification performance

9.7.4.1 Background efficiency vs. signal efficiency

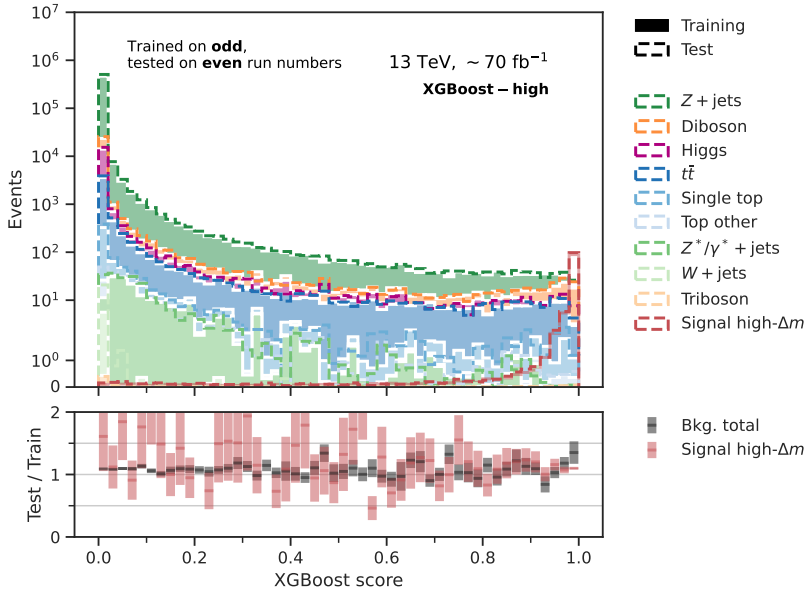
The ROC curves in Figure 9.24 show that one can retain close to 100% background efficiency for almost any choice of signal efficiency, and vice versa. This indicates that a large share of the class distributions are almost entirely positioned at, or very close to, the true class values. The AUC of both the training and test ROC curves amount to 0.998, which is very close to the maximum value of one.

9.7.4.2 Background rejection vs. signal efficiency

In terms of the background rejection, the ROC curves in Figure 9.25 show that only about 1 in 10,000 background events will be mis-classified as signal, given a signal efficiency of 0.5. This is roughly an order of magnitude better than the XGBoost-int models and two orders of magnitude better than the XGBoost-low models.

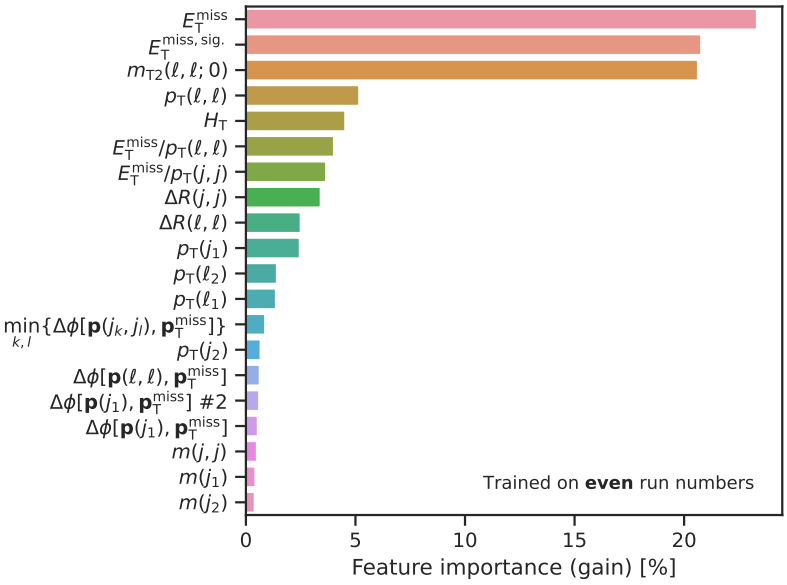


(a) Model trained on even and tested on odd run numbers.

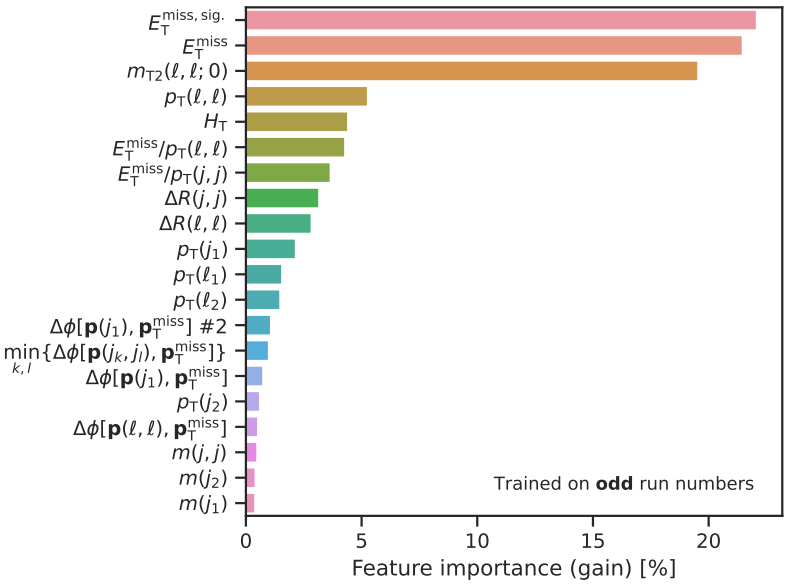


(b) Model trained on odd and tested on even run numbers.

Figure 9.22: XGBoost output scores on the training (filled histograms/lines) and test (dashed lines) sets, for all backgrounds and the high-Δm signal samples. The bottom subplots show the bin-wise test-to-training ratio of the sum of background contributions in black and the sum of high-Δm signal samples in red. The uncertainty bands represent statistical uncertainties only.

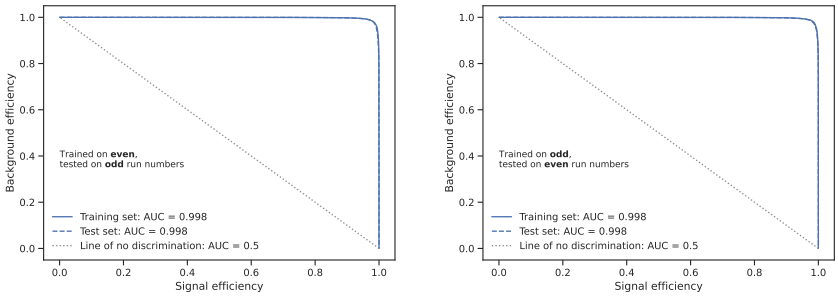


(a) Model trained on even and tested on odd run numbers



(b) Model trained on odd and tested on even run numbers

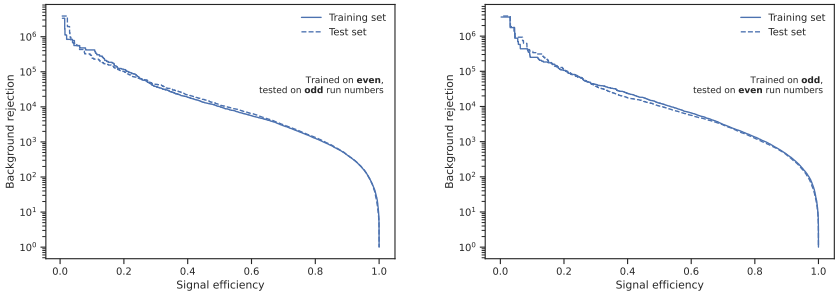
Figure 9.23: Top 20 most important features in the high-Δm models, in terms of the total gain obtained in the full models (all added trees included).



(a) Model trained on even and tested on odd run numbers

(b) Model trained on odd and tested on even run numbers

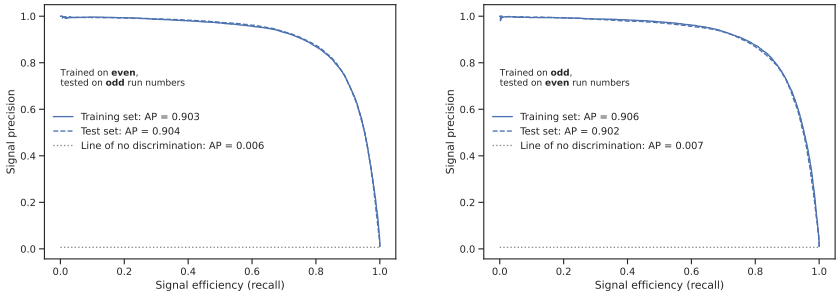
Figure 9.24: ROC curves for models targeting high-Δm signal: background efficiency vs. signal efficiency.



(a) Model trained on even and tested on odd run numbers

(b) Model trained on odd and tested on even run numbers

Figure 9.25: ROC curves for models targeting high-Δm signal: background rejection vs. signal efficiency, where background rejection is equal to $1/(1 -$ background efficiency), and represents the number of correctly rejected background events for every background event that ends up in the signal region (plus one).



(a) Model trained on even and tested on odd run numbers

(b) Model trained on odd and tested on even run numbers

Figure 9.26: Precision vs. recall curves for models targeting high-Δm signal.

9.7.4.3 Signal precision vs. signal efficiency

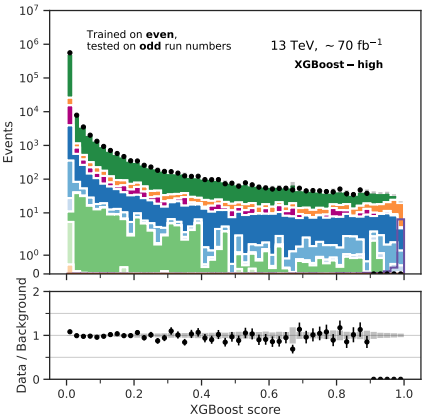
The more representative precision-recall curves of the XGBoost-high models, which directly compare the numbers of events from both classes in the signal region, are shown in Figure 9.26. For a signal efficiency, i.e., recall, of 0.5, the signal precision obtained by the XGBoost-high models is about 0.95, compared with 0.9 of the XGBoost-int and 0.1 of the XGBoost-low models. The average precision of the XGBoost-high models are approximately 0.90 on both the training and test sets, compared with 0.84 of the XGBoost-int and 0.16-0.17 of the XGBoost-low models.

9.7.5 Data vs. background

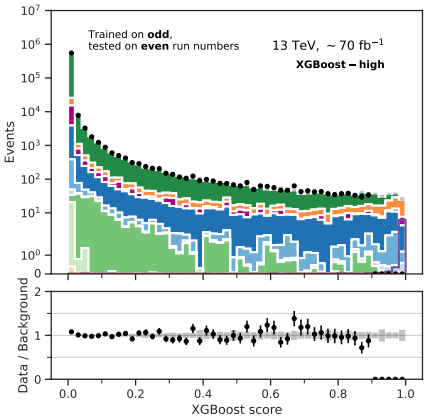
Similarly to the data vs. background comparisons of the XGBoost-low and -int models, the corresponding plots of XGBoost-high in Figure 9.27 show good agreement between data and the sum of background sources. Also in this case, the uncertainties on the total background are statistical only, and therefore underestimates the total uncertainties that would include systematic uncertainties from the simulations, reconstruction and MVA modeling. The experimental data are not shown in the score interval $[0.9, 1]$, where the signal could become non-negligible, if present in the data.

9.7.6 Benchmark signal vs. background

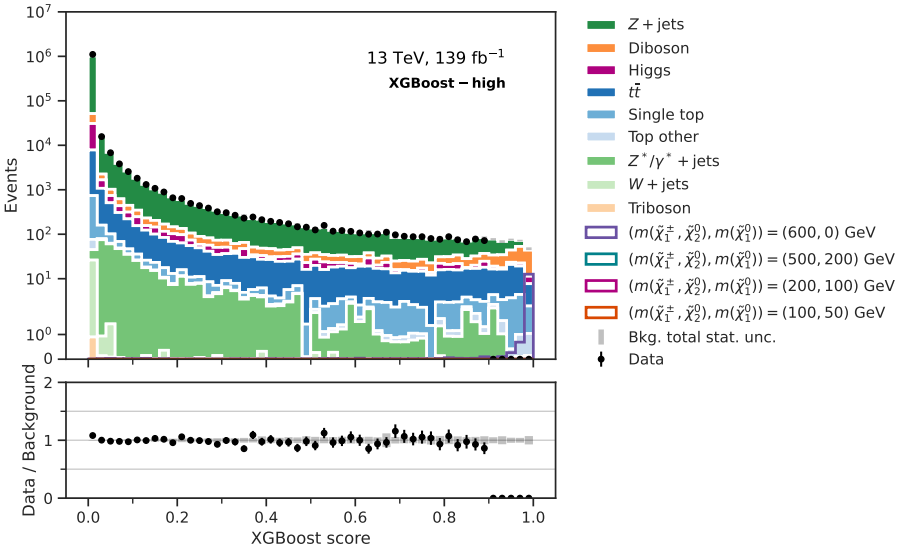
Figure 9.28 shows the score distributions of the background and the high-Δm benchmark signal, $(600, 0)$ GeV, of the two XGBoost-high models in (a) and (b), and the sum of them, which amounts to the integrated luminosity of Run 2, in (c). Both the number of signal events, in the upper panel of (c), and the signal significance, in the lower panel, peak in the rightmost bin with XGBoost scores > 0.995. The background in this bin is from dibosons only. The Z_N significance in the XGBoost-high SR, with XGBoost scores > 0.995, is 3.35, which is greater



(a) Model trained on even and tested on odd run numbers

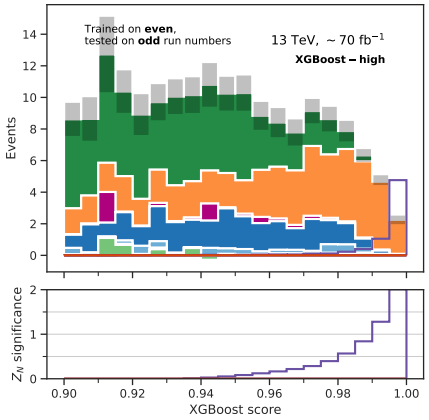


(b) Model trained on odd and tested on even run numbers

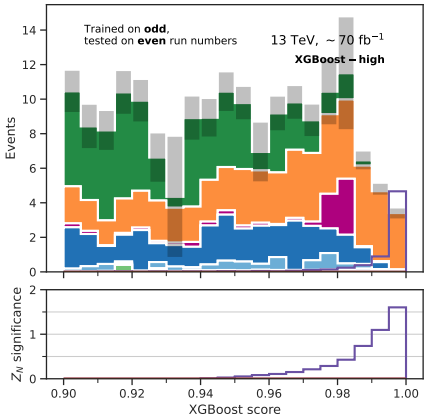


(c) Combination of both models

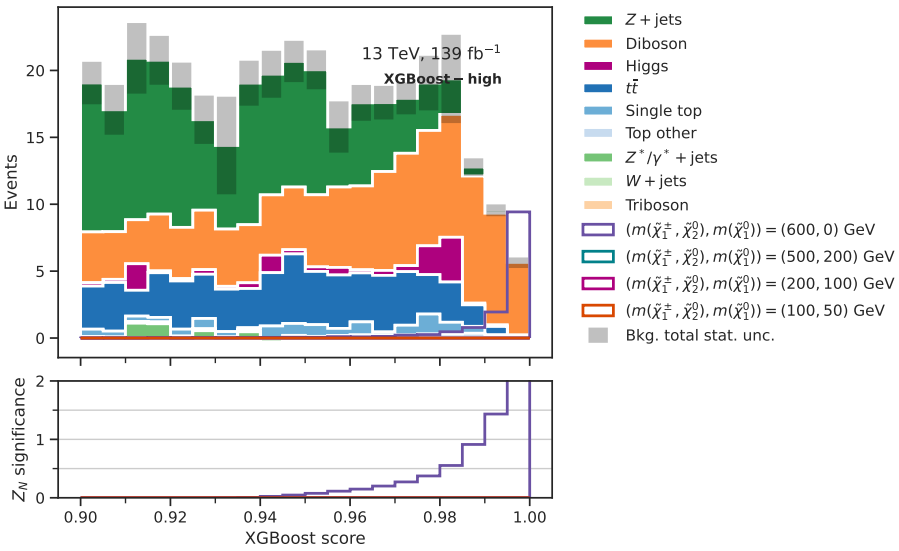
Figure 9.27: Data vs. expected background in the background-dominated XGBoost score region of $[0, 0.9]$, using the XGBoost-high models. The data points in the region with XGBoost score > 0.9 are not shown ("blinded")



(a) Model trained on even and tested on odd run numbers



(b) Model trained on odd and tested on even run numbers



(c) Combination of both models

Figure 9.28: High-Δm benchmark signal sample vs. expected background in the highly signal-like XGBoost score region of $[0.9, 1]$, using the XGBoost-high model.

than 1.64. This means that we expect to be able to exclude the $(600, 0)$ GeV signal at 95% CL, if no excess is observed in the data above the background.

9.8 Expected signal mass exclusions

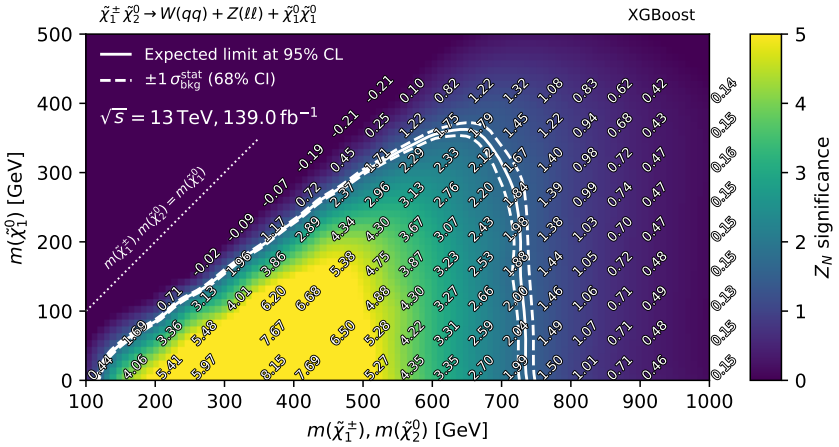
With all the six XGBoost models trained, that is, XGBoost-low, -int and -high, all three with one model trained on events from even run numbers and another model trained on events from odd run numbers, we can feed all the signal samples of the full signal mass-grid through the models that correspond to their respective mass-splittings. The models trained on even run numbers are passed the events from odd run numbers, and vice versa. The outputs from each model are then summed together, and the number of signal and background events that fall into the SRs are collected. With the number of signal and background events in the SRs in place, the corresponding Z_N significance can be calculated, applying a conservative estimate of 30% systematic uncertainty to the calculation.

Figure 9.29 shows the Z_N significance calculated for all samples in the signal mass-grid, using the XGBoost models in (a) and the single-bin discovery regions (DRs) from Table 8.5 in (b) for comparison. Whereas the expected exclusion limits of the conventional analysis, shown in Figure 8.19, make use of shape fits from multi-bin SRs, the exclusion limits of Figure 9.29 only make use of simple cut-and-count analysis on single-bin SRs. I will first have a look at the performance of the single-bin to the multi-bin SRs for the conventional analysis approach, before addressing the performance of the single-bin XGBoost SRs w.r.t. the conventional ones.

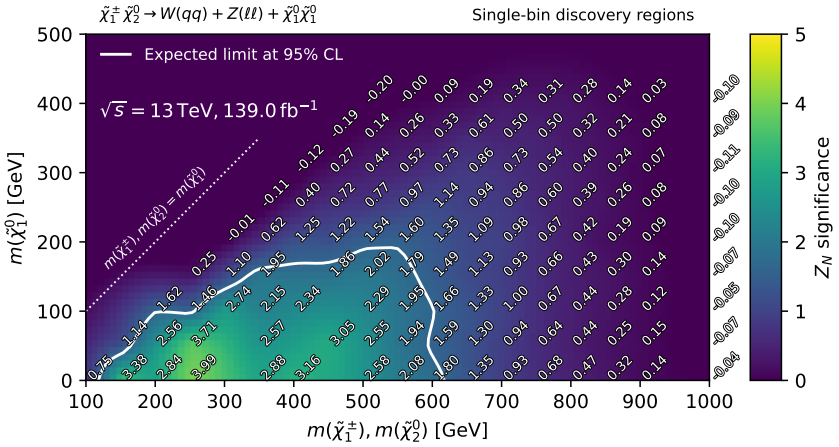
Comparing the multi-bin results of the conventional analysis in Figure 8.19 with the conventional single-bin results (which make no use of the shape of the signal and background distributions in the SRs) in Figure 9.29b, the added sensitivity from multi-bin shape fits are striking. The single-bin DRs obtain expected exclusion limits up to about 600 GeV in $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$-mass, whereas the multi-bin SRs reach all the way up to 750 GeV. When it comes to the expected exclusion reach on the $\tilde{\chi}_1^0$-mass, the single-bin DRs go up to nearly 200 GeV, whereas the multi-bin SRs reach just above 300 GeV. The multi-bin SRs also have better coverage in the more compressed mass-splitting region, where they obtain expected exclusion sensitivity to the $\Delta m = 100$ GeV points of $(100, 0)$, $(150, 50)$ and $(200, 100)$ GeV, whereas the single-bin DRs do not reach expected exclusion at all for these mass-points.

On the other hand, comparing the single-bin DRs, with 139.0 fb^{-1} of integrated luminosity, to previously published results which also made use of single-bin SRs, the expected exclusion limits continue to improve with more data. The single-bin SR-Zjets, shown in Figure 7(a) of Reference [213], with 20.3 fb^{-1} of 7 TeV data from LHC Run 1, obtained expected exclusion limits up to $m_{\tilde{\chi}_1^{\pm},\tilde{\chi}_2^0} = 350$ GeV and $m_{\tilde{\chi}_1^0} = 80$ GeV, and excluded only one signal point with $\Delta m = 150$ GeV, none below. The single-bin SRs used to obtain the expected exclusion limits of Figure 7.6b, with 36.1 fb^{-1} of 13 TeV data, reached up to $m_{\tilde{\chi}_1^{\pm},\tilde{\chi}_2^0} = 530$ GeV and $m_{\tilde{\chi}_1^0} = 180$ GeV. The single-bin DRs do not improve the expected exclusion reach on the $\tilde{\chi}_1^0$-mass, but extend the reach on the $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$-mass from 530 to 600 GeV.

To be fair, the DRs themselves have not been optimized for exclusion potential



(a) XGBoost-low, -int and -high



(b) DR-low, -int and -high

Figure 9.29: Z_N significance, for each of the signal mass-points in the signal grid, are given in numbers at the location of the corresponding mass-points. Interpolation of Z_N between the grid points has been carried out to obtain contours at $Z_N = 1.64$, which correspond to the exclusion limit at 95% CL. All mass-points within these contours are expected to be excluded at 95% CL, if no excess is observed in the data above the background. The results from single-bin XGBoost SRs are shown in (a), and compared to corresponding results from single-bin DRs (discovery regions) in (b). The DRs are defined in Table 8.5. In (a), there are also $\pm 1\,\sigma_{\text{bkg}}^{\text{stat}}$ error bands, which correspond to the exclusion contours when the number of background events in the SRs have their statistical uncertainty added (inner dashed line) or subtracted (outer dashed line). (Uncertainty bands were not calculated for the DR exclusion limits.)

on the signal model considered in this chapter, in the same way as the previous single-bin SRs and the multi-bin SRs for the full Run 2 data have been. They are rather combinations of the multi-bin SRs, to be used for model-independent signal interpretations. In this comparison, they are only taken as rough indications of how conventional, single-bin, cut-and-count SRs may compare to multi-bin and MVA SRs. They may be slightly conservative in terms of expected exclusion reach for optimized single-bin SRs.

Now, we compare the performance of the single-bin ML analysis to the single- and multi-bin conventional analyses. When it comes to the XGBoost models, the expected exclusion potential seems to be a lot stronger than the single-bin DRs, but comparable to the multi-bin SRs. The $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$-mass reach is slightly lower for the XGBoost models, of roughly 725 GeV, compared to 750 GeV of the multi-bin SRs. The $\tilde{\chi}_1^0$-mass reach is, however, slightly better with the XGBoost models, of roughly 350 GeV, compared to the multi-bin SRs with just above 300 GeV. In the more compressed mass-splitting region around $\Delta m = 100$ GeV, the XGBoost-low models only obtain expected exclusion for the $(150, 50)$ mass-point, whereas the multi-bin SRs also manage to get exclusion sensitivity for the $(100, 0)$ and $(200, 100)$ GeV mass-points as well.

Given that the XGBoost output distributions for signal and background events have different shapes for all the models, in particular in the highly signal-like region (XGBoost scores > 0.9), there may be more signal sensitivity to gain by performing multi-bin shape-fits on the output distributions, instead of the simple single-bin cut-and-count SRs considered in this work. As the single-bin XGBoost SRs show exclusion reach comparable to the multi-bin conventional analysis SRs, multi-bin shape-fits on the XGBoost output could potentially become even more powerful than the multi-bin conventional analysis SRs. This could be a promising topic for future studies.

9.9 Summary

A sensitivity study has been carried out using XGBoost MVA models, performing binary classification of signal vs. background, to target $\tilde{\chi}_1^{\pm}$ $\tilde{\chi}_2^0$-production in final states with two leptons, two jets and $E_{\mathrm{T}}^{\mathrm{miss}}$. Six XGBoost models have been trained on MC simulated signal and background samples. The two XGBoost-low models are trained only on the signal samples in the low-Δm category, where one model is trained on events with even run numbers only and the other on events with odd run numbers only. The XGBoost-int and -high models are trained correspondingly on signal samples in the int- and high-Δm categories, respectively. On final evaluation (test), the signal samples are passed through the XGBoost model trained on the relevant mass-splitting category, where signal and bakground events from even run numbers are passed through the model trained on odd run numbers, and events from odd run numbers are passed through the model trained on even run numbers. Finally, the XGBoost score distributions from even and odd run numbers are combined (summed) for each of the XGBoost-low, -int and -high models.

The expected exclusion contour at 95% CL in the signal mass-grid shows that single-bin XGBoost SRs obtain significantly higher signal sensitivity compared to the single-bin DRs of Chapter 8, but are comparable to the performance of the multi-bin SRs of Chapter 8. This suggests that performing multi-bin shape-fits on the XGBoost output distributions could potentially expand the exclusion reach beyond that of the multi-bin SRs of the conventional (non-MVA) analysis.

Chapter 10

Conclusions

With the proton-proton collision data of the LHC Run 2, we have covered new ground in the search for the existence of supersymmetric particles. The part of the full Run 2 dataset which is deemed "good for physics" amounts to $139\,\text{fb}^{-1}$, collected at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. This is more than 5x the integrated luminosity of the full Run 1 dataset, which consists of $4.7\,\text{fb}^{-1}$ taken at $\sqrt{s} = 7\,\text{TeV}$ and $20\,\text{fb}^{-1}$ taken at $\sqrt{s} = 8\,\text{TeV}$.

On the one hand, the nearly doubling of collision energy from Run 1 to Run 2 increases the potential mass reach for heavier particles. On the other hand, the larger dataset increases the statistical significance of the physics analyses, and thereby increases the sensitivity to rare events.

The first thesis search involved checking the ATLAS sensitivity to pair-production of the lightest chargino ($\tilde{\chi}_1^\pm\ \tilde{\chi}_1^\mp$), with decays via sleptons ($\tilde{\ell}$) or sneutrinos ($\tilde{\nu}$) into final-states with two leptons (e or μ) and E_T^miss from two neutrinos (ν) and two lightest neutralinos ($\tilde{\chi}_1^0$). My findings showed that we expected to go beyond the Run 1 sensitivity of ATLAS to the same model already with $6\,\text{fb}^{-1}$ of integrated luminosity in Run 2. The search for this model was therefore worth targeting for publication with early Run 2 data. The ATLAS publication [158] with $13.3\,\text{fb}^{-1}$ increased the exclusion limit on the $\tilde{\chi}_1^\pm$ mass up to 620 GeV (for a massless $\tilde{\chi}_1^0$), compared to 480 GeV with the Run 1 data. The exclusion reach for the $\tilde{\chi}_1^0$ mass increased from 180 GeV in Run 1 up to 260 GeV with the early data of Run 2.

For the remainder of thesis work, I moved on to another simplified model, namely pair-production of the lightest chargino along with the next-to-lightest neutralino ($\tilde{\chi}_1^\pm\ \tilde{\chi}_2^0$), with decays via W and Z bosons into final-states with two leptons (e or μ), two light-flavor jets (not b-tagged) and E_T^miss from two lightest neutralinos ($\tilde{\chi}_1^0$). The first search with this model involved defining and optimizing signal regions for publication with the two first years of data-taking in Run 2 (2015 and 2016). This amounted to $36.1\,\text{fb}^{-1}$ of integrated luminosity and was the first analysis of this model with Run 2 data. The ATLAS publication [165, 166] increased the exclusion limit on the $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ mass from 420 GeV with Run 1 data to 580 GeV (for a massless $\tilde{\chi}_1^0$) with the 2015+2016 dataset of Run 2. The exclusion reach for the $\tilde{\chi}_1^0$ mass increased from 150 GeV to 230 GeV w.r.t. the Run 1 limits.

My third search was a continuation of the previous project, but this time using the full Run 2 dataset (data-taking from 2015 up to and including 2018), which amounted to $139\,\text{fb}^{-1}$. Much of my contributions were in the form of preparations for the final analysis, in terms of requesting and producing simulated samples, developing analysis code, chairing working group meetings and giving analysis updates in the ATLAS SUSY sub-groups working on electroweak and

strong production of sparticles. On the analysis side, I continued working on defining and optimizing signal regions, with accompanying control and validation regions for background modeling. At the time of writing, the complete ATLAS publication has been submitted to a journal for peer-review and has been published as a pre-print [193]. In addition to the increase in statistics w.r.t. to the previous publication, from 36.1 fb^{-1} to 139 fb^{-1}, the statistical modeling also developed from using single-bin to multi-bin signal regions, which make use of the shape of the background and signal vs. the data distributions. The expected mass exclusion reach thereby increased from 580 to 750 GeV for $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ and from 230 to 320 GeV for $\tilde{\chi}_1^0$. The observed limits ended up being stronger, due to observing fewer events than expected in several of the signal regions. $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ masses up to 820 GeV and $\tilde{\chi}_1^0$ masses below 380 GeV ended up being excluded.

Finally, I present a sensitivity study using the machine learning algorithm XGBoost, where I make use of gradient boosted decision trees for binary classification of background- and signal-like events, using the integrated luminosity corresponding to the full Run 2 dataset. By making a single cut on each of the XGBoost distributions, the sensitivity to the signal increases significantly compared to conventional single-bin signal/discovery regions, from 600 to 700 GeV for $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ and from 150 to 350 GeV for $\tilde{\chi}_1^0$. The single-bin XGBoost regions achieve comparable sensitivity to the conventional multi-bin signal regions, which suggests that there may be even more sensitivity to gain by considering multi-bin signal regions also for the XGBoost models.

There are exciting times ahead. The LHC Run 3 is expected to produce the first *pp* collisions for physics analysis, at the (yet again) record-breaking energy of 13.6 TeV, on July 5, 2022, just a couple of weeks after submitting this thesis. It is planned to be operational until 2026 and produce a new dataset of similar size to that of Run 2. Run 3 will also be a stepping stone toward the high-luminosity LHC (HL-LHC) [214], which will become the LHC Run 4, and is currently expected to be operational from 2029 through 2032 and to produce a significant increase in the instantaneous luminosity. This will pose great technological challenges, such as increased radiation exposure for the detectors and a dramatic increase in demand of data throughput and processing capacity. In return, this will also significantly increase the event rates, which means more statistics faster. The HL-LHC is expected to operate at $\sqrt{s} = 13.6 - 14$ TeV and produce more than 10x more data than the LHC Run 1 and Run 2 combined, namely $3000{-}4000\,\text{fb}^{-1}$.

Even though some people have started to lose faith in the realization of Supersymmetry in nature, the more than 100-dimensional parameter space is so vast that it is practically impossible to reject it with any kind of certainty. However, what we do know with certainty, is that there are phenomena in nature we can't describe with our current knowledge, and that there are in fact more to be discovered. Will it be dark matter, or maybe something we have not even thought about yet?

Appendices

Appendix A

Authorship qualification

As a member of the ATLAS Collaboration, one has to complete a so-called qualification task in order to qualify as a signing author of ATLAS publications. The idea is that an ATLAS member should have made a significant contribution to the operation of the experiment before being allowed to sign the scientific results published by the collaboration. For the qualification work to be approved, the qualifying member should have spent at least 80 full working days on the task, in principle within the scope of one year.

Given that the high-energy physics group at the University of Oslo is invested in the software and computing operations in ATLAS, and thereby has expertise on the topic, it was natural for me to choose a task related to software and computing.

A.1 Task description

The qualification task assigned to me was *to work toward a reproducible Monte Carlo (MC) simulation setup in heterogeneous environments on the Worldwide LHC Computing Grid*. The task description was formulated as follows:

> The reproducibility of various Monte-Carlo production steps is critical for production in heterogeneous distributed environment. The divergence of results comes either from software sensitivity to cpu architecture or from different execution environments and conditions on the grid. The primary task is to extend the technical validation platform to cover different architectures and to develop a capability to execute the payload in targeted distributed environment in a reproducible way. The job execution should set the initial conditions for various production steps to be the same on per-event basis regardless of the execution platform.

The description refers to the fact that MC simulations produced by ATLAS are run on different computing sites (servers) around the world, on different CPU types, running on different (versions of) operating systems, etc., which may introduce differences in the simulation outputs depending on where they are produced.

Simulation outputs from two or more setups and/or grid sites were compared on two levels: (i) bit-wise comparisons, i.e., looking for identical results, which would be the optimal outcome (fully reproducible), and (ii) statistical comparisons, which means evaluating the goodness-of-fit between histograms of simulated physics variables, i.e., whether the simulations reproduce the same physical distributions (satisfyingly reproducible physics-wise).

My technical supervisor on the task, working at CERN, was Jose Enrique Garcia Navarro (Instituto de Fisica Corpuscular (IFIC), Centro Mixto Universidad de Valencia - CSIC), and my local supervisor, working in Oslo, was Eirik Gramstad (University of Oslo).

A.2 Simulated samples

The physics process considered for the simulations was $t\bar{t}$-production, which is a common process in pp-collisions at the LHC. It captures a wide range of phenomena in the subsequent decays of the top and anti-top quarks, involving leptons, missing energy from neutrinos and hadronic jets, both in terms of light-flavor jets and b-jets, where the latter give rise to secondary, displaced vertices.

The complete process of producing the MC simulation samples can be broken down into four separate steps:

1. **Event generation:** An event generator picks a random tree-level process (and certain higher-order corrections, depending on the setup) per event, based on the scattering processes available in the setup and the relative probabilities (scattering amplitudes) of these processes. So-called pile-up events are simulated separately and overlaid the collision events.

2. **Simulation:** The particles from the hard-scatter processes of step 1. are now evolved/simulated from the collision point to the point where they reach the detectors. This involves hadronization and parton-showering of the final-state quarks and gluons as well as particle interactions ("hits") with the detector material.

3. **Digitization and reconstruction:** In this step, the detector interactions/hits from the simulation step are converted to digital signals. Then, the event reconstruction algorithms are run on the detector readouts to produce reconstructed particles/objects for physics analyses.

4. **Merging:** Typically, step 2. and 3. are split into multiple jobs that run in parallel, handling different subsets of the generated events from step 1. In step 4., the results from each of the jobs in step 3. are merged into one large dataset.

For the qualification task, we made use of 100.000 pre-generated $t\bar{t}$-events, which were common to all comparisons. Step 1. is therefore the same in all comparisons, and is therefore effectively not part of the evaluation. Since step 4. only combines the outputs from step 3., the merge step does not include any form of simulation and therefore does not introduce changes to the data from step 3. The comparisons therefore only evaluate the effects of the simulation (step 2.) and the digitazation and reconstruction (step 3.) Only the effects of one of these steps are varied at a time, where the other is treated identically (kept constant) between the results being compared.

A.3 Bit-wise reproducibility

We start by investigating whether we can obtain a fully reproducible simulation setup on two different Grid sites, i.e., identical simulation outputs, event by event. This was done by running a Python script called `acmd.py`, which is part of the ATLAS software. Commands on the following form were run:

```
acmd.py diff-root --error-mode resilient <file1> <file2>
```

which runs a "diff" operation comparing two Root files, `<file1>` and `<file2>`. An additional command-line option `--entries <range>` was also applied in order to run the comparisons on smaller chunks of events in the two files.

A.3.1 Toward a reproducible setup

Mainly three simulation settings were applied to facilitate reproducibility on different Grid sites:

- *Use the same math library on all CPU types.* The Intel math library was already set up to be loaded by default in the simulation transform.
- *Use a fixed seed to the generator of random numbers.* The random seed offset was disabled in the simulation transform.
- *Set reproducible ordering of events in the merging of outputs from multiprocessing jobs.* The Round Robin strategy was used in AthenaMP [215].

A.3.2 Simulation

The outputs of the simulation step, so-called HITS files, were produced at two different Grid sites running Intel CPUs, namely Brookhaven Naitonal Laboratory (BNL) and Rutherford Appleton Laboratory (RAL), from identical inputs.

The data stored in the simulation output files from the two sites only differed in the information about the CPU time spent on the tasks. The simulated physics variables were identically reproduced from both sites.

Thereby, the so-called simulation transform was found to be bit-wise compatible when run at two different Intel sites.

A.3.3 Digitization

The two sets of HITS files from BNL and RAL in the previous step were then passed on to the digitization step, which was done at BNL. The outputs from the digitization process are called Raw Data Output (RDO) files.

The digitization of the HITS files from BNL and RAL were also found to be identical, except for differences in the CPU time spent on the tasks.

Thereby, the digitization of identical HITS files were found to be bit-wise compatible when run at the same Grid site.

A.3.4 Reconstruction

After digitization of the simulation outputs, the RDO files are passed on to the reconstruction step. Also the reconstruction of the simulations from BNL and RAL are run at the same site, namely at BNL.

The reconstruction process consists of two substeps: first producing a set of Event Summary Data (ESD) files with detailed outputs of the reconstruction, from which higher-level analysis objects are constructed and stored to Analysis Object Data (AOD) files.

Even though the RDO files from the digitization step had identical physics values in the two datasets, a few differences arose in the reconstructed data in the ESD files. The differences were found in the following physics variables:

- `MET_EMTopoRegions`: `mpx`, `mpy`, `name`, `sumet`
- `MET_LocHadTopoRegions`: `mpx`, `mpy`, `name`, `sumet`
- `Muons`: `rpcHitTime`
- `MuonSegments`: `clusterTime`

where the two first bullet points represent missing E_{T} (MET) objects calculated using measurements from the electromagnetic (EM) and hadronic (LocHad) calorimeters, respectively; with `mpx`, `mpy`, `name`, `sumet` representing the x- and y-components, name and total sum of MET, respectively.

A comparison of print-outs of content in the two AOD files are shown in Figure A.1. The diff view highlights the differences between the two files. One can see an event with no $E_{\mathrm{T}}^{\mathrm{miss}}$ in the file to the left, but with $E_{\mathrm{T}}^{\mathrm{miss}}$ contributions in the file to the right.

After the ESD files had been processed into AOD files, the same differences that were found between the ESDs were found again in the AODs, in addition to differences in yet another variable:

- `HLT_xAOD__TrackParticleContainer_InDetTrigTrackingxAODCnv_Electron_L2IDAux`: `numberOfSCTSharedHits`

which represents the number of shared hits in the semiconductor tracker (SCT) of the electron high-level trigger (HLT) object.

Figure A.2 shows a summary of the differences found by the Python script comparing the two AOD files. The number of different values (leaves) per variable are also given.

This means that the reconstruction step is not completely reproducible when run on the same Grid site on the same input. However, the number of differences introduced are fairly small, as only five out of the 2000 (0.25%) events included in the comparison contain differences in the AOD files, where only a small fraction of the physics variables of these five events are different.

The findings of this study were reported to the relevant people in ATLAS for further follow-up.

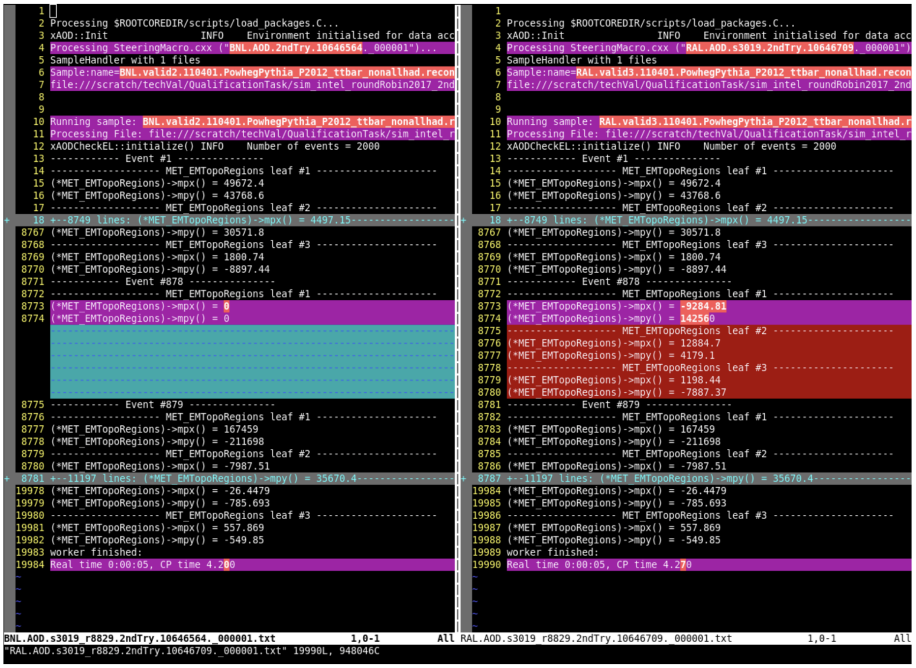


Figure A.1: Comparison of print-out logs from selected physics variables in the AODs, where differences are highlighted in red.

```
Py:diff-root        INFO Found [392836552] identical leaves
Py:diff-root        INFO Found [41055] different leaves
Py:diff-root        INFO  [ESDtoAOD_mems.timings]: 7677 leaves differ
Py:diff-root        INFO  [ESDtoAOD_timings.timings]: 7977 leaves differ
Py:diff-root        INFO  [EVNTtoHITS_timings.timings]: 4016 leaves differ
Py:diff-root        INFO  [HITStoRDO_timings.timings]: 4008 leaves differ
Py:diff-root        INFO  [HLT_xAOD__TrackParticleContainer_InDetTrigTrackingxAODCnv_Electron_L2IDAux.numberOfSCTSharedHits]: 3 leaves differ
Py:diff-root        INFO  [MET_EMTopoRegionsAux]: 2 leaves differ
Py:diff-root        INFO  [MET_EMTopoRegionsAux.mpx]: 3 leaves differ
Py:diff-root        INFO  [MET_EMTopoRegionsAux.mpy]: 3 leaves differ
Py:diff-root        INFO  [MET_EMTopoRegionsAux.name]: 3 leaves differ
Py:diff-root        INFO  [MET_EMTopoRegionsAux.sumet]: 3 leaves differ
Py:diff-root        INFO  [MET_LocHadTopoRegionsAux]: 4 leaves differ
Py:diff-root        INFO  [MET_LocHadTopoRegionsAux.mpx]: 6 leaves differ
Py:diff-root        INFO  [MET_LocHadTopoRegionsAux.mpy]: 6 leaves differ
Py:diff-root        INFO  [MET_LocHadTopoRegionsAux.name]: 6 leaves differ
Py:diff-root        INFO  [MET_LocHadTopoRegionsAux.sumet]: 6 leaves differ
Py:diff-root        INFO  [MuonSegmentsAuxDyn.clusterTime]: 3 leaves differ
Py:diff-root        INFO  [MuonsAuxDyn.rpcHitTime]: 4 leaves differ
Py:diff-root        INFO  [RAWtoESD_mems.timings]: 9327 leaves differ
Py:diff-root        INFO  [RAWtoESD_timings.timings]: 8004 leaves differ
Py:diff-root        INFO NOTE: there were errors during the dump
Py:diff-root        INFO fold.allgood: False
Py:diff-root        INFO fnew.allgood: False
Py:diff-root        INFO files differ!
```

Affected physics variables

Figure A.2: Summary of the diff after running the Python script on the two AOD files. The physics variables are framed in, whereas the remaining differences come from information about the run times.

A.4 Physics validation framework

For the statistical comparisons of distributions of physics variables, we made use of an existing framework for performing physics validation. The framework takes a specific ntuple-format as input, which contains histograms of a wide range of physical characteristics of the reconstructed physics objects, 5170 histograms (variables) in total.

The validation framework uses one dataset as reference and performs goodness-of-fit tests between the reference and each of the other datasets to be compared, one test for each of the histograms in the ntuples. The chi-square test is used to measure the goodness-of-fit between the histograms.

A.4.1 Chi-square goodness-of-fit

The chi-square (χ^2) value is calculated by summing the squared difference between the comparison and the reference, divided by the variance[1] of the reference, bin by bin in each histogram:

$$\chi^2 = \sum_{i=1}^{k} \frac{(\text{comparison}_i - \text{reference}_i)^2}{\text{reference}_i}, \tag{A.1}$$

where k denotes the number of bins in the histogram and i is an index running over the bins, $i \in \{1, 2, ..., k\}$.

A.4.2 The χ^2-distribution

The chi-square distribution is illustrated for various numbers of degrees of freedom (NDF) in Figure A.3. For a χ^2 goodness-of-fit, the NDF is equal to the number of bins in the given histogram. One can show that the expectation value for a χ^2-distribution is equal to the NDF of that distribution

$$\langle \chi^2 \rangle = \text{NDF}, \tag{A.2}$$

with variance

$$\sigma^2_{\chi^2} = 2 \times \text{NDF}, \tag{A.3}$$

i.e., each term in the sum of Equation (A.1) on average contributes a value of one, given that the comparison dataset is consistent with the reference dataset. What we mean by this is that the comparison histogram has only Gaussian distributed deviations away from the bin-values of the reference histogram, with variances equal to the variance of the bins of the reference histogram.

[1] The number of entries in each bin follows a Poisson distribution, where the standard deviation is equal to the square-root of the number of entries in the bin, i.e., for the reference histogram the standard deviation in bin i is $\sigma_{\text{Poisson}} = \sqrt{\text{reference}_i}$. Since the variance is just the standard deviation squared, the variance of the number of entries in bin i of the reference histogram is $\sigma^2_{\text{Poisson}} = \sqrt{\text{reference}_i}^2 = \text{reference}_i$.

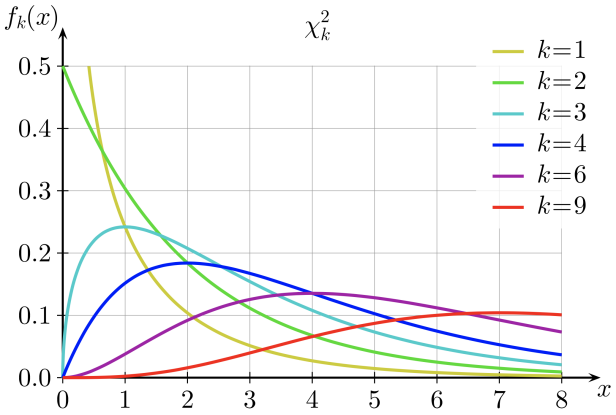


Figure A.3: Probability density functions (PDFs), $f_k(x)$, of chi-square values, χ^2_k (denoted x in the plot), for various numbers of degrees of freedom, k [216].

A.4.3 χ^2 per number of degrees of freedom

Since the histograms of the physics variables in the ntuples have varying numbers of bins, the sizes of the χ^2-values are not directly comparable. In order to collect and compare the χ^2-values of all the physics variables in a single histogram, we divide the χ^2-values by their corresponding NDF. Since the expectation value of χ^2 is equal to NDF, we thereby have that

$$\left\langle \frac{\chi^2}{\text{NDF}} \right\rangle = 1, \tag{A.4}$$

with variance

$$\sigma^2_{\chi^2/\text{NDF}} = 2 \tag{A.5}$$

and standard deviation

$$\sigma_{\chi^2/\text{NDF}} = \sqrt{2} \approx 1.41. \tag{A.6}$$

A.4.4 Three regimes

When it comes to comparing two simulated datasets, there are effectively three distinct χ^2/NDF-regimes based on the nature of the deviations:

- **Identical:** If the two simulations are in fact identical, we expect

$$\left\langle \frac{\chi^2}{\text{NDF}} \right\rangle = 0. \tag{A.7}$$

- **Random errors:** If the two simulations are inherently equivalent physics-wise, except for random errors such as slightly different methods for calculating the same quantity, we expect the results to be χ^2/NDF-distributed with

$$\left\langle \frac{\chi^2}{\text{NDF}} \right\rangle = 1. \tag{A.8}$$

- **Inherently different:** The third and final option is that the two simulations are inherently different, i.e., that the deviations are not only random fluctuations about the same expectation value, but rather different treatments which result in physics results that are not consistent. We then expect

$$\left\langle \frac{\chi^2}{\text{NDF}} \right\rangle \gg 1. \tag{A.9}$$

A.5 Statistical reproducibility

A.5.1 Comparisons of non-equivalent simulations

In order to get a handle on what comparisons of non-equivalent simulations may look like in terms of χ^2/NDF, I produced two pairs of simulations where the reference and comparison datasets were deliberately produced with different conditions. Figure A.4 shows comparisons of two simulations produced with different software versions and different pile-up conditions (black line), and another comparison of two simulations produced either with or entirely without the addition of pile-up contributions to the hard-scatter events.

The black histogram clearly shows examples of all the three regimes mentioned above. The far left bin, with $\chi^2/\text{NDF} < 0.05$, holds the results of comparisons between histograms which are identical, or very close to identical, in the datasets being compared. There is also a broad peak around $\chi^2/\text{NDF} = 1$, which most likely stem from distributions that are physics-wise consistent, but has minor adjustments in the implementation between software versions. Finally, the far right bin, with $\chi^2/\text{NDF} > 1.95$, holds the results of comparisons between distributions which are significantly different, with deviations much greater than one standard deviation in the reference dataset.

Similar observations can be made for the blue histogram as well, in particular when it comes to examples of both identical and significantly different distributions in the two datasets being compared. There is, however, not the same clear peak around $\chi^2/\text{NDF} = 1$ as was the case for the black histogram. This may indicate that there are less random fluctuations about the same expectation values in the two datasets, but rather more systematic deviations. The addition of pile-up contributions to the events should significantly affect the physics variables and measurements, and it is not surprising that this shifts the background levels in a systematic way.

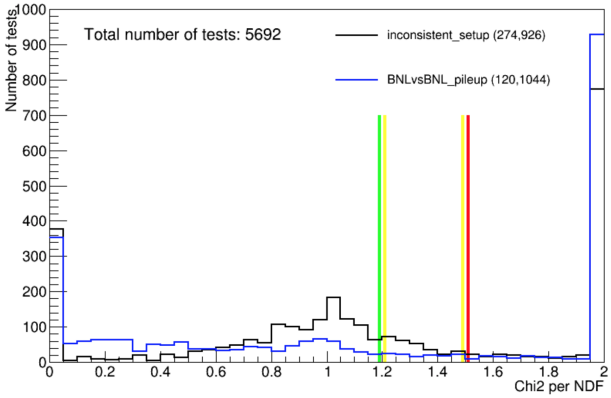


Figure A.4: Examples of comparisons between datasets produced with different conditions. The black line shows the histogram of χ^2/NDF-values for the comparisons between two simulations produced with different software versions (Athena release 19 vs. 21) and different pile-up conditions (mc15c vs. mc16). The blue line shows the results of comparisons between two simulations produced at the same computing site (Brookhaven National Lab, BNL), but where the referance dataset has no pile-up (minimum bias events) overlaid the hard-scatter events and the comparison dataset does have pile-up. The parentheses in the legends give two numbers: the first represents the number of values/tests which fall in the "yellow" region with $\chi^2/\text{NDF} \in [1.2, 1.5)$, and the second number represents the number of values/tests which fall in the "red" region with $\chi^2/\text{NDF} > 1.5$. These classifications come as part of the general physics validation framework, and are not considered explicitly in this work, as we focus more on the distributions of the 40 bins of the histogram rather than the number of occurrences of the three classes.

A.5.2 Intel vs. AMD math libraries

Most Grid sites run only on central processing units (CPUs) manufatured by Intel, but there are also a few sites that make use of AMD CPUs. Both the Intel and AMD CPUs use the x86 instruction set architecture. However, the two companies have developed their own math libraries, the Intel Math Kernel Library (MKL) and the AMD Math Library (LibM), respectively, which are basic math routines optimized for their own CPUs. Since both the Intel and AMD CPUs use the same instruction sets, their math libraries can be used interchangeably, possibly at the cost of somewhat reduced performance.

In order to check if the choice of math library could affect reproducibility on the Grid, a comparison was made between one simulation produced using the Intel math library and another using the AMD math library. The simulation tasks were run on the SiGNET (Slovenian Grid NETwork) site, which is one of the Grid sites with AMD CPUs.

The results of the comparison are shown in Figure A.5. Most of the tests

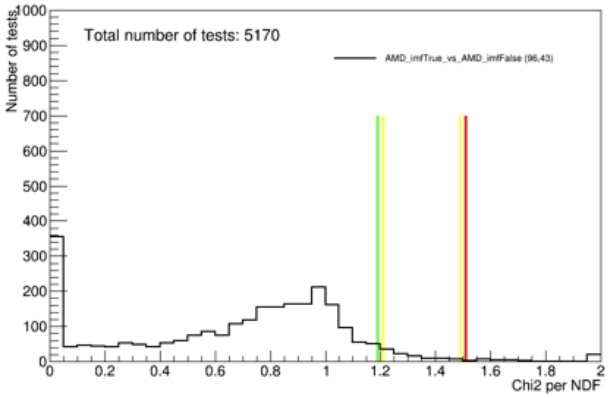


Figure A.5: Effects of using Intel vs. AMD math libraries on comparison of physics variable histogram.

that are not (close to) identical, i.e., in the far left bin, have χ^2/NDF-values around 1, which indicates differences of random nature. Therefore, the Intel and AMD math libraries produce results that are either identical or statistically similar, and should thereby give compatible physics results, but they are not all identical.

A.5.3 Intel site 1 vs. Intel site 2

Next, we compare the results of simulation (step 2.) run at two Grid sites with Intel CPUs: Brookhaven National Laboratory (BNL) in the US and Rutherford Appleton Laboratory (RAL) in the UK. The digitization and reconstruction (step 3.) is run at BNL in both cases.

A comparison between the two sites, with simulations of the hard-scatter events only, is plotted in the black histogram of Figure A.6. The vast majority of variable comparisons produce χ^2/NDF-values less than 1, with the 0-0.05 bin being the most populous. The simulations produced at two different Grid sites using the same CPUs therefore result in physics variables with either identical or statistically compatible distributions.

The blue histogram of Figure A.6 shows comparisons between simulations run at the same two Grid sites, but here with the addition of pile-up interactions in the hard-scatter events. The addition of pile-up shifts the distribution of χ^2/NDF to somewhat larger values, with more weight around $\chi^2/\text{NDF} \approx 1$. This indicates more randomness in the variable distributions, but without becoming incompatible physics-wise. This is not surprising, as the pile-up interactions, which are overlaid the hard-scatter events, are independent of the hard-scatter interactions and are drawn randomly from a pool of pre-generated minimum-bias events.

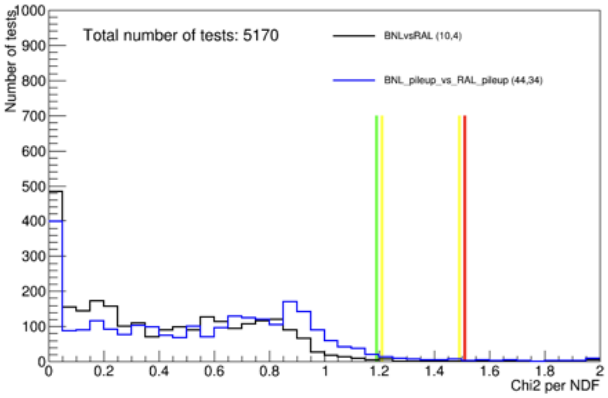


Figure A.6: Comparison of simulations run at two Grid sites with Intel CPUs (BNL and RAL), without the addition of pile-up in blue and with the addition of pile-up in black.

A.5.4 Intel site vs. Intel+AMD site vs. AMD site

After having compared the output of the simulation step (step 2.) at two different sites running the same CPU type, the next sets of comparisons are between sites running different types of CPUs. We compare the output from the following sites: (i) BNL (reference) and RAL (comparison) with Intel CPUs only, (ii) Instituto Nazionale di Fisica Nucleare (INFN), in Italy, with both Intel and AMD CPUs (mixed CPU site), and (iii) SiGNET with AMD CPUs only. The digitization and reconstruction (step 3.) is run at BNL for all simulations.

The comparisons are shown in Figure A.7. The black histogram compares the output from the two Intel sites, BNL and RAL, the blue histogram compares an Intel site (BNL) with the mixed Intel and AMD site (INFN), whereas the pink histogram compares an Intel site (BNL) with the AMD site (SiGNET).

The Intel vs. Intel comparison (black) shows results which are identical or with differences which are smaller than the standard deviation of the reference sample (BNL). The Intel vs mixed Intel+AMD comparison (blue) is somewhat similar in distribution to the pile-up comparisons between Intel sites in Figure A.6 (blue histogram). In a similar manner to how pile-up seems to add physics-wise compatible randomness to the simulations, the same seems to apply when using AMD instead of Intel CPUs. The tendency of shifting the χ^2/NDF-distribution toward the value of 1 is even more prominent in the Intel vs. AMD only comparison (pink), which supports the hypothesis that AMD CPUs produce simulations that are not all identical to results from Intel CPUs, but that the distributions of physics variables are statistically compatible.

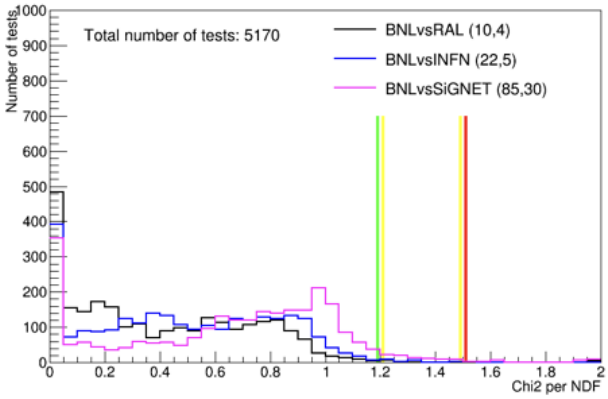


Figure A.7: Effect of running simulations on Intel vs. Intel+AMD vs. AMD sites, on comparison of physics variable histograms.

A.5.5 Comparing step 2. and 3. on Intel and AMD sites

Finally, we compare the output of both the simulation (step 2.) and digitization and reconstruction (step 3.) when run on sites with different CPU types. The Grid site with Intel CPUs is RAL and the site with AMD CPUs is SiGNET. All of the following four permutations are compared: (i) Intel in step 2. + Intel in step 3. (reference), (ii) Intel+AMD, (iii) AMD+Intel and (iv) AMD+AMD.

The results of these comparisons are shown in Figure A.8: Intel+Intel vs. Intel+AMD in the black histogram, Intel+Intel vs. AMD+Intel in the blue histogram and Intel+Intel vs. AMD+AMD in the pink histogram. All three distributions show identical or statistically compatible simulations. The key take-away from this figure seems to be that the simulation (step 2.) introduces more differences compared to the digitization and reconstruction (step 3.), when run on Grid sites with different types of CPU. I.e., step 2. introduces more randomness, and is thereby less strictly reproducible, than step 3. However, any combination will produce sane physics results.

A.5.6 Reproducibility across OS versions

As part of a physics validation of simulations in connection with the upgrade of operating system at the Grid sites, from Scientific Linux CERN 6 (SLC6) to Community ENTerprise Linux Operating System 7 (CentOS 7), I produced simulations on both OSes and compared the resulting physics distributions. The comparisons are shown in Figure A.9, with differences in the hard-scatter processes only in (a) and the addition of pile-up interactions in (b).

Without pile-up, in (a), the distributions are either identical or have minor differences which are significantly smaller than random statistical fluctuations. With the addition of pile-up, in (b), the distributions are either identical or have

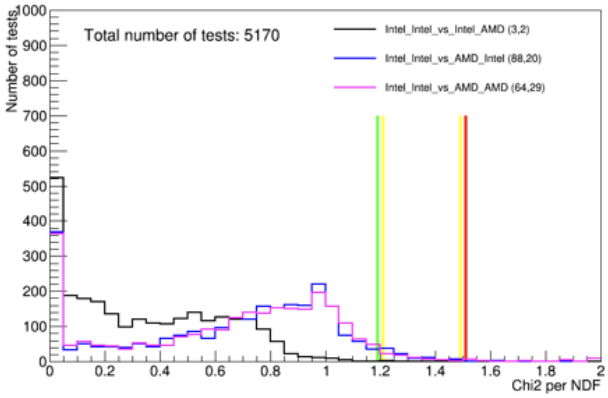
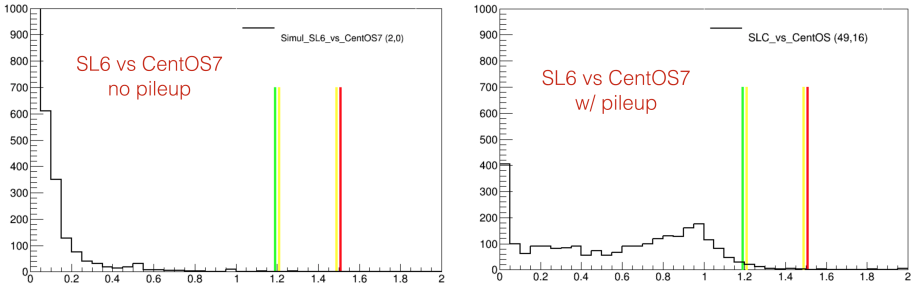


Figure A.8: Effect of running simulation and digitization+reconstruction on all permutations of Intel and AMD CPUs, on comparison of physics variable histograms.



(a) No pile-up

(b) With pile-up

Figure A.9: Physics validation of simulations run on the CentOS 7 operating system, compared with Scientific Linux CERN 6 (SL(C)6).

differences that are consistent with random statistical deviations. In both cases, the physics results should be statistically similar.

A.6 Conclusions

Three simulation settings have been identified as important in order to facilitate complete reproducibility of simulation in heterogeneous and distributed environments on the Worldwide LHC Computing Grid: (i) make use of the same (Intel) math library regardless of CPU type at the site, (ii) use a fixed seed to the random number generator used for the Monte Carlo simulations, and (iii) use the RoundRobin strategy in AthenaMP for reproducible ordering of events when merging the outputs from multiple tasks processed in parallel.

Bit-wise compatibility was achieved when running simulations at two different Intel sites on the Grid, and digitization at the same Intel site. However, bit-wise compatibility was partly tampered in the reconstruction step, as a few events were reconstructed with different physics-related values, given the same input. These findings were reported to the relevant people in ATLAS for follow-up.

Simulations produced at Grid sites running Intel CPUs were found to give either identical or statistically consistent distributions of physics variables, but were not bit-wise compatible.

When running on AMD rather than Intel CPUs, the simulation step was found to introduce more differences than the digitization and reconstruction step. The differences were, however, consistent with at most being random fluctuations with a standard deviation similar to that of the reference (Intel) sample. These findings were reported to the team working on the ATLAS detector simulations using the Geant4 toolkit [55].

Simulations produced at different sites on the Grid can thereby not (yet) be guaranteed to be identical, but they are found to reproduce statistically consistent distributions of physics variables.

Appendix B

Triggers

The trigger names contain information about their definition. All of the triggers listed in Table B.1 and B.2 are so-called high level triggers, and therefore their names are all prefixed by 'HLT'.

The next part of the trigger names indicate the type of leptons they trigger on, the lepton multiplicity and the online p_{T} thresholds. E.g., '2e12' means that the trigger fires if it finds at least two electrons with $E_{\mathrm{T}} > 12\,\mathrm{GeV}$, whereas 'mu18_mu8noL1' fires if it finds at least one muon with $E_{\mathrm{T}} > 18\,\mathrm{GeV}$ in addition to at least one other muon with $E_{\mathrm{T}} > 8\,\mathrm{GeV}$.

Triggers involving electrons [191] also include the name of the so-called identification working point, at which they decide whether an object passes certain quality requirements for being an electron or not. E.g., 'lhloose' uses the *Loose* likelihood-based identification requirements, whereas 'lhvloose' makes use of the *Very loose* likelihood-based identification requirements. The addition of the 'nod0' label in some cases means that no information about the transverse impact parameter d_0 entered into the likelihood function for identification. Some of the electron triggers also include the name of the level 1 (L1) trigger which seeded the HLT. E.g., 'L12EM15VHI' indicates that this 'L1' trigger was fired by '2' objects in the 'EM' calorimeter, with $E_{\mathrm{T}} >$ '15' GeV, where the E_{T} threshold is varied ('V') with η to account for energy loss, and both hadronic core isolation ('H') and elecromagnetic isolation ('I') has been applied.

Some of the triggers involving muons [192] adds a 'noL1' label after a low-p_{T} muon threshold. E.g., 'mu18_mu8noL1' indicates that at least one muon with $p_{\mathrm{T}} > 18\,\mathrm{GeV}$ was found in event information passed from an L1 trigger, by only looking in a specified region of interest, whereas at least one muon with $p_{\mathrm{T}} > 8\,\mathrm{GeV}$ was found in a so-called full scan of the whole muon system, a more CPU intensive search which does not make use of L1 information.

Table B.1: Triggers used in the early Run 2 analyses of the 2015+2016 dataset only.

Year	Type	Trigger
2015	ee	HLT_2e12_lhloose_L12EM10VH
	$\mu\mu$	HLT_mu18_mu8noL1
	$e\mu$	HLT_e17_lhloose_mu14
2016	ee	HLT_2e17_lhvloose_nod0
	$\mu\mu$	HLT_mu22_mu8noL1
	$e\mu$	HLT_e17_lhloose_nod0_mu14

Table B.2: Triggers used in the analyses of the full Run 2 dataset.

Year	Type	Trigger
2015	ee	HLT_2e12_lhloose_L12EM10VH
	$\mu\mu$	HLT_2mu10 HLT_mu18_mu8noL1
	$e\mu$	HLT_e17_lhloose_mu14 HLT_e7_lhmedium_mu24
2016	ee	HLT_2e15_lhvloose_nod0_L12EM13VH HLT_2e17_lhvloose_nod0
	$\mu\mu$	HLT_2mu10 HLT_2mu14 HLT_mu20_mu8noL1 HLT_mu22_mu8noL1
	$e\mu$	HLT_e17_lhloose_nod0_mu14 HLT_e24_lhmedium_nod0_L1EM20VHI_mu8noL1 HLT_e7_lhmedium_nod0_mu24
2017	ee	HLT_2e17_lhvloose_nod0_L12EM15VHI HLT_2e24_lhvloose_nod0
	$\mu\mu$	HLT_2mu14 HLT_mu22_mu8noL1
	$e\mu$	HLT_e17_lhloose_nod0_mu14 HLT_e26_lhmedium_nod0_mu8noL1 HLT_e7_lhmedium_nod0_mu24
2018	ee	HLT_2e17_lhvloose_nod0_L12EM15VHI HLT_2e24_lhvloose_nod0
	$\mu\mu$	HLT_2mu14 HLT_mu22_mu8noL1
	$e\mu$	HLT_e17_lhloose_nod0_mu14 HLT_e26_lhmedium_nod0_mu8noL1 HLT_e7_lhmedium_nod0_mu24

Bibliography

[1] CERN Bulletin,
The Standard Model infographic developed at the CERN webfest, 2012,
URL: https://cds.cern.ch/journal/CERNBulletin/2012/35/News%20Articles/1473657, (accessed: 29.06.2021).

[2] 't Hooft, Gerard,
Renormalizable Lagrangians for Massive Yang-Mills Fields,
Nucl. Phys. B **35** (1971) 167, ed. by Taylor, J.C.

[3] Fritzsch, H., Gell-Mann, Murray, and Leutwyler, H.,
Advantages of the Color Octet Gluon Picture,
Phys. Lett. B **47** (1973) 365.

[4] Gross, David J. and Wilczek, Frank,
Ultraviolet Behavior of Nonabelian Gauge Theories,
Phys. Rev. Lett. **30** (1973) 1343, ed. by Taylor, J.C.

[5] Politzer, H.David, *Reliable Perturbative Results for Strong Interactions?*,
Phys. Rev. Lett. **30** (1973) 1346, ed. by Taylor, J.C.

[6] Glashow, Sheldon L., *The renormalizability of vector meson interactions*,
Nucl. Phys. **10** (1959) 107.

[7] Salam, Abdus and Ward, John Clive,
Weak and electromagnetic interactions, Nuovo Cim. **11** (1959) 568.

[8] Weinberg, Steven, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[9] Englert, F. and Brout, R.,
Broken Symmetry and the Mass of Gauge Vector Mesons,
Phys. Rev. Lett. **13** (1964) 321, ed. by Taylor, J.C.

[10] Higgs, Peter W., *Broken Symmetries and the Masses of Gauge Bosons*,
Phys. Rev. Lett. **13** (1964) 508, ed. by Taylor, J.C.

[11] Guralnik, G.S., Hagen, C.R., and Kibble, T.W.B.,
Global Conservation Laws and Massless Particles,
Phys. Rev. Lett. **13** (1964) 585, ed. by Taylor, J.C.

[12] Tomonaga, S., *On a relativistically invariant formulation of the quantum theory of wave fields*, Prog. Theor. Phys. **1** (1946) 27.

[13] Schwinger, Julian,
Quantum Electrodynamics. I. A Covariant Formulation,
Phys. Rev. **74** (10 1948) 1439,
URL: https://link.aps.org/doi/10.1103/PhysRev.74.1439.

[14] Feynman, R. P., *Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction*, Phys. Rev. **80** (3 1950) 440, URL: https://link.aps.org/doi/10.1103/PhysRev.80.440.

[15] Cabibbo, Nicola, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531.

[16] Kobayashi, Makoto and Maskawa, Toshihide, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

[17] Sakharov, A. D., *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32, [Usp. Fiz. Nauk161,no.5,61(1991)].

[18] 't Hooft, Gerard, *Symmetry Breaking Through Bell-Jackiw Anomalies*, Phys. Rev. Lett. **37** (1976) 8, ed. by Shifman, Mikhail A.

[19] Zyla, P A et al., *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2020** (2020), 083C01, eprint: https://academic.oup.com/ptep/article-pdf/2020/8/083C01/33653179/ptaa104.pdf, URL: https://doi.org/10.1093/ptep/ptaa104.

[20] Riess, Adam G. et al., *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, Astron. J. **116** (1998) 1009, arXiv: astro-ph/9805201 [astro-ph].

[21] Planck Collaboration, *Planck 2018 results. VI. Cosmological parameters*, (2018), arXiv: 1807.06209 [astro-ph.CO].

[22] Super-Kamiokande Collaboration, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562, arXiv: hep-ex/9807003 [hep-ex].

[23] 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** (2001) 071301, arXiv: nucl-ex/0106015 [nucl-ex].

[24] Gouvea, Andre de, Hernandez, Daniel, and Tait, Tim M. P., *Criteria for Natural Hierarchies*, Phys. Rev. **D89** (2014) 115005, arXiv: 1402.2658 [hep-ph].

[25] Rubin, Vera C. and Ford W. Kent, Jr., *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, ApJ **159** (1970) 379.

[26] Zwicky, F., *Die Rotverschiebung von extragalaktischen Nebeln*, Helvetica Physica Acta **6** (1933) 110.

[27] Taylor, A. N., Dye, S., Broadhurst, T. J., Benitez, N., and Kampen, E. van,
Gravitational Lens Magnification and the Mass of Abell 1689,
The Astrophysical Journal **501** (1998) 539,
URL: http://dx.doi.org/10.1086/305827.

[28] Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M.,
The evolution of large-scale structure in a universe dominated by cold dark matter, ApJ **292** (1985) 371.

[29] Georgi, H. and Glashow, S. L., *Unity of All Elementary Particle Forces*,
Phys. Rev. Lett. **32** (1974) 438.

[30] Batzing, Paul and Raklev, Are,
Supersymmetry – Lecture notes for FYS5190/FYS9190 at UiO, 2019,
URL: https://www.uio.no/studier/emner/matnat/fys/FYS5190/h19/pensumliste/notes.pdf, (accessed: 30.06.2021).

[31] Müller-Kirsten, H.J.W. and Wiedemann, A.,
Introduction to Supersymmetry, World Scientific lecture notes in physics,
World Scientific, 2010,
URL: https://books.google.no/books?id=65DkngEACAAJ.

[32] Quevedo, Fernando, Krippendorf, Sven, and Schlotterer, Oliver,
Cambridge Lectures on Supersymmetry and Extra Dimensions, (2010),
arXiv: 1011.1491 [hep-th].

[33] Coleman, Sidney R. and Mandula, J.,
All Possible Symmetries of the S Matrix,
Phys. Rev. **159** (1967) 1251, ed. by Zichichi, A.

[34] Haag, Rudolf, Lopuszanski, Jan T., and Sohnius, Martin,
All Possible Generators of Supersymmetries of the s Matrix,
Nucl. Phys. B **88** (1975) 257.

[35] Martin, Stephen P., *A Supersymmetry primer*,
(1997) 1, [Adv. Ser. Direct. High Energy Phys.18,1(1998)],
arXiv: hep-ph/9709356 [hep-ph].

[36] Alves, Daniele et al., *Simplified models for LHC new physics searches*,
Journal of Physics G: Nuclear and Particle Physics **39** (2012) 105005,
URL: http://dx.doi.org/10.1088/0954-3899/39/10/105005.

[37] Athron, Peter et al.,
Combined collider constraints on neutralinos and charginos,
The European Physical Journal C **79** (2019),
URL: http://dx.doi.org/10.1140/epjc/s10052-019-6837-x.

[38] Covi, Laura, Kim, Jihn E., and Roszkowski, Leszek,
Axinos as cold dark matter, Phys. Rev. Lett. **82** (1999) 4180,
arXiv: hep-ph/9905212.

[39] ATLAS Collaboration,
Summary of the ATLAS experiment's sensitivity to supersymmetry after LHC Run 1 — interpreted in the phenomenological MSSM,
JHEP **10** (2015) 134, arXiv: 1508.06608 [hep-ex].

[40] Evans, Lyndon and Bryant, Philip, *LHC Machine*,
JINST **3** (2008) S08001.

[41] CERN, *The Large Hadron Collider*,
URL: https://home.cern/science/accelerators/large-hadron-collider,
(accessed: 02.07.2021).

[42] Mobs, Esma, *The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019*, (2019), General Photo,
URL: https://cds.cern.ch/record/2684277.

[43] Thomson, Mark, *Modern Particle Physics*,
Cambridge University Press, 2013.

[44] ATLAS Collaboration,
Summary plots from the ATLAS Standard Model physics group,
URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/ATLAS_n_SMSummary_SqrtS/ATLAS_n_SMSummary_SqrtS.png.

[45] LHC SUSY Cross Section Working Group, *SUSY cross sections*, URL:
https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections.

[46] Schott, Matthias and Dunford, Monica,
Review of single vector boson production in pp collisions at $\sqrt{s} = 7$ *TeV*,
Eur. Phys. J. C **74** (2014) 60 p, Comments: 60 pages, 64 figures, For Eur. Phys. J. C, URL: http://cds.cern.ch/record/1699952.

[47] ATLAS Collaboration,
The ATLAS Experiment at the CERN Large Hadron Collider,
JINST **3** (2008) S08003.

[48] ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*,
JINST **15** (2020) P10004, arXiv: 2007.12539 [physics.ins-det].

[49] ATLAS Collaboration, *Luminosity determination in pp collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*, (2019).

[50] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, JINST **15** (2020) P04003,
arXiv: 1911.04632 [physics.ins-det].

[51] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*,
The European Physical Journal C **70** (2010) 823,
URL: http://dx.doi.org/10.1140/epjc/s10052-010-1429-9.

[52] Buckley, Andy et al., *General-purpose event generators for LHC physics*,
Phys. Rept. **504** (2011) 145, arXiv: 1101.2599 [hep-ph].

[53] Gleisberg, T. et al., *Event generation with SHERPA 1.1*,
JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph].

[54] ATLAS Collaboration, *Number of interactions per bunch crossing used in mc16 production (Run 2)*,
URL: https://atlas.web.cern.ch/Atlas/GROUPS/DATAPREPARATION/PublicPlots/2018/General/PLOT-DAPR-2019-01-01.pdf.

[55] Agostinelli, S. et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250, URL: http://www.sciencedirect.com/science/article/pii/S0168900203013688.

[56] ATLAS, Coll et al., *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*,
tech. rep. ATL-PHYS-PUB-2010-013, CERN, 2010,
URL: https://cds.cern.ch/record/1300517.

[57] Alwall, J. et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079,
arXiv: 1405.0301 [hep-ph].

[58] Sjöstrand, T., Mrenna, S., and Skands, P.,
A brief introduction to PYTHIA 8.1,
Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph].

[59] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, (2014).

[60] Ball, Richard D. et al., *Parton distributions with LHC data*,
Nucl. Phys. B **867** (2013) 244, arXiv: 1207.1303 [hep-ph].

[61] Lönnblad, Leif and Prestel, Stefan,
Matching tree-level matrix elements with interleaved showers,
JHEP **03** (2012) 019, arXiv: 1109.4829 [hep-ph].

[62] Beenakker, W., Hopker, R., Spira, M., and Zerwas, P. M.,
Squark and gluino production at hadron colliders,
Nucl. Phys. B **492** (1997) 51, arXiv: hep-ph/9610490.

[63] Kulesza, A. and Motyka, L., *Threshold resummation for squark-antisquark and gluino-pair production at the LHC*,
Phys. Rev. Lett. **102** (2009) 111802, arXiv: 0807.2405 [hep-ph].

[64] Kulesza, A. and Motyka, L., *Soft gluon resummation for the production of gluino-gluino and squark-antisquark pairs at the LHC*,
Phys. Rev. D **80** (2009) 095004, arXiv: 0905.4749 [hep-ph].

[65] Beenakker, Wim et al.,
Soft-gluon resummation for squark and gluino hadroproduction,
JHEP **12** (2009) 041, arXiv: 0909.4418 [hep-ph].

[66] Beenakker, W. et al., *Squark and Gluino Hadroproduction*,
Int. J. Mod. Phys. A **26** (2011) 2637, arXiv: 1105.1110 [hep-ph].

[67] Borschensky, Christoph 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** (2014) 3174, arXiv: 1407.5066 [hep-ph].

[68] Bothmann, Enrico et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034, arXiv: 1905.09127 [hep-ph].

[69] Ball, Richard D. et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph].

[70] Lai, H.-L. et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, arXiv: 1007.2241 [hep-ph].

[71] Gleisberg, Tanju and Höche, Stefan, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph].

[72] Buccioni, Federico et al., *OpenLoops 2*, Eur. Phys. J. C **79** (2019) 866, arXiv: 1907.13071 [hep-ph].

[73] Cascioli, Fabio, Maierhöfer, Philipp, and Pozzorini, Stefano, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv: 1111.5206 [hep-ph].

[74] Denner, Ansgar, Dittmaier, Stefan, and Hofer, Lars, COLLIER*: A fortran-based complex one-loop library in extended regularizations*, Comput. Phys. Commun. **212** (2017) 220, arXiv: 1604.06792 [hep-ph].

[75] Höche, Stefan, Krauss, Frank, Schumann, Steffen, and Siegert, Frank, *QCD matrix elements and truncated showers*, JHEP **05** (2009) 053, arXiv: 0903.1219 [hep-ph].

[76] Höche, Stefan, Krauss, Frank, Schönherr, Marek, and Siegert, Frank, *A critical appraisal of NLO+PS matching methods*, JHEP **09** (2012) 049, arXiv: 1111.1220 [hep-ph].

[77] Höche, Stefan, Krauss, Frank, Schönherr, Marek, and Siegert, Frank, *QCD matrix elements + parton showers. The NLO case*, JHEP **04** (2013) 027, arXiv: 1207.5030 [hep-ph].

[78] Catani, S., Krauss, F., Kuhn, R., and Webber, B. R., *QCD Matrix Elements + Parton Showers*, JHEP **11** (2001) 063, arXiv: hep-ph/0109231.

[79] Schumann, Steffen and Krauss, Frank, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv: 0709.1027 [hep-ph].

[80] Anastasiou, Charalampos, Dixon, Lance J., Melnikov, Kirill, and Petriello, Frank, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order*, Phys. Rev. D **69** (2004) 094008, arXiv: hep-ph/0312266.

[81] Nason, Paolo, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146.

[82] Frixione, Stefano, Laenen, Eric, Motylinski, Patrick, and Webber, Bryan R., *Single-top production in MC@NLO*, JHEP **03** (2006) 092, arXiv: hep-ph/0512250.

[83] Alioli, Simone, Nason, Paolo, Oleari, Carlo, and Re, Emanuele, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph].

[84] Frixione, Stefano, Nason, Paolo, and Ridolfi, Giovanni, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph].

[85] Frixione, Stefano, Nason, Paolo, and Oleari, Carlo, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph].

[86] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016, URL: https://cds.cern.ch/record/2216168.

[87] Sjöstrand, Torbjörn et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph].

[88] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419.

[89] Lange, D. J., *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152.

[90] Re, Emanuele, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv: 1009.2450 [hep-ph].

[91] Frixione, Stefano, Laenen, Eric, Motylinski, Patrick, White, Chris, and Webber, Bryan R., *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) 029, arXiv: 0805.3067 [hep-ph].

[92] Frederix, Rikkert, Re, Emanuele, and Torrielli, Paolo, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, JHEP **09** (2012) 130, arXiv: 1207.5391 [hep-ph].

[93] Alioli, Simone, Nason, Paolo, Oleari, Carlo, and Re, Emanuele, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **09** (2009) 111, arXiv: 0907.4076 [hep-ph], Erratum: JHEP **02** (2010) 011.

[94] Nason, Paolo and Oleari, Carlo, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **02** (2010) 037, arXiv: 0911.5299 [hep-ph].

[95] Hamilton, Keith, Nason, Paolo, Re, Emanuele, and Zanderighi, Giulia, *NNLOPS simulation of Higgs boson production*, JHEP **10** (2013) 222, arXiv: 1309.0017 [hep-ph].

[96] Hamilton, Keith, Nason, Paolo, and Zanderighi, Giulia, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, JHEP **05** (2015) 140, arXiv: 1501.04637 [hep-ph].

[97] Hamilton, Keith, Nason, Paolo, and Zanderighi, Giulia, *MINLO: Multi-Scale Improved NLO*, JHEP **10** (2012) 155, arXiv: 1206.3572 [hep-ph].

[98] Campbell, John M. et al., *NLO Higgs Boson Production Plus One and Two Jets Using the POWHEG BOX, MadGraph4 and MCFM*, JHEP **07** (2012) 092, arXiv: 1202.5475 [hep-ph].

[99] Hamilton, Keith, Nason, Paolo, Oleari, Carlo, and Zanderighi, Giulia, *Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching*, JHEP **05** (2013) 082, arXiv: 1212.4504 [hep-ph].

[100] Catani, Stefano and Grazzini, Massimiliano, *An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC*, Phys. Rev. Lett. **98** (2007) 222002, arXiv: hep-ph/0703012 [hep-ph].

[101] Butterworth, Jon et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph].

[102] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv: 1406.3660 [hep-ex].

[103] Florian, D. de et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, (2016), arXiv: 1610.07922 [hep-ph].

[104] Anastasiou, Charalampos et al., *High precision determination of the gluon fusion Higgs boson cross-section at the LHC*, JHEP **05** (2016) 058, arXiv: 1602.00695 [hep-ph].

[105] Anastasiou, Charalampos, Duhr, Claude, Dulat, Falko, Herzog, Franz, and Mistlberger, Bernhard, *Higgs Boson Gluon-Fusion Production in QCD at Three Loops*, Phys. Rev. Lett. **114** (2015) 212001, arXiv: 1503.06056 [hep-ph].

[106] Dulat, Falko, Lazopoulos, Achilleas, and Mistlberger, Bernhard, *iHixs 2 – Inclusive Higgs cross sections*, Comput. Phys. Commun. **233** (2018) 243, arXiv: 1802.00827 [hep-ph].

[107] Harlander, Robert V. and Ozeren, Kemal J., *Finite top mass effects for hadronic Higgs production at next-to-next-to-leading order*, JHEP **11** (2009) 088, arXiv: 0909.3420 [hep-ph].

[108] Harlander, Robert V. and Ozeren, Kemal J., *Top mass effects in Higgs production at next-to-next-to-leading order QCD: Virtual corrections*, Phys. Lett. B **679** (2009) 467, arXiv: 0907.2997 [hep-ph].

[109] Harlander, Robert V., Mantler, Hendrik, Marzani, Simone, and Ozeren, Kemal J., *Higgs production in gluon fusion at next-to-next-to-leading order QCD for finite top mass*, Eur. Phys. J. C **66** (2010) 359, arXiv: 0912.2104 [hep-ph].

[110] Pak, Alexey, Rogal, Mikhail, and Steinhauser, Matthias, *Finite top quark mass effects in NNLO Higgs boson production at LHC*, JHEP **02** (2010) 025, arXiv: 0911.4662 [hep-ph].

[111] Actis, Stefano, Passarino, Giampiero, Sturm, Christian, and Uccirati, Sandro, *NLO electroweak corrections to Higgs boson production at hadron colliders*, Phys. Lett. B **670** (2008) 12, arXiv: 0809.1301 [hep-ph].

[112] Actis, Stefano, Passarino, Giampiero, Sturm, Christian, and Uccirati, Sandro, *NNLO computational techniques: The cases $H \to \gamma\gamma$ and $H \to gg$*, Nucl. Phys. B **811** (2009) 182, arXiv: 0809.3667 [hep-ph].

[113] Bonetti, Marco, Melnikov, Kirill, and Tancredi, Lorenzo, *Higher order corrections to mixed QCD-EW contributions to Higgs boson production in gluon fusion*, Phys. Rev. D **97** (2018) 056017, arXiv: 1801.10403 [hep-ph], Erratum: Phys. Rev. D **97** (2018) 099906.

[114] Djouadi, A., Kalinowski, J., and Spira, M., *HDECAY: A program for Higgs boson decays in the Standard Model and its supersymmetric extension*, Comput. Phys. Commun. **108** (1998) 56, arXiv: hep-ph/9704448.

[115] Spira, Michael, *QCD Effects in Higgs physics*, Fortsch. Phys. **46** (1998) 203, arXiv: hep-ph/9705337.

[116] Djouadi, A., Mühlleitner, M. M., and Spira, M., *Decays of supersymmetric particles: The Program SUSY-HIT (SUspect-SdecaY-Hdecay-InTerface)*, Acta Phys. Polon. B **38** (2007) 635, arXiv: hep-ph/0609292.

[117] Bredenstein, A., Denner, Ansgar, Dittmaier, S., and Weber, M. M., *Radiative corrections to the semileptonic and hadronic Higgs-boson decays $H \to WW/ZZ \to 4$ fermions*, JHEP **02** (2007) 080, arXiv: hep-ph/0611234.

[118] Bredenstein, A., Denner, Ansgar, Dittmaier, S., and Weber, M. M., *Precise predictions for the Higgs-boson decay H —> WW/ZZ —> 4 leptons*, Phys. Rev. D **74** (2006) 013004, arXiv: hep-ph/0604011 [hep-ph].

[119] Bredenstein, A., Denner, Ansgar, Dittmaier, S., and Weber, M. M., *Precision calculations for the Higgs decays H —> ZZ/WW —> 4 leptons*, Nucl. Phys. Proc. Suppl. **160** (2006) 131, [,131(2006)], arXiv: hep-ph/0607060 [hep-ph].

[120] Ciccolini, M., Denner, Ansgar, and Dittmaier, S., *Strong and Electroweak Corrections to the Production of Higgs + 2 Jets via Weak Interactions at the Large Hadron Collider*, Phys. Rev. Lett. **99** (2007) 161803, arXiv: 0707.0381 [hep-ph].

[121] Ciccolini, Mariano, Denner, Ansgar, and Dittmaier, Stefan, *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the CERN LHC*, Phys. Rev. D **77** (2008) 013002, arXiv: 0710.4749 [hep-ph].

[122] Bolzoni, Paolo, Maltoni, Fabio, Moch, Sven-Olaf, and Zaro, Marco, *Higgs Boson Production via Vector-Boson Fusion at Next-to-Next-to-Leading Order in QCD*, Phys. Rev. Lett. **105** (2010) 011801, arXiv: 1003.4451 [hep-ph].

[123] Ciccolini, M. L., Dittmaier, S., and Krämer, M., *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, Phys. Rev. D **68** (2003) 073003, arXiv: hep-ph/0306234 [hep-ph].

[124] Brein, Oliver, Djouadi, Abdelhak, and Harlander, Robert, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, Phys. Lett. B **579** (2004) 149, arXiv: hep-ph/0307206.

[125] Brein, Oliver, Harlander, Robert, Wiesemann, Marius, and Zirke, Tom, *Top-Quark Mediated Effects in Hadronic Higgs-Strahlung*, Eur. Phys. J. C **72** (2012) 1868, arXiv: 1111.0761 [hep-ph].

[126] Altenkamp, Lukas, Dittmaier, Stefan, Harlander, Robert V., Rzehak, Heidi, and Zirke, Tom J. E., *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, JHEP **02** (2013) 078, arXiv: 1211.5015 [hep-ph].

[127] Denner, Ansgar, Dittmaier, Stefan, Kallweit, Stefan, and Mück, Alexander, *HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders*, Comput. Phys. Commun. **195** (2015) 161, arXiv: 1412.5390 [hep-ph].

[128] Brein, Oliver, Harlander, Robert V., and Zirke, Tom J. E., *vh@nnlo – Higgs Strahlung at hadron colliders*, Comput. Phys. Commun. **184** (2013) 998, arXiv: 1210.5347 [hep-ph].

[129] Harlander, Robert V., Kulesza, Anna, Theeuwes, Vincent, and Zirke, Tom, *Soft gluon resummation for gluon-induced Higgs Strahlung*, JHEP **11** (2014) 082, arXiv: 1410.0217 [hep-ph].

[130] Hartanto, Heribertus B., Jäger, Barbara, Reina, Laura, and Wackeroth, Doreen, *Higgs boson production in association with top quarks in the POWHEG BOX*, Phys. Rev. D **91** (2015) 094003, arXiv: 1501.04498 [hep-ph].

[131] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, (2012).

[132] Martin, A. D., Stirling, W. J., Thorne, R. S., and Watt, G., *Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections*, Eur. Phys. J. C **64** (2009) 653, arXiv: 0905.3531 [hep-ph].

[133] Martin, A. D., Stirling, W. J., Thorne, R. S., and Watt, G., *Parton distributions for the LHC*, Eur. Phys. J. C **63** (2009) 189, arXiv: 0901.0002 [hep-ph].

[134] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016, URL: https://cds.cern.ch/record/2206965.

[135] ATLAS Collaboration, *Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. C **71** (2011) 1577, arXiv: 1012.1792 [hep-ex].

[136] ATLAS Collaboration, *Electron identification measurements in ATLAS using $\sqrt{s}$ = 13 TeV data with 50 ns bunch spacing*, tech. rep., CERN, 2015, URL: https://cds.cern.ch/record/2048202.

[137] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data*, Eur. Phys. J. C **77** (2017) 195, arXiv: 1612.01456 [hep-ex].

[138] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex].

[139] Cacciari, Matteo, Salam, Gavin P., and Soyez, Gregory, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph].

[140] Cacciari, Matteo, Salam, Gavin P., and Soyez, Gregory, *FastJet User Manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: 1111.6097 [hep-ph].

[141] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex].

[142] 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. C **76** (2016) 581, arXiv: 1510.03823 [hep-ex].

[143] Cacciari, Matteo, Salam, Gavin P., and Soyez, Gregory, *The Catchment Area of Jets*, JHEP **04** (2008) 005, arXiv: 0802.1188 [hep-ph].

[144] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, tech. rep. ATL-PHYS-PUB-2016-012, CERN, 2016, URL: https://cds.cern.ch/record/2160731.

[145] ATLAS Collaboration,
Performance of b-Jet Identification in the ATLAS Experiment,
JINST **11** (2016) P04008, arXiv: 1512.01094 [hep-ex].

[146] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data*,
Eur. Phys. J. C **76** (2016) 666, arXiv: 1606.01813 [hep-ex].

[147] ATLAS Collaboration,
Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data,
JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex].

[148] Salzburger, Andreas,
"Track Simulation and Reconstruction in the ATLAS experiment",
PhD thesis: Innsbruck U., 2008.

[149] Cranmer, Kyle,
Statistical challenges for searches for new physics at the LHC, Statistical Problems in Particle Physics, Astrophysics and Cosmology (2006),
URL: http://dx.doi.org/10.1142/9781860948985_0026.

[150] Cousins, Robert D., Linnemann, James T., and Tucker, Jordan,
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 Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **595** (2008) 480,
URL: http://dx.doi.org/10.1016/j.nima.2008.07.086.

[151] Cousins, Robert D., Hymes, Kathryn E., and Tucker, Jordan,
Frequentist evaluation of intervals estimated for a binomial parameter and for the ratio of Poisson means, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **612** (2010) 388,
URL: http://dx.doi.org/10.1016/j.nima.2009.10.156.

[152] ROOT, *Special functions*, URL: https://root.cern.ch/doc/v608/group__SpecFunc.html#ga56454f7cd31cecca1d252ce03a5b83f0,
(accessed: 26.05.2022).

[153] Moneta, Lorenzo et al., *The RooStats Project*, 2011,
arXiv: 1009.1003 [physics.data-an].

[154] Baak, M. et al.,
HistFitter software framework for statistical data analysis,
The European Physical Journal C **75** (2015),
URL: http://dx.doi.org/10.1140/epjc/s10052-015-3327-7.

[155] Verkerke, Wouter and Kirkby, David,
The RooFit toolkit for data modeling, 2003,
arXiv: physics/0306116 [physics.data-an].

[156] Read, Alexander L., *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693.

[157] CERN, *LHC experiments back in business at record energy*, URL: https://home.cern/news/news/accelerators/lhc-experiments-back-business-record-energy-0, (accessed: 14.03.2022).

[158] ATLAS Collaboration, *Search for supersymmetry with two and three leptons and missing transverse momentum in the final state at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector*, ATLAS-CONF-2016-096, 2016, URL: https://cds.cern.ch/record/2212162.

[159] ATLAS Collaboration, *Search for direct production of charginos, neutralinos and sleptons in final states with two leptons and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **05** (2014) 071, arXiv: 1403.5294 [hep-ex].

[160] ATLAS Collaboration, *Search for the electroweak production of supersymmetric particles in $\sqrt{s}$=8 TeV pp collisions with the ATLAS detector*, Phys. Rev. D **93** (2016) 052002, arXiv: 1509.07152 [hep-ex].

[161] CMS Collaboration, *Search for new phenomena in final states with two opposite-charge, same-flavor leptons, jets, and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV*, JHEP **03** (2018) 076, arXiv: 1709.08908 [hep-ex].

[162] ATLAS Collaboration, *Search for electroweak production of charginos and sleptons decaying into final states with two leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector*, Eur. Phys. J. C **80** (2020) 123, arXiv: 1908.08215 [hep-ex].

[163] CERN, *2016: an exceptional year for the LHC*, 2016, URL: https://home.cern/news/news/accelerators/2016-exceptional-year-lhc, (accessed: 24.07.2021).

[164] ATLAS Collaboration, *What is Z_n ?*, tech. rep., 2013, URL: https://indico.cern.ch/event/712580/contributions/2926912/attachments/1627409/2644704/Zn.pdf.

[165] ATLAS Collaboration, *Search for electroweak production of supersymmetric particles in the two and three lepton final state at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector*, ATLAS-CONF-2017-039, 2017, URL: https://cds.cern.ch/record/2267406.

[166] ATLAS Collaboration, *Search for electroweak production of supersymmetric particles in final states with two or three leptons at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector*, Eur. Phys. J. C **78** (2018) 995, arXiv: 1803.02762 [hep-ex], URL: https://cds.cern.ch/record/2307399.

[167] CMS Collaboration, *Combined search for electroweak production of charginos and neutralinos in proton-proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **03** (2018) 160, arXiv: 1801.03957 [hep-ex].

[168] ATLAS Collaboration, *Search for chargino-neutralino production using recursive jigsaw reconstruction in final states with two or three charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **98** (2018) 092012, arXiv: 1806.02293 [hep-ex].

[169] CERN, *A successful conclusion to Run 2*, 2018, URL: https://home.cern/news/opinion/accelerators/successful-conclusion-run-2, (accessed: 21.08.2021).

[170] CERN, *Final lap of the LHC track for protons in 2018*, 2018, URL: https://home.cern/news/news/accelerators/final-lap-lhc-track-protons-2018, (accessed: 21.08.2021).

[171] ATLAS Collaboration, *ATLAS Experiment – Public Results > Annual Plots > 2018 pp Collisions > Pileup Interactions and Data Taking Efficiency*, 2018, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#2018_pp_Collisions, (accessed: 21.08.2021).

[172] 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*, Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [physics.ins-det].

[173] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, tech. rep. ATLAS-CONF-2020-030, CERN, 2020, URL: http://cds.cern.ch/record/2725736.

[174] 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** (2018) 903, arXiv: 1802.08168 [hep-ex].

[175] ATLAS Collaboration, *E_T^{miss} performance in the ATLAS detector using 2015-2016 LHC p-p collisions*, tech. rep., CERN, 2018, URL: https://cds.cern.ch/record/2625233.

[176] ATLAS Collaboration, *Object-based missing transverse momentum significance in the ATLAS detector*, (2018).

[177] Collins, John C., Soper, Davison E., and Sterman, George, *Factorization of Hard Processes in QCD*, 2004, arXiv: hep-ph/0409313 [hep-ph].

[178] Gribov, V. N. and Lipatov, L. N., *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438.

[179] Altarelli, Guido and Parisi, G., *Asymptotic Freedom in Parton Language*, Nucl. Phys. B **126** (1977) 298.

[180] Dokshitzer, Yuri L., *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov. Phys. JETP **46** (1977) 641.

[181] Zyla, P. A. et al., *Review of Particle Physics*, PTEP **2020** (2020) 083C01.

[182] Butterworth, J. M. et al., "THE TOOLS AND MONTE CARLO WORKING GROUP Summary Report from the Les Houches 2009 Workshop on TeV Colliders", *6th Les Houches Workshop on Physics at TeV Colliders*, 2010, arXiv: 1003.1643 [hep-ph].

[183] LHEF, *LHEF: Les Houches Event File*, URL: http://home.thep.lu.se/~leif/LHEF/, (accessed: 12.06.2022).

[184] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, **2/2017** (2016), arXiv: 1610.07922 [hep-ph].

[185] ATLAS Collaboration, *Multi-Boson Simulation for 13 TeV ATLAS Analyses*, (2016).

[186] ATLAS Collaboration, *Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (2020), arXiv: 2007.02645 [hep-ex].

[187] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data*, JINST **14** (2019) P03017, arXiv: 1812.03848 [hep-ex].

[188] ATLAS Collaboration, *Tagging and suppression of pileup jets*, (2014).

[189] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 580, [Erratum: Eur.Phys.J.C 77, 712 (2017)], arXiv: 1705.02211 [hep-ex].

[190] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 970, arXiv: 1907.05120 [hep-ex].

[191] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, Eur. Phys. J. C **80** (2020) 47, arXiv: 1909.00761 [hep-ex].

[192] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, JINST **15** (2020) P09015, arXiv: 2004.13447 [physics.ins-det].

[193] ATLAS Collaboration, *Searches for new phenomena in events with two leptons, jets, and missing transverse momentum in* 139 fb^{-1} *of* $\sqrt{s} = 13$ *TeV pp collisions with the ATLAS detector*, (2022), arXiv: 2204.13072 [hep-ex].

[194] Cousins, Robert D., Linnemann, James T., and Tucker, Jordan, *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*, Nucl. Instrum. Meth. A **595** (2008) 480, arXiv: physics/0702156 [physics.data-an].

[195] ATLAS Collaboration, *Search for electroweak production of supersymmetric particles in final states with two or three leptons at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Eur. Phys. J. C **78** (2018) 995, arXiv: 1803.02762 [hep-ex].

[196] ATLAS Collaboration, *Search for direct production of charginos, neutralinos and sleptons in final states with two leptons and missing transverse momentum in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, JHEP **05** (2014) 071, arXiv: 1403.5294 [hep-ex].

[197] CMS Collaboration, *Search for supersymmetry in final states with two oppositely charged same-flavor leptons and missing transverse momentum in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **04** (2021) 123, arXiv: 2012.08600 [hep-ex].

[198] D0 Collaboration, *Evidence for production of single top quarks*, Phys. Rev. D **78** (2008) 012005, arXiv: 0803.0739 [hep-ex].

[199] D0 Collaboration, *Observation of Single Top Quark Production*, Phys. Rev. Lett. **103** (2009) 092001, arXiv: 0903.0850 [hep-ex].

[200] CDF Collaboration, *First Observation of Electroweak Single Top Quark Production*, Phys. Rev. Lett. **103** (2009) 092002, arXiv: 0903.0885 [hep-ex].

[201] Baldi, Pierre, Sadowski, Peter, and Whiteson, Daniel, *Searching for Exotic Particles in High-Energy Physics with Deep Learning*, Nature Commun. **5** (2014) 4308, arXiv: 1402.4735 [hep-ph].

[202] Chen, Tianqi and Guestrin, Carlos, *XGBoost*, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016), URL: http://dx.doi.org/10.1145/2939672.2939785.

[203] Kaggle, *Kaggle: Your Machine Learning and Data Science Community*, URL: https://kaggle.com, (accessed: 07.11.2021).

[204] Adam-Bourdarios, Claire et al.,
"The Higgs boson machine learning challenge", *Proceedings of the NIPS 2014 Workshop on High-energy Physics and Machine Learning*,
ed. by Cowan, Glen, Germain, Cécile, Guyon, Isabelle, Kégl, Balázs, and Rousseau, David, vol. 42, Proceedings of Machine Learning Research,
PMLR, 2015 19, URL: https://proceedings.mlr.press/v42/cowa14.html.

[205] HiggsML,
HEP meets ML award | The Higgs Machine Learning Challenge, 2014,
URL: https://higgsml.ijclab.in2p3.fr/prizes-and-award/award/,
(accessed: 07.11.2021).

[206] ATLAS Collaboration, *Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector*, tech. rep., CERN, 2017,
URL: https://cds.cern.ch/record/2275641.

[207] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, tech. rep., CERN, 2019,
URL: http://cds.cern.ch/record/2688062.

[208] Hoecker, Andreas et al.,
"TMVA - Toolkit for Multivariate Data Analysis", 2007.

[209] Breiman, Leo, *Bagging Predictors*, Machine Learning **24** (1996) 123.

[210] Ho, Tin Kam, "Random decision forests", *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, 1995 278.

[211] Sharpr and kakau, *Receiver operating characteristic*,
URL: https://en.wikipedia.org/wiki/Receiver_operating_characteristic,
(accessed: 28.11.2021).

[212] Walber, *Precision and recall*,
URL: https://en.wikipedia.org/wiki/Precision_and_recall,
(accessed: 28.11.2021).

[213] ATLAS Collaboration,
Search for direct production of charginos, neutralinos and sleptons in final states with two leptons and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, JHEP **05** (2014) 071,
arXiv: 1403.5294 [hep-ex].

[214] HiLumiLHC, *The HL-LHC project*,
URL: https://hilumilhc.web.cern.ch/content/hl-lhc-project,
(accessed: 27.03.2022).

[215] Binet, S., Calafiura, P., Snyder, S., Wiedenmann, W., and Winklmeier, F.,
Harnessing multicores: Strategies and implementations in ATLAS,
J. Phys. Conf. Ser. **219** (2010) 042002, ed. by Gruntorad, Jan and Lokajicek, Milos.

[216] Geek3, *Chi-squared distribution*,
URL: https://en.wikipedia.org/wiki/Chi-squared_distribution,
(accessed: 30.01.2022).