Université de Montréal

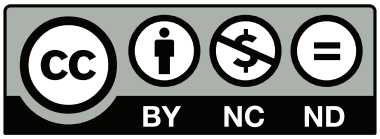
Searching for supersymmetry using deep learning with the ATLAS detector

par
Louis-Guillaume Gagnon

Département de physique
Faculté des arts et science

Thèse présentée en vue de l'obtention du grade de
Philosophiae doctor (Ph.D.) en physique

7 juillet 2020



Université de Montréal
Département de physique, Faculté des arts et science

Cette thèse intitulée

Searching for supersymmetry using deep learning with the ATLAS detector

Présentée par

Louis-Guillaume Gagnon

A été évaluée par un jury composé des personnes suivantes

Laurence Perreault Levasseur
Présidente

Jean-François Arguin
Directeur de recherche

Yashar Hezaveh
Membre du jury

Dag Gilberg (Carleton)
Examinateur externe

Résumé

Le Modèle Standard de la physique des particules (MS) est une théorie fondamentale de la nature dont la validité a été largement établie par diverses expériences. Par contre, quelques problèmes théoriques et expérimentaux subsistent, ce qui motive la recherche de théories alternatives. La Supersymétrie (SUSY), famille de théories dans laquelle une nouvelle particule est associée à chaque particules du MS, est une des théories ayant les meilleures motivations pour étendre la portée du modèle. Par exemple, plusieurs théories supersymétriques prédisent de nouvelles particules stables et interagissant seulement par la force faible, ce qui pourrait expliquer les observations astronomiques de la matière sombre. La découverte de SUSY représenterait aussi une importante étape dans le chemin vers une théorie unifiée de l'univers. Les recherches de supersymétrie sont au coeur du programme expérimental de la collaboration ATLAS, qui exploite un détecteur de particules installé au Grand Collisioneur de Hadrons (LHC) au CERN à Genève, mais à ce jours aucune preuve en faveur de la supersymétrie n'a été enregistrée par les présentes analyses, largement basées sur des techniques simples et bien comprises.

Cette thèse documente l'implémentation d'une nouvelle approche à la recherche de particules basée sur l'apprentissage profond, utilisant seulement les quadri-impulsions comme variables discriminatoires; cette analyse utilise l'ensemble complet de données d'ATLAS enregistré en 2015–2018. Les problèmes de la naturalité du MS et de la matière sombre orientent la recherche vers les partenaires supersymétriques du gluon (le gluino), des quarks de troisième génération (stop et sbottom), ainsi que des bosons de gauge (le neutralino). Plusieurs techniques récentes sont employées, telles que l'utilisation directe des quadri-impulsions reconstruites à partir des données enregistrées par le détecteur ATLAS ainsi que la paramétrisation d'un réseau de neurone avec les masses des particules recherchées, ce qui permet d'atteindre une performance optimale quelle que soit l'hypothèse de masses. Cette méthode améliore la signification statistique par un facteur 85 par rapport au dernier résultat d'ATLAS pour certaines hypothèses de masses, et ce avec la même luminosité.

Aucun excès signifif au-delà du Modèle Standard n'est observé. Les masses du gluino en deçà de 2.45 TeV et du neutralino en deça de 1.7 TeV sont exclues à un niveau de confiance de 95%, ce qui étend largement les limites précédentes sur deux modèles de productions de paires de gluinos faisant intervenir des stops et des sbottoms, respectivement.

Mots-clés: Physique des particules, Supersymétrie, LHC, ATLAS, Apprentissage machine, Apprentissage profond, Réseaux de neurones.

Abstract

The Standard Model of particle physics (SM) is a fundamental theory of nature whose validity has been extensively confirmed by experiments. However, some theoretical and experimental problems subsist, which motivates searches for alternative theories to supersede it. Supersymmetry (SUSY), which associate new fundamental particles to each SM particle, is one of the best-motivated such theory and could solve some of the biggest outstanding problems with the SM. For example, many SUSY scenarios predict stable neutral particles that could explain observations of dark matter in the universe. The discovery of SUSY would also represent a huge step towards a unified theory of the universe. Searches for SUSY are at the heart of the experimental program of the ATLAS collaboration, which exploits a state-of-the-art particle detector installed at the Large Hadron Collider (LHC) at CERN in Geneva. The probability to observe many supersymmetric particles went up when the LHC ramped up its collision energy to 13 TeV, the highest ever achieved in laboratory, but so far no evidence for SUSY has been recorded by current searches, which are mostly based on well-known simple techniques such as counting experiments.

This thesis documents the implementation of a novel deep learning-based approach using only the four-momenta of selected physics objects, and its application to the search for supersymmetric particles using the full ATLAS 2015–2018 $\sqrt{s} = 13$ TeV dataset. Motivated by naturalness considerations as well as by the problem of dark matter, the search focuses on finding evidence for supersymmetric partners of the gluon (the gluino), third generation quarks (the stop and the sbottom), and gauge bosons (the neutralino). Many recently introduced physics-specific machine learning developments are employed, such as directly using detector-recorded energies and momenta of produced particles instead of first deriving a restricted set of physically motivated variables and parametrizing the classification model with the masses of the particles searched for, which allows optimal sensitivity for all mass hypothesis. This method improves the statistical significance of the search by up to 85 times that of the previous ATLAS analysis for some mass hypotheses, after accounting for the luminosity difference.

No significant excesses above the SM background are recorded. Gluino masses below 2.45 TeV and neutralino masses below 1.7 TeV are excluded at the 95% confidence level, greatly increasing the previous limit on two simplified models of gluino pair production with off-shell stops and sbottoms, respectively.

Keywords: Particle physics, Supersymmetry, LHC, ATLAS, Machine learning, Deep learning, Neural networks.

Contents

List of Tables

List of Figures

List of abbreviations

AI Artificial Intelligence

ADAM Adaptive Moments algorithm

ATLAS A Toroidal LHC Apparatus

BDT Boosted Decision Tree

BSM Beyond the Standard Model

CERN European Organization for Nuclear Research

CCA Cut-And-Count Analysis

CDM Cold Dark Matter

CKM Cabibbo-Kobayashi-Maskawa matrix

CL Confidence Level

CMB Cosmic Microwave Background

CMS Compact Muon Solenoid

CoM Center-of-mass

CP Charge–Parity

CR Control Region

CSC Cathode Strip Chambers

DMDD Dark Matter Direct Detection

EWK Electroweak

FWHM Full width at half-maximum

GMSB Gauge-Mediated Supersymmetry Breaking

GPU Graphical Processing Unit

GUT Grand Unified Theory

HLT High-Level Trigger

KF Kalman Filter

IBL Insertable B-Layer

ID Inner Detector

IP Interaction Point

L1 Level-1 trigger

LHC Large Hadron Collider

LEE Look-Elsewhere Effect

LEP Large Electron-Positron Collider

LLR Log Likelihood Ratio

LSP Lightest Supersymmetric Particle

MC Monte Carlo

MDT Monitored Drift Tubes

MOND MOdified Newtonian Dynamics

MSSM Minimal Supersymmetric Standard Model

NN Neural Network

OR Overlap Removal

PDF Parton Distribution Function

PS Proton–Synchrotron

PSB Proton–Synchrotron Booster

QCD Quantum Chromodynamics

QED Quantum Electrodynamics

QFT Quantum Field Theory

QM Quantum Mechanics

ReLU Rectified Linear Unit

RF Radio frequency

RoI Region of Interest

ROC Receiver Operating Characteristic

RPC R-Parity Conservation

RNN Recurrent Neural Network

SCT Semiconductor Tracker

SGD Stochastic Gradient Descent

SM Standard Model

SPS Super-Proton–Synchrotron

S$p\overline{p}$S Super-Proton–Antiproton–Synchrotron

SR Signal Region

SUSY Supersymmetry

TDAQ Trigger and Data Acquisition

TRT Transition Radiation Tracker

UV Ultraviolet

VR Validation Region

WIMP Weakly-Interacting Massive Particle

mSUGRA minimal SuperGravity

pMSSM phenomenological MSSM

Remerciements

J'aimerais tout d'abord remercier mes parents, Marc et Isabelle, pour m'avoir toujours encouragé et supporté, ainsi que mes deux soeurs, Laurence et Émilie. Un gros merci à tous mes amis à qui j'ai cassé les oreilles avec mes projets académiques, particulièrement à mes camarades de S'cuse[1], Arno et Sédrick ; la musique donne un sens plus fort à la vie que n'importe quelle science. Et un merci tout spécial à Chloé, qui a été à mes côtés durant toute cette aventure ; sans toi le chemin aurait été très difficile, voire même impraticable. Merci de m'inspirer.

J'aimerais aussi remercier tous mes collègues de l'Université de Montréal, spécialement ceux et celles du groupe ATLAS-UdeM ; les discussions que nous avons partagé à travers les années ont été très enrichissantes et, je l'espère, continuerons de l'être dans le futur malgré les chemins parfois très différents que nous choisissons. Je me dois de mentionner mon directeur de recherche, Jean-François Arguin ; merci d'avoir cru en moi et de m'avoir donné l'espace et la confiance nécessaire pour m'épanouir en tant que chercheur. Je t'en serai éternellement reconnaissant.

Finally, I'd like to thank all of my colleagues from the CERN and ATLAS worlds, especially members of the CTIDE and multi-b-jets groups. In particular, I want to thank Max and Chiara; without your patient help and your expertise, I don't think this thesis would have seen the light of day.

Merci!

[1] `https://scuse.bandcamp.com`

Statement of personal contributions

As the ATLAS experiment is a huge collaboration of more than 3000 physicists, it can be tricky to disentangle personal contributions to the densely connected web of knowledge woven by the numerous analyzers who are active on any given project.

During my graduate studies, I mainly worked on two subjects, both of which are treated in this thesis. First of all, I've been significantly involved in the search for supersymmetric particles in events with many b-jets and missing transverse energy. The so-called "strong multi-b" analysis group has already produced a paper and a public result with partial datasets of 36.1 fb^{-1} and 79.9 fb^{-1}, respectively; the search presented in this thesis will eventually be published in an upcoming paper, this time with the full ATLAS 2015–2018 dataset (139 fb^{-1}). My biggest contribution to this analysis is the implementation and optimization of a deep learning classifier that can distinguish supersymmetric events from Standard Model background processes and is treated in detail in Chapter 5 of this thesis. I have also been involved in many other aspects of the analysis, such as configuring and coordinating the production of large (terabyte-scale) simulated particle collision datasets.

During the first three years of my doctoral studies, I've also contributed to the functioning of the ATLAS experiment by working extensively on a machine learning algorithm that aims to recognize charge clusters originating from more than one particle in the ATLAS pixel sub-detector. While I have not participated to the original implementation of the method, I have performed many optimization and characterization studies; these are documented in Chapter 4.

« *La science a cette salutaire capacité à nous montrer que l'Univers n'est pas qu'une construction conventionnelle : quelque chose d'étrange, souvent de magnifique, parfois de stupéfiant, se révèle à nous. Penser en scientifique, c'est d'abord accepter de se laisser surprendre ; c'est ne pas enclore le réel dans ce que nous souhaitons qu'il soit ; c'est vouloir penser au-delà de nos fantasmes et de nos croyances. En demeurant conscient des limites évidentes de cette démarche : la pensée ne se distancie jamais d'elle même. C'est cet effort teinté d'impossible qui, sans doute, fonde le geste.* »

Aurélien Barreau

Introduction

In the year 2020, it seems that the field of particle physics is in very good shape. The Standard Model of particle physics, an elegant and very successful fundamental theory of nature, is theoretically complete since the 70s, and all its constituent fundamental particles have been observed in various experiments, from the discovery of the electron in the late 19th century [1] to the recent observation of the Higgs boson [2, 3] by the ATLAS [4] and CMS [5] collaborations at the LHC [6]. And yet, searches for so-called Beyond-the-Standard-Model physics are still ongoing in full force; there are even serious discussions about building bigger and more powerful particle accelerators to carry out this work in the future. Beyond the need to validate the Standard Model with high-precision measurements, which might still uncover flaws in the theory, there is still a number of theoretical problems with the Standard Model that motivate this state of affairs. For instance, the model does not explain astronomical observations of dark matter; nor does it explain the huge energy difference between the characteristic scale of weak interactions and that of gravitational phenomena. In fact, the Standard Model does not even try to describe gravity, instead positing that a new model is needed for energy scales above the Plank mass.

Supersymmetry, an extension of the fundamental space–time symmetries that associates new bosons to Standard Model fermions and vice-versa, was sometimes touted in pre-LHC times as an easy and elegant way to fix such problems. Indeed, the projected cross-sections for a large class of supersymmetric processes would lead to clear effects at the LHC if realized in nature; to this date, no clear and unambiguous sign of such effects have been detected by any of the LHC collaborations, or by any other experiments. Faced with this situation, we then have, at least, three paths moving forward. Firstly, we could simply abandon altogether supersymmetry searches and focus on alternate Beyond-the-Standard-Model theories. Secondly, we could also just note that after its second data-taking run, the LHC has only produced about five percent of its eventual final dataset and be patient, hoping that supersymmetry lurks right around the corner and that more data will clarify the situation. The third option, which is not mutually exclusive to the second

one, is to improve existing search methodology or to implement new search strategies to maximize the discovery potential given the data that we have right now.

Recent developments in the field of artificial intelligence, most notably the so-called *deep learning revolution* [7], have pushed the boundaries of machine learning techniques further than would have been thought possible a few decades ago. Granted, machine learning has already been in extensive usage throughout the field of high-energy physics for already quite some time, but the recent renaissance of the field of neural network research has lead to many new possibilities. The reasons are many: the existence of well-understood models such as deep neural networks along with enough data to train them; ubiquity of accelerated hardware in the form of graphical processing units (GPU); and the availability of high-quality free-software libraries enabling researchers to tap into the power of modern AI without reinventing the wheel.

With these reasons in mind, this thesis will apply deep learning techniques to the problem of the search for supersymmetry at the LHC, using data recorded by the ATLAS detector. It will build upon previous work by the ATLAS collaboration in which supersymmetric partners of the gluon, third generation quarks, and gauge bosons are searched for in events with a significant amount of missing transverse energy and many b-jets [8, 9]. More specifically, a single neural network is trained to classify events as originating from a supersymmetric signal or from a Standard Model background, using low-level (four-momenta) inputs reconstructed from ATLAS detector data.

In Chapter 1.1, the theoretical underpinnings of the Standard Model are briefly reviewed, including a discussion of some of the aforementioned problems, before presenting in Chapter 2 the proposed solution, supersymmetry. The LHC and the ATLAS detector are presented in Chapter 3. In Chapter 4, the theory of neural networks is presented, along with an example application in the context of track reconstruction. Finally, the search itself is presented in Chapter 5, before concluding.

Chapter 1

The Standard Model

1.1 $SU(3) \times SU(2) \times U(1)$: Theoretical overview

Note: unless otherwise noted, this section is based on References [10–13]

Modern particle physics is formulated within a framework known as *Quantum Field Theory*, or QFT. In this framework, pretty much everything is unsurprisingly described in terms of quantum fields; what we call particles are really just excitations of fundamental quantum fields away from their ground states [14]. There are two basic classes of fields, distinguished by the nature of the spins of the particles they carry. *Fermionic* fields allow for half-integer spin states, while *bosonic* fields allow integer spin states. What we commonly refer to as matter are excitations in fermionic fields with spin = $1/2$. What we think of as forces are due to excitations in the bosonic fields; local interactions arise due to the exchange of force carrying *gauge bosons* that have unit spin. Another important field is the spinless bosonic field, which describes *scalar bosons*; it enters the theory as the *Higgs field*.

But how is the Standard Model built from these ingredients? The basic recipe is to start with non-interacting matter only, and requiring something called *local gauge invariance*, which implies the existence of the force carrying gauge bosons. In Section 1.1.1, we review the fundamental concept of local gauge invariance, and take a look at how it can give rise to a realistic theory: Quantum Electrodynamics, or QED. We then follow the same blueprint to build the Standard Model of particle physics (SM) itself: starting with the SM fermions, we use gauge theory to describe the two sub-theories of the SM; Quantum Chromodynamics (QCD) in Section 1.1.2 and the electroweak model in Section 1.1.3.

1.1.1 Local gauge invariance

> *"[Local gauge invariance] is the essential concept out of which the Standard Model is built: a concept that has all the features of a fundamental principle of nature."*
>
> – Gian Guidice [15]

As promised, we start from one of the simplest possible situations: a single free fermion with mass m. For familiarity let's call it an *electron*. From QFT, we know that we need to use the Dirac Lagrangian to derive the equation of motion of spin-$1/2$ fermions. In the position basis:

$$\mathcal{L} = \overline{\psi}(i\partial\!\!\!/ - m)\psi, \tag{1.1}$$

in which we used the "Feynman slash notation", $\gamma^\mu \partial_\mu \equiv \partial\!\!\!/$, to contract the four-momentum with the Dirac matrices γ^μ. Using the Euler-Lagrange equations, we obtain the Dirac equation:

$$(i\partial\!\!\!/ - m)\psi = 0, \tag{1.2}$$

which is a differential equation with a plane wave solution of the form:

$$\psi = \sum_p \omega(p) e^{ip_\mu x^\mu}, \tag{1.3}$$

where the $\omega(p)$ terms are the normalization coefficients for each momentum mode.

Let's now try a first gauge transformation, that of the *global* unitary group of order one, $U(1)$. This group describes phase transformations of the wave-function without any change in normalization such that unitarity is preserved. Applying this to a wave with a single momentum mode:

$$\psi \to e^{i\theta}\psi = \omega p e^{ip_\mu x^\mu + \theta}, \tag{1.4}$$

which is still a solution of the Dirac equation. Let's consider the effect of the transformation on the Lagrangian:

$$\mathcal{L} = \overline{\psi}(i\not{\partial} - m)\psi \rightarrow \overline{\psi}e^{-i\theta}(i\not{\partial} - m)e^{i\theta}\psi = \mathcal{L}'. \tag{1.5}$$

Since this is a global transformation, $e^{i\theta}$ is constant; thus, the contributions from $\overline{\psi}$ and ψ cancel out and $\mathcal{L} = \mathcal{L}'$. Such transformations that leave the Lagrangian unchanged are called *global gauge symmetries*, and are symptomatic of fundamental redundancies in the mathematical description of the system.

This doesn't seem very important or interesting. We simply pick the most convenient representation for the task at hand from the gauge group once and for all; the choice doesn't fundamentally matter, since all these representations are exactly equivalent. But if indeed they are, it's not a big stretch to wonder why we have to pick only one representation for all space-time; it might be logically or aesthetically pleasing to be able to smoothly interpolate between different representations throughout space-time. In other words, it's almost unavoidable that we would ask ourselves what happens when we use a *local* transformation group rather than a global one:

$$\psi \rightarrow \Theta(x)\psi = e^{iq\theta(x)}\psi. \tag{1.6}$$

An explicit scale factor q for the phase transformation was introduced, for reasons that will soon become apparent. The dependence on space-time makes the Lagrangian non-invariant under the transformation, since we get an extra term when taking the derivative:

$$\partial_\mu\psi \rightarrow \partial(\Theta\psi) = (\partial\Theta)\psi + \Theta\partial\psi, \tag{1.7}$$

$$\implies \mathcal{L} \rightarrow \mathcal{L} - \overline{\psi}q\not{\partial}\theta(x)\psi. \tag{1.8}$$

In plain English: the theory of the free electron does not satisfy local gauge invariance. To make it so, the model needs to be extended; let's start by introducing a gauge boson, A_μ, which we will call the *photon*, for familiarity. Our first task is to incorporate it in the Lagrangian:

$$\mathcal{L} = \overline{\psi}(i\not{\partial} - q\not{A} - m)\psi. \tag{1.9}$$

Our second and final task is to modify the transformation to simultaneously act on both

the electron and the photon:

$$\begin{cases} \psi & \to e^{iq\theta(x)}\psi, \\ A_\mu & \to A_\mu - \partial\theta(x). \end{cases} \tag{1.10}$$

Clearly, the extra $-\partial\theta(x)$ term from the transformation of the photon cancels the unwanted term in Eq. (1.8). Thus, the Lagrangian of Eq. (1.9) is invariant under the local gauge transformation of Eq. (1.10). After having done so, the theory is no longer free since the following term appeared in the Lagrangian:

$$iq\not{A}\overline{\psi}\psi, \tag{1.11}$$

which is an interaction term between electrons and the force-carrying photon, with overall strength governed by q, which we can call the *electric charge*, for familiarity.

In summary, we've started from the theory of the free electron, and we've seen that requiring local gauge invariance forced us to introduce the photon in the theory. As a consequence, there appeared an unavoidable interaction between electrons, which also gained a charge. The resulting theory is called *Quantum Electrodynamics* [16–21], or QED, and it is sufficient to accurately describe all electromagnetic phenomena. It is without a doubt one of the most successful physical theories ever devised; for instance, the experimental measurement of the electron's magnetic moment matches the theoretical prediction up to a $10^{-7}\%$ deviation [22]!

In the next few sections we follow the same simple procedure to derive the Standard Model itself, in which QED is embedded.

1.1.2 SU(3): Quantum Chromodynamics

We've just seen how we can derive a realistic theory by starting from a free electron and requiring invariance under the $U(1)$ local gauge group. Let's now try something a little different – starting from free *quarks* [23], we'll require invariance under local gauge transformations described by *the special unitary group of order 3*, $SU(3)$.

Instead of a single $\theta(x)$ scalar term, the transformations corresponding to $SU(3)$ are generated by a basis of eight[1] 3×3 matrices:

[1] $3 \times 3 = 9$, so why not nine parts? What makes this gauge group "special" is that the transformation matrix has unit determinant, which removes one degree of freedom.

$$S = \lambda_1\phi_1 + \lambda_2\phi_2 \ldots \lambda_8\phi_8. \tag{1.12}$$

Here, the λ terms are the Gell-Mann matrices. Let's write the transformation a little more explicitly for the k-th of the eight components in exponential form:

$$\psi \rightarrow S_k\psi = e^{-iq\lambda_k\phi_k}\psi. \tag{1.13}$$

We already know what will happen, since it is analogous to Eq. 1.8; the free fermionic Lagrangian acquires an extra term:

$$\mathcal{L} \rightarrow \mathcal{L} - iq\overline{\psi}\not{\partial}(\lambda_k\phi_k)\psi. \tag{1.14}$$

We know what to do. First, add a gauge boson Λ_k and obtain:

$$\mathcal{L} = \overline{\psi}(i\not{\partial} - iq\lambda\not{\Lambda}_k - m)\psi. \tag{1.15}$$

Then, add the Λ_k transformation term, which takes care of the extra term. Are we done now? If so, then the situation is quite similar to the QED case, with eight new bosons instead of a single one. There is, however, one final piece missing. When we derived QED, we obtained the Lagrangian of Eq. 1.9, but it is incomplete; it is missing the part for the free boson, $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. This was of no great consequence for this term is trivially gauge invariant, but the situation is different in the SU(3) case. There are now eight gauge bosons, each represented by a generator *matrix*; since matrix multiplication is not commutative, the new free Lagrangian has a mixed terms that forbids us to think of each of the eight transformations in purely separate terms:

$$G_k^{\mu\nu} = \partial^\mu\Lambda_k^\nu - \partial^\nu\Lambda_k^\mu - 2q\sum_{i,j=1}^{8} f_{kij}\Lambda_i^\mu\Lambda_j^\nu, \tag{1.16}$$

where f_{kij} are the structure constants of SU(3). The gauge transformation must account for this extra term, and we end up with:

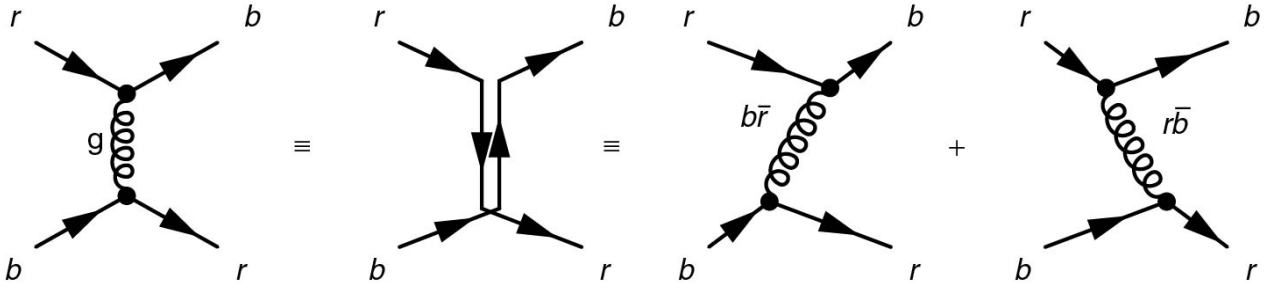


Figure 1.1 – Color-changing interaction via gluon exchange. In the basis of the left diagram, the gluon is in a superposition of color states, $g = \frac{1}{\sqrt{2}}(b\bar{r} + r\bar{b})$ [11].

$$\begin{cases} \psi & \to e^{iq\lambda_k\phi_k}\psi, \\ \Lambda_k^\mu & \to \Lambda_k^\mu - \partial\phi_k - 2q\sum_{i,j=1}^{8} f_{kij}\phi_i\Lambda_j. \end{cases} \tag{1.17}$$

The implications are rich. First, since the $SU(3)$ transformations are represented by 3×3 matrices, the wave-function ψ must have a three-component part that represents the charge of the interaction (much like the electron became electrically charged when we added-in the photons). The three components of the charge $SU(3)$ charge basis are labeled red, green, and blue; the resulting theory is called *Quantum Chromodynamics* [24–26], or QCD. Particles that carry such charges are *colored* particles, and colored fermions are precisely the three Standard Model quarks. Each color has to be represented, and since each quark has single unit of color, there actually are three "copies" of each quark, one for each charge. The theory gained eight bosons: the *gluons*, which form a color octet and allow for color changing interactions as seen in Figure 1.1. Moreover, the appearance of terms involving the $SU(3)$ structure constant imply that gluons are self-coupled, that is, there exist 3- and 4-gluon interactions.

The full QCD Lagrangian is:

$$\mathcal{L}_{\text{QCD}} = \sum_{q,c} \overline{\psi}_{q,c}(i\not{\partial} - m_q)\psi_{q,c} \tag{1.18}$$

$$- \sum_{q,c,c',k} \overline{\psi}_{q,c}\alpha_s\lambda_{kcc'}\not{\Lambda}_k\psi_{q,c'} \tag{1.19}$$

$$- \frac{1}{4}G_{k,\mu\nu}G_k^{\mu\nu}, \tag{1.20}$$

where q indexes the quark flavor (6 in total), c and c' are color indices and k is the gluon

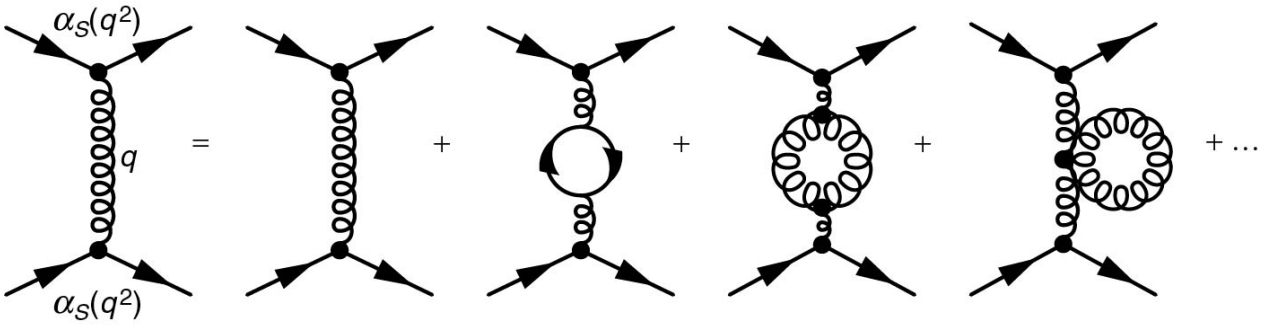


Figure 1.2 – Through the renormalization procedure, a sum of a number of diagrams involving the bare QCD coupling constant is replaced by a single diagram with an energy-dependent coupling [11].

index. The second term represent the quark-quark-gluon color-changing interactions while the third term allows three- and four-gluon interactions, with the definition of the G term given by Eq. 1.16. The α_s factor is the *strong coupling constant*, scaling the overall strength of the interaction.

Taking α_s to be the QCD coupling constant is, however, misleading. This constant enters mathematical expressions corresponding to individual Feynman diagrams (one power of α_s for each vertex), but many such diagrams have the same initial and final states and must be summed together when computing the amplitude for the corresponding process. Moreover, diagrams with many vertices only contribute significantly when there is enough energy to distribute between all its branches, and so the number of diagrams that have to be considered in a computation is dependent on the momentum transfer Q of an interaction. This behavior can instead be absorbed in the definition of the coupling constant through the *renormalization* procedure [27], in which the sum of all diagrams with fixed couplings is replaced with a single diagram but with an energy dependent *running* coupling, as can be seen in Figure 1.2.

This behavior is exhibited by all of the forces in the Standard Model, but the dependence of α_s on the energy scale, shown in Figure 1.3, is a bit peculiar: as the energy scale increases, the strong interaction amplitude decreases – a behavior termed *asymptotic freedom*. In natural units, lengths are equivalent to Energy^{-1}, and so high energy scales are equivalent to small length scales. The implication for color interactions is that they are very weak over small length scales; for instance, colored particles do not interact much within the proton nucleus. Towards the low-energy limit, the behavior is reversed: as the energy (length) scale decreases (increases), the color interaction amplitude increases. Correspondingly, as colored particles radiate out of an interaction point, they tend to exchange gluons and sprout quark–antiquark pairs, and ultimately form colorless hadrons

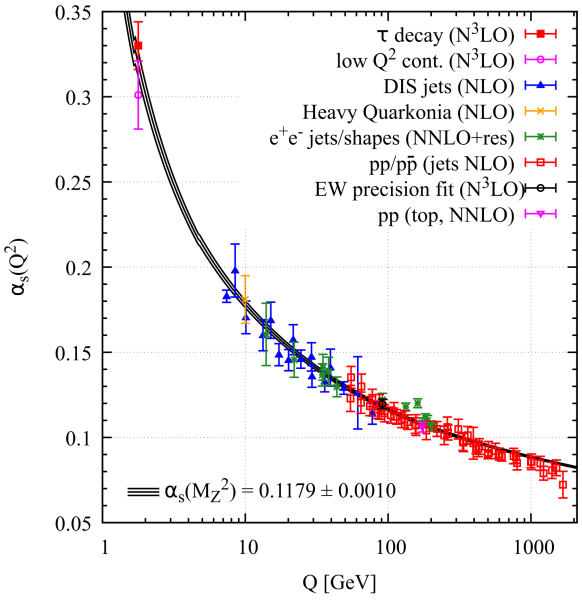


Figure 1.3 – Dependence of the strong coupling constant α_s on the energy scale of the interaction, exhibiting the peculiar behavior of the strong force known as asymptotic freedom: the strong force actually gets weaker and weaker as the energy scale of the interaction increases [28].

through the color confinement hypothesis; this process is known as hadronization and can be seen in Figure 1.4. The resulting collections of hadronic cascades, called *jets*, are hallmarks of hadronic colliders. Their phenomenology is a bit problematic however, since at energy scales characteristic of hadronization, α_s can become $\gtrapprox 1$ and consequently perturbative expansions cannot be used; instead, parameterized models such as the *Lund string model* [29] must be used. Nevertheless, predictions based on such models have been extensively studied at the LHC with satisfactory results; see, for example, Figure 1.5.

Another important consequence of confinement is that collisions between initial state hadrons such as protons can give rise to interactions between gluons and quarks or antiquarks beyond the valence quark content. In order to take this into account in theoretical computations, a non-perturbative part must be factored out of the computation and into experimentally-determined *parton distribution functions*, or PDFs, which describe the quark/gluon densities as a function of the parton's momentum fraction of the parton and the energy scale of the interaction; see Figure 1.6 for an example.

1.1.3 $SU(2) \times U(1)$: The electroweak model

At this point in our journey, we have two forces: the Strong force of QCD, or the $SU(3)$ part of the SM, and the electrical force of QED, which needs to be incorporated into the SM. By

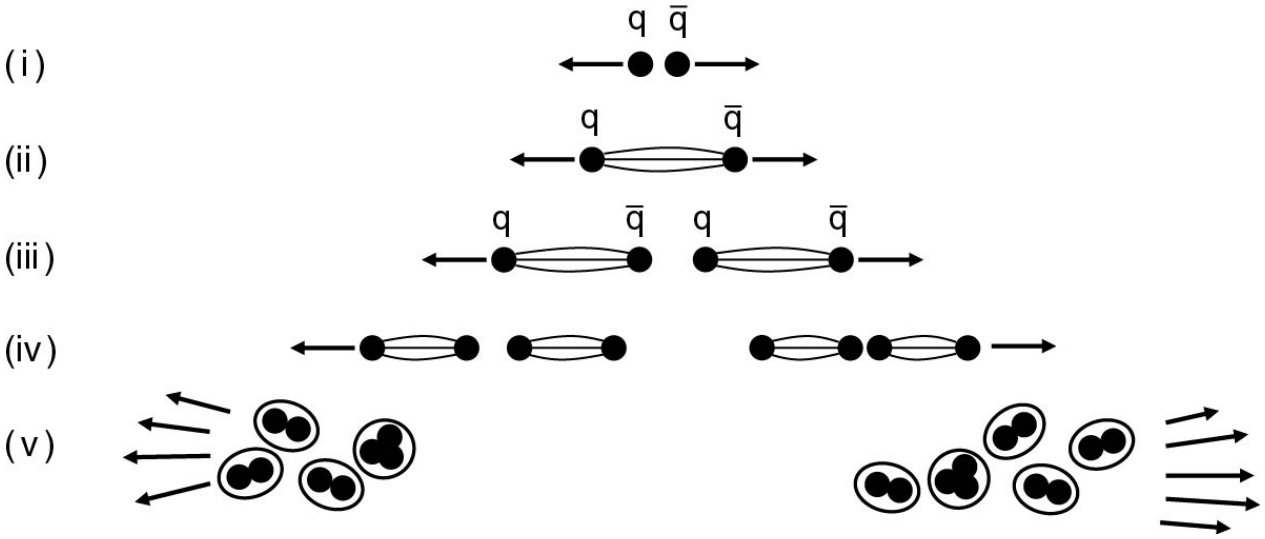
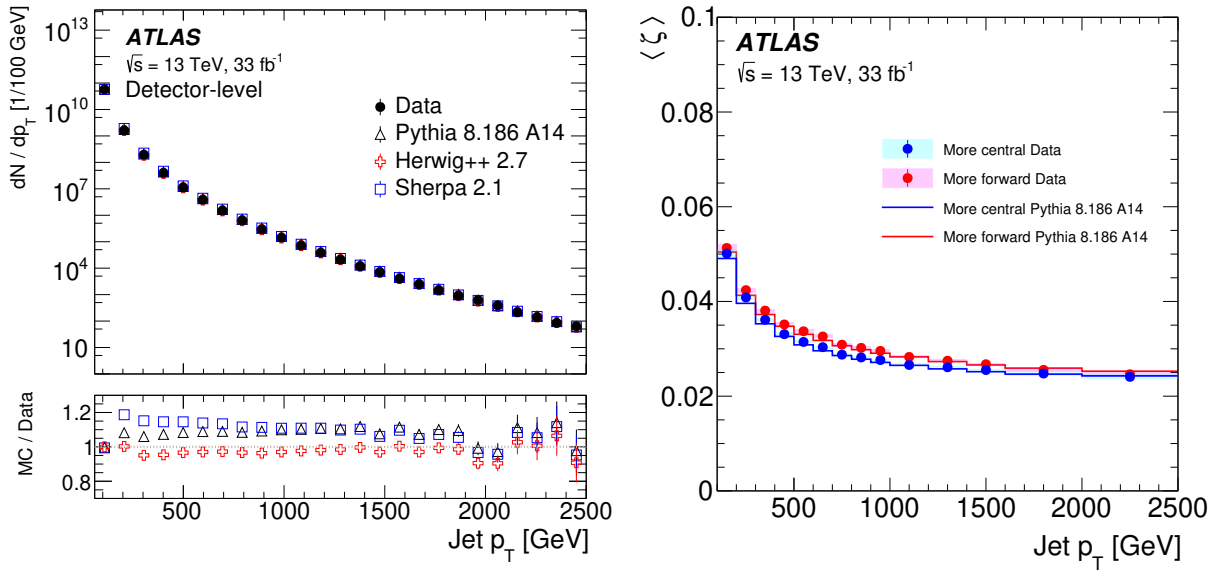


Figure 1.4 – Hadronization via gluon exchange [11].



(a) (b)

Figure 1.5 – (a) Comparison of the detector-level jet p_{T} spectrum from different generators and LHC data for an inclusive jet sample in the $|\eta| < 2.1$ region. (b) Average transverse momentum fraction $\zeta = p_{\mathrm{T}}^{\mathrm{particle}}/p_{\mathrm{T}}^{\mathrm{jet}}$, after unfolding to particle-level. Both plots show a good agreement between data and simulation, confirming the soundness of the hadronization model [30].

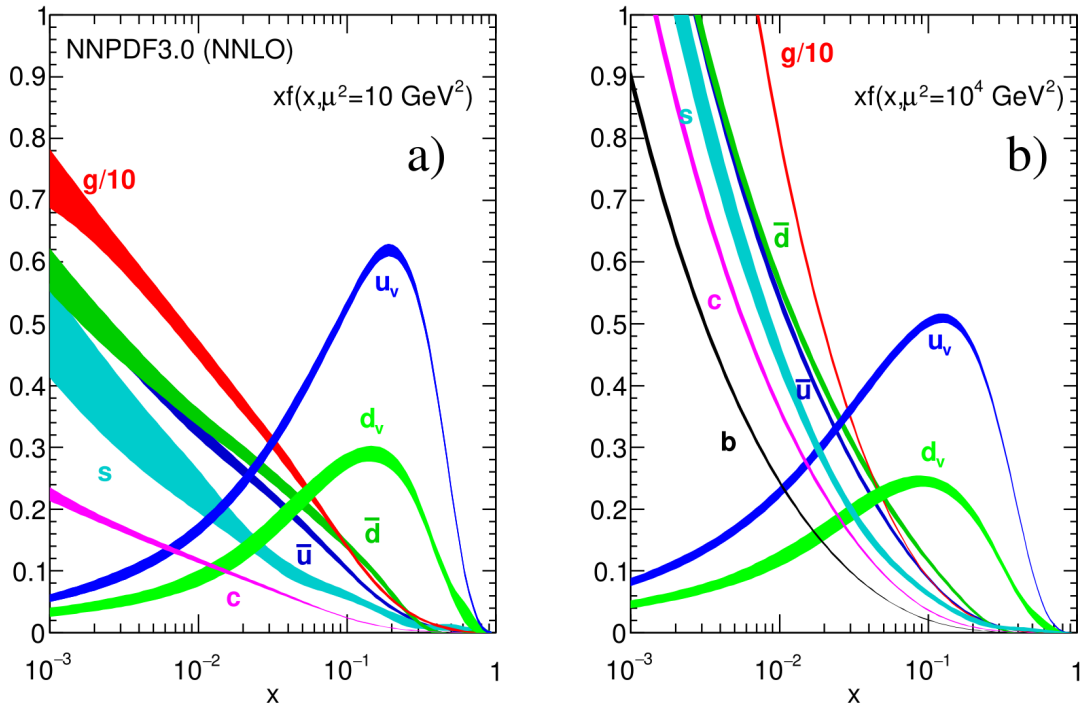


Figure 1.6 – Graphical representation of the `NNPDF3.0` proton parton distribution function [31], measured at two different energy scales (μ^2), as a function of an individual parton's fraction of the overall proton momentum (x) [28]. Interestingly, the valance quark content of the proton (two up-quarks and one down-quark) only dominates at large x, while the gluon is by far the most abundant at low x.

construction, QCD and QED are remarkably similar; pretty much the same procedure was followed in both cases, with two different gauge groups. In these two models, interaction vertices enter amplitude computations as *vector* currents proportional to $\overline{\psi}_1 \gamma^\mu \psi_2$. As a consequence, these interactions are *parity invariant*, meaning that they proceed in exactly the same way under a sign change of the spatial coordinates. This seems almost trivially true; and yet:

> *The physics community was stunned to learn in the 1950s that some events, unlike billiard ball collisions, follow different rules in their mirror-image versions. [32]*

This quotation refers to the 1957 experiment on β-decay performed by C.S. Wu *et al* [33], which showed that parity isn't always conserved. This discovery has a big implication for particle physics: since both QED and QCD are structurally constrained to be transparent to parity, then there must be another kind of interaction that doesn't have this limitation.

In order to stay as close as possible to our previous path, we still want to proceed via exchange of vector bosons, customarily called the Ws. To break parity invariance, the new interaction needs to have an *axial*-vector part, that is, $\propto \overline{\psi}_1 \gamma^\mu \gamma^5 \psi_2$, since axial-vectors do not change sign under parity transformations. Experimentally, the structure is determined to be "Vector minus Axial", or *V-A*:

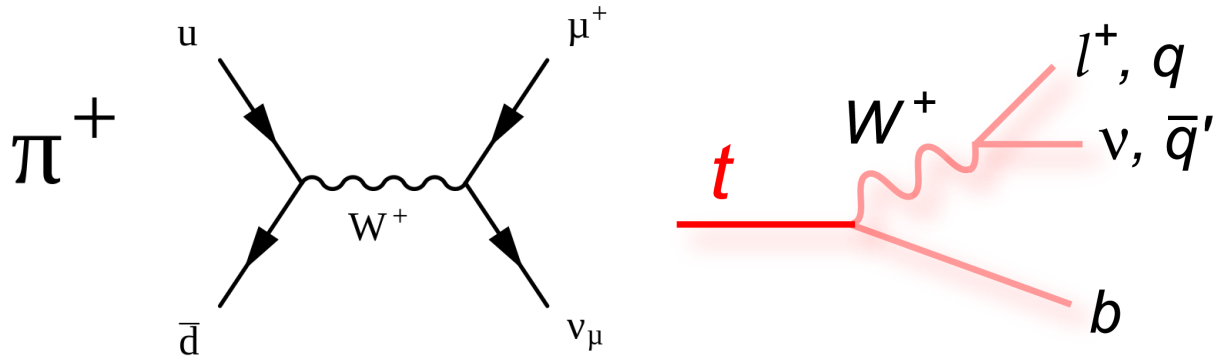
$$\gamma^\mu - \gamma^\mu \gamma^5 = \gamma^\mu (1 - \gamma^5). \tag{1.21}$$

In this expression, we find the *left-handed chirality projection* operator:

$$P_L \equiv \frac{1}{2}(1 - \gamma^5). \tag{1.22}$$

This means that the W boson will only couple to left-handed particles (or right-handed anti-particles): the Standard Model is a *chiral* theory.

The W is observed to couple together particles differing by a unit of electric charge; therefore, there are really two such bosons, the W^+ and $W^- \equiv W^\pm$. The particles they couple to are grouped in left-handed "doublets", ϕ_L. Members of a given doublet have the same value of *weak isospin*, I_W, the charge governing this interaction, and consequently the term "isospin doublet" is sometimes used. Doublets are formed either by lepton–neutrino pairs or by different-flavor quark pairs; it is the only SM interaction that allows such couplings, allowing phenomena such as charged pion decay to leptons and neutrinos, or top decay to a bottom quark, as seen in Figure 1.7.



(a) (b)

Figure 1.7 – (a) Charged pion and (b) top quark decays via the charged weak interaction.

Analogously to QED and QCD, we can derive the existence of the $W^{\pm}$ bosons by writing a Lagrangian for free isospin doublets and requiring local gauge invariance under the $SU(2)$ local gauge group. The derivation is very similar to the $SU(3)$ case; the main difference, of course, is that there are three generators instead of eight: W_1, W_2, W_3. Since these bosons act on the doublet ϕ_L, they don't correspond directly to the physical bosons, which couples to actual particles; the physical bosons are actually linear combinations of $W_{1,2,3}$ that pick out particles from the doublets. Following Thomson [11]:

$$\frac{1}{\sqrt{2}}(\sigma_1 + i\sigma_2) = \sqrt{2}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{1.23}$$

$$\frac{1}{\sqrt{2}}(\sigma_1 - i\sigma_2) = \sqrt{2}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{1.24}$$

And so the two corresponding currents will pick out particles from the doublets:

$$j^{\mu}_{+} \propto (\overline{\nu}_L \quad \overline{e}_L)\gamma^{\mu}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \nu_L \\ e_L \end{pmatrix} = \overline{\nu}_L\gamma^{\mu}e_L = \overline{\nu}\gamma^{\mu}P_L\, e, \tag{1.25}$$

$$j^{\mu}_{-} \propto (\overline{\nu}_L \quad \overline{e}_L)\gamma^{\mu}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} \nu_L \\ e_L \end{pmatrix} = \overline{e}_L\gamma^{\mu}\nu_L = \overline{e}\gamma^{\mu}P_L\, \nu. \tag{1.26}$$

This is exactly what is needed. Therefore, the physical W bosons are:

$$W^{\pm} = \frac{1}{\sqrt{2}}(W_1 \mp iW_2). \tag{1.27}$$

What about the W_3? It cannot account by itself for the experimentally-motivated neutral current since it needs to also couple to right-handed particles. The solution is simple:

$$SU(2) \rightarrow SU(2) \times U(1)_Y. \tag{1.28}$$

That is, we need yet another $U(1)$ gauge boson, the B_μ. It can't just be the QED photon since experiments pointed to a *massive* neutral boson (as for the $W^{\pm}$), hence the Y suffix appended to the gauge group in the equation below which stands for the *hypercharge*, the charge governing the neutral part of the electroweak model.

Amazingly, by including this boson, the need for QED as a standalone theory is completely removed. The price to pay for this *electroweak unification* is to introduce a new free parameter in the model, the *weak mixing angle*, θ_{W}. Then,

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_{\mathrm{W}} & \sin\theta_{\mathrm{W}} \\ -\sin\theta_{\mathrm{W}} & \cos\theta_{\mathrm{W}} \end{pmatrix} \begin{pmatrix} B_\mu \\ W_{3,\mu} \end{pmatrix}, \tag{1.29}$$

where A_μ is the photon and Z_μ is the required neutral weak boson.

There is, however, one thing that needs to be taken care of before calling it a day. We've mentioned how the weak gauge bosons are massive, and that they entered the theory via local gauge invariance, as they should. This poses a great problem: local gauge invariance *forbids* such massive force carriers. It even forbids fermion masses, and we know for sure that electrons are massive! Do we get rid of the local gauge principle, then? We don't need to; we simply add the missing piece, which takes us from a locally gauge invariant model to the real world : the *Higgs mechanism* [34–36].

In the Higgs mechanism, two scalar fields, one charged and one neutral, enter the theory in a single complex doublet ϕ_h. Analogously to what we did before, we start with the Lagrangian for a non-interacting Higgs:

$$\mathcal{L}_{\mathrm{h}} = (\partial^\mu \phi_{\mathrm{h}})(\partial_\mu \phi_{\mathrm{h}})^\dagger - V(\phi_{\mathrm{h}}). \tag{1.30}$$

It is customary to work in the *unitary gauge*, in which the charged Higgs and the imaginary

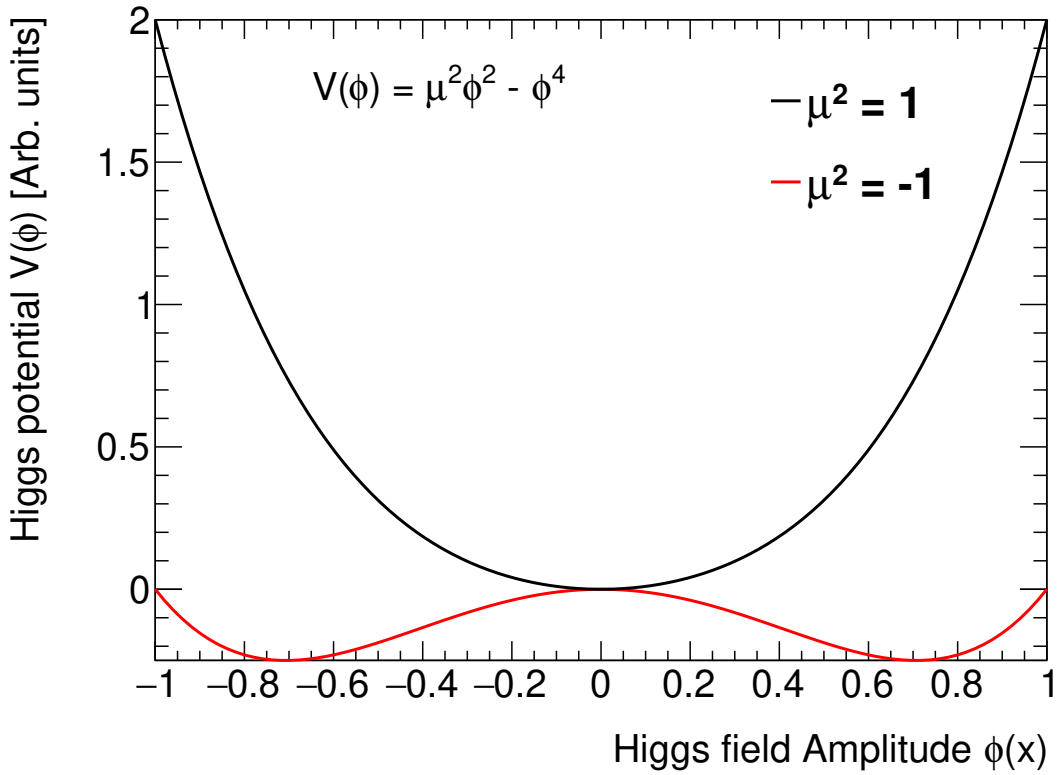


Figure 1.8 – Higgs field potential, for different μ^2 regimes. The $\mu^2 < 1$ case is realized in nature, and the resulting potential exhibits a so-called "Mexican hat" shape, which implies a non-zero vacuum expectation value of the Higgs field.

part of the neutral Higgs are absorbed as longitudinal polarization modes of the W and Z bosons:

$$\phi_{\mathrm{h}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \phi^0(x) \end{pmatrix}, \phi^0(x) \in \mathbb{R}. \tag{1.31}$$

The potential $V(\phi^0)$ has the form:

$$V(\phi^0) = \mu^2 \phi\phi^\dagger + \lambda(\phi\phi^\dagger)^2. \tag{1.32}$$

As seen in Figure 1.8, this potential potential has minima away from zero for the $\mu^2 < 0$ case, which is realized in nature:

$$\langle 0|\phi_h^0|0\rangle = v. \tag{1.33}$$

This motivates a change of variable of the Higgs as a perturbation away from this minimum:

$$\phi^0 = v + h(x). \tag{1.34}$$

Therefore, any particle that couples to the Higgs field then acquires a mass proportional to the vacuum expectation value. For a vector boson V^μ this would look like:

$$\frac{1}{2}\phi_h(x)V_\mu V^\mu = \frac{1}{2}vV_\mu V^\mu + \frac{1}{2}h(x)V_\mu V^\mu. \tag{1.35}$$

Indeed, the first term on the right is a kinetic energy term with mass $m_V = v$.

Let's see how this works out in our model. First, the $W_{1,2,3}$ and B_μ bosons are introduced in the Higgs Lagrangian such that the whole thing is locally gauge invariant. Using the *covariant derivative* notation:

$$\mathcal{L}_h \to (D^\mu\phi_h)(D_\mu\phi_h)^\dagger - V(\phi_h), \tag{1.36}$$

$$D^\mu \equiv \partial^\mu + ig_{\mathrm{w}}\frac{\sigma_i \cdot W_i^\mu}{2} + ig'\frac{Y}{2}B^\mu. \tag{1.37}$$

Expanding this Lagrangian, we get the following kinetic term for $W_{1,2}$:

$$\frac{v^2}{8}g_w^2\begin{pmatrix}W_1 & W_2\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}W_1\\ W_2\end{pmatrix} = \frac{1}{2}\begin{pmatrix}W^+ & W^-\end{pmatrix}\begin{pmatrix}m_W^2 & 0\\ 0 & m_W^2\end{pmatrix}\begin{pmatrix}W^+\\ W^-\end{pmatrix}. \tag{1.38}$$

And so:

$$m_W = \frac{1}{2}g_{\mathrm{w}}v. \tag{1.39}$$

A similar procedure for W_3 and B yields:

$$m_{\mathrm{Z}} = \frac{1}{2} v \sqrt{g_{\mathrm{w}}^2 + g'^2}. \tag{1.40}$$

Thus, through *electroweak symmetry breaking* by the Higgs mechanism, the W and Z bosons are permitted to have a mass, but the Lagrangian of Eq. 1.36 *is still locally gauge invariant*[2].

Comparing Equations 1.39 and 1.40:

$$m_{\mathrm{Z}} = m_{\mathrm{W}} \frac{\sqrt{g_{\mathrm{w}}^2 + g'^2}}{g_{\mathrm{w}}} \equiv \frac{m_{\mathrm{W}}}{\cos\theta_{\mathrm{W}}}. \tag{1.41}$$

Here, θ_{W} is the same "weak mixing angle" than that of Eq. 1.29; it allows the electromagnetic coupling and charge of QED to be expressed in terms of the electroweak parameters:

$$e = g_{\mathrm{w}} \sin\theta_{\mathrm{W}} = g' \cos\theta_{\mathrm{W}}, \tag{1.42}$$

$$q = \frac{Y}{2} + I_W. \tag{1.43}$$

Let's briefly recap. Left-handed fermions of the Standard Model were placed in left-handed doublets; requiring that these doublets' Lagrangians be invariant under $SU(2)$ local gauge transformation, we deduced the existence of the $W_{1,2,3}$ bosons. Correspondingly, the model gained a charge: weak isospin, I_w. Another gauge group, $U(1)_Y$, was used to imply the existence of the B_μ bosons, which mixes with the W_3 to give rise to the Z boson and the photon. Another charge appeared, the hypercharge Y. Local gauge invariance of the Higgs field under $SU(2) \times U(1)_Y$ was then required, which explains the origin of masses, and the breaking of the $SU(2) \times U(1)$ sector into weak charged, weak neutral, and electromagnetic interactions. This is the *Electroweak* model [38–40]; the physical Higgs boson was observed in 2012 by the ATLAS and CMS collaborations [2, 3], finally confirming its validity.

1.1.4 Summary: The complete Standard Model

We are now ready to write the full standard model Lagrangian. Schematically, we can break it down into the following constituent parts:

$$\mathcal{L}_{\mathrm{SM}} = \mathcal{L}_{\mathrm{matter}} + \mathcal{L}_{\mathrm{QCD}} + \mathcal{L}_{\mathrm{EWK}} + \mathcal{L}_{\mathrm{Higgs}}. \tag{1.44}$$

[2] *"The global minimum of the theory defines the ground state, and spontaneous symmetry breaking implies that there is a (global and/or local) symmetry of the system that is not respected by the ground state."* [37]

The first part, $\mathcal{L}_{\text{matter}}$, is the kinetic term for all Standard Model fermions:

$$\mathcal{L}_{\text{matter}} = \sum_{f \in \text{SM}} \overline{\psi}(i\partial\!\!\!/ - m_f)\psi. \tag{1.45}$$

Here, the sum runs over all fermions. Their masses are generated by interactions with the Higgs field, as is the case for the electroweak bosons; however, the SM does not provide prediction for their observed values and so they enter the model as free parameters constrained by experiment: three for the charged leptons (electron, e; muon, μ; tau, τ) and six for the quarks (up, u; down, d; strange, s; charm, c; bottom, b; top, t). The SM also includes an anti-particle for each fermion, differing by one unit of electric charge – both constituent of particle–antiparticle pairs have the same mass. Similarly, there are three color-differing copies of each quark with the same mass. In the vanilla Standard Model, the three neutrinos (ν_e, ν_μ and ν_τ) are assumed to be massless[3]. The SM fermions can be grouped into three generations ordered by increasing mass, separately for the leptons and the quarks:

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}, \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix}, \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix}, \tag{1.46}$$

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix}. \tag{1.47}$$

The $\mathcal{L}_{\text{QCD}}$ term describes color-changing interaction via gluon exchange:

$$\mathcal{L}_{\text{QCD}} = -\sum_{q,c,c',k} \overline{\psi}_{q,c} \alpha_s \lambda_{kcc'} A\!\!\!/_k \psi_{q,c'} - \frac{1}{4} G_{k,\mu\nu} G_k^{\mu\nu}. \tag{1.48}$$

The sum runs over quarks (q), color (c and c') and gluons (k). This term contains another free parameter: α_s, which governs the strength of the strong interaction; the gluon itself remains massless.

The $\mathcal{L}_{\text{EWK}}$ term represents the electroweak interaction before gauge symmetry breaking:

[3]Of course, they are known to be massive. The actual values are under-constrained, but they are known to be extremely small; therefore, in collider physics it is of no great consequence to ignore this complication.

$$\mathcal{L}_{\text{EWK}} = -\frac{g_{\text{w}}}{2}\sum_{\phi_L}\overline{\phi}_L(\sigma_{1,2,3}\cdot W_{1,2,3;\mu})\phi_L - \frac{g'}{2}\sum_f \overline{\psi}_f(Y_W B_\mu)\psi_f. \tag{1.49}$$

Here we have the remaining two free parameters associated to gauge couplings, g_{w} and g'. The part involving the W bosons couple together "left-handed doublets" and is the chiral part of the model. Flavor-changing interactions involving quark doublets are under-constrained, and four additional parameters are needed to describe them through the CKM matrix [41, 42].

Finally, we have the Higgs sector:

$$\mathcal{L}_{\text{Higgs}} = (D\phi)(D\phi)^\dagger - \mu^2\phi\phi^\dagger + \lambda(\phi\phi^\dagger)^2, \tag{1.50}$$

$$D \equiv \partial + ig_{\text{w}}\frac{\sigma_i\cdot W_i}{2} + ig'\frac{Y}{2}B_\mu. \tag{1.51}$$

We find here the remaining SM parameters – The Higgs parameter, μ^2, which sets the "bare" (non-renormalized) mass of the Higgs, $m_h^{(0)} = \sqrt{2\mu^2}$. The observed mass of the Higgs, which differs from the bare value due to higher-order loop diagrams, is our last parameter of interest, m_h.

So there we have it, the Standard Model in all its glory: 15 fermions, eight gluons, three electroweak gauge bosons and one Higgs boson, interacting together through the chromodynamic, electroweak, and Higgs sectors. Its gauge-theory based theoretical structure is parameterized by 18 free variables[4]. This theory has passed a large number of experimental tests with great success, as can be seen for example in Figure 1.9.

1.2 Problems with the Standard Model

In the previous few sections, we've seen how the Standard Model is a theoretically sophisticated and experimentally successful theory of nature. However, it is not perfect; there are a some problems that still subsist. To name a few:

- The energy gap between the Plank and weak scales, M_P and M_W, is very large, which

[4]There is an additional parameter, the "QCD vacuum angle" (θ_{QCD}), which in principle allows for CP-violation from QCD interactions; this parameter is experimentally constrained to be ≈ 0. There are also further parameters describing non-zero neutrino masses (grouped into the the PMNS matrix), which do not arise out-of-the-box in the Standard Model but are needed to explain phenomena such as neutrino oscillations. These topics are out of the scope of this thesis.

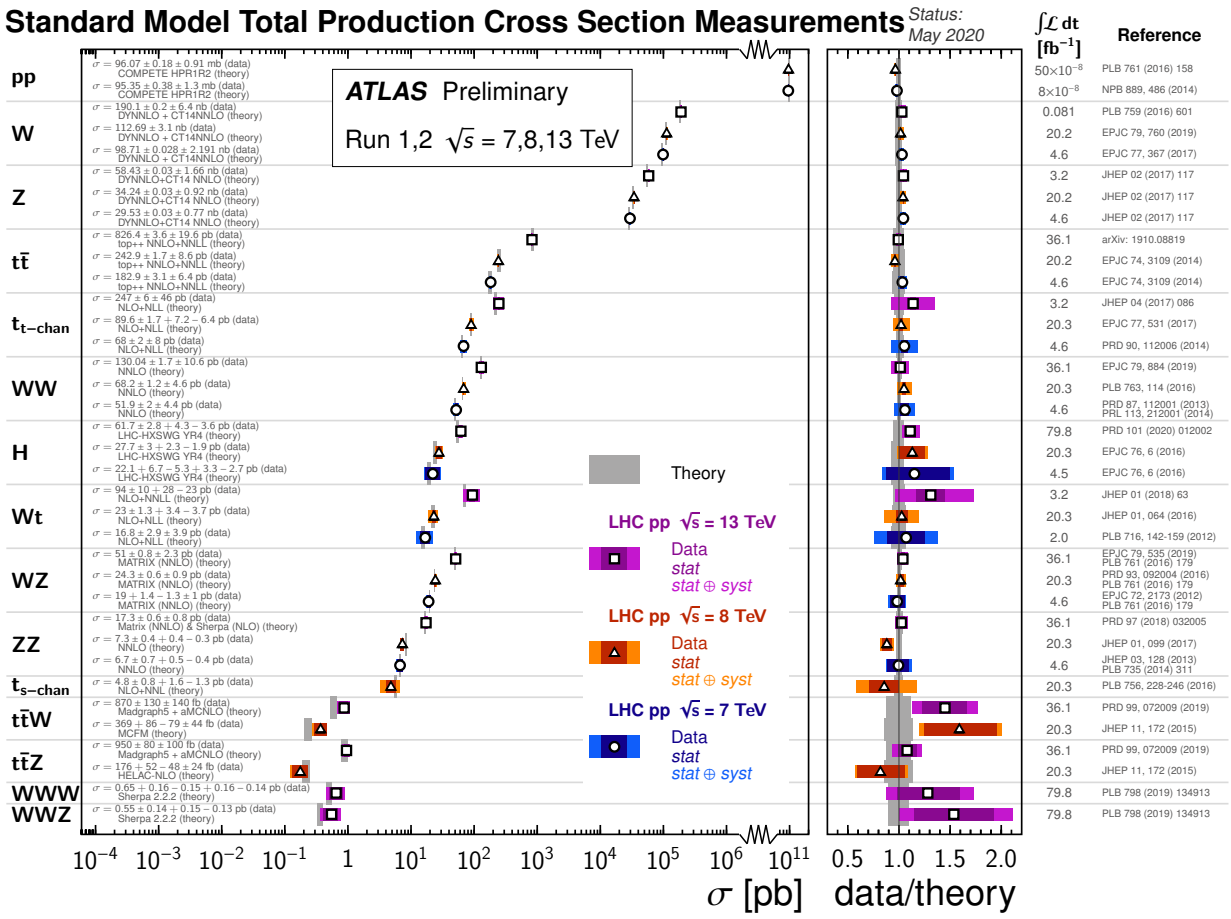


Figure 1.9 – Summary of several ATLAS Standard Model total production cross section measurements, corrected for branching fractions, compared to the corresponding theoretical expectations and ratio with respect to best prediction. [43].

implies an unnatural fine-tuning of the Higgs mass;

- Astronomical observations of *Dark Matter* and *Dark Energy* are not accounted for;
- Gravity is not accounted for;
- The theory is not completely unified, as the electroweak model and QCD require separate descriptions;
- The matter–antimatter asymmetry observed in our universe require additional sources of CP violation beyond the SM mechanisms;
- Neutrino masses are not accounted for;
- The energy gap between the SM vacuum expectation value and the observed cosmological constant Λ_0 is extremely large, implying an unnatural fine-tuning of Λ_0 to 120 orders of magnitude.

For this thesis however, we only need to look into the first two problems in details. The hierarchy problem and the closely related Higgs fine-tuning problem and numerical naturalness principle are discussed in Section 1.2.1, and the subject of Dark Matter is treated in Section 1.2.2.

1.2.1 Naturalness

Let's first define the concept of *numerical naturalness*, as it relates the the construction of physical theories. In his 1985 paper "Naturalness in theoretical physics" [44], Philip Nelson states the problem:

> *" (...) we have a strong naturalness problem whenever the set of theories which even remotely resemble our world is a tiny subset of all the acceptable theories. We must cure the problem by slicing the latter class down to size. This entails finding some new principle which renders most of its members unacceptable, leaving only a few — including of course at least one of the desired theories. In this way, theorists often permit the introduction of new structures into their theories, even when they are not strictly called for by observation. "*

For example, remember the QCD model introduced in Section 1.1.2. We could have just started by positing the existence of the eight gluons and not bother with local gauge symmetry – this principle has no direct phenomenological consequence after the appearance of the gauge fields and the subsequent gauge symmetry breaking. However, this leaves us

with a little too much freedom: the resulting equations still need to have the same exact form for the theory to match observations, but there is no underlying principle to help us understand why these are the right equations. Different versions of these equations, with different numbers of gluons for example, might lead to universes very different from our own, which means that we live in a very specific realization. Why this one instead of another? Requiring that QCD be an $SU(3)$ local gauge theory allowed us to start with less assumptions, and automatically exclude a lot of competing theories, even if this is not strictly needed after we have obtained the Lagrangian. It makes the theory more natural in the sense that the appearance of exactly the right eight gluons is *unavoidable* once we require invariance under the right gauge group.

However, there is at least one area of the model that suffers from a naturalness problem. It is related to what is known as the *hierarchy problem* [45–47], that is, the extremely large energy gap between the Planck and weak energy scales. Since the Standard Model is an effective theory, it is only valid up to its ultraviolet cutoff, Λ_{UV}. Above this energy, there needs to be a new theory in which the SM is embedded; However, as Stephen Martin explains, the Higgs potential exhibits "a disturbing sensitivity to new physics in almost any imaginable extension of the Standard Model" [48].

Let's make this more concrete. The expression of the Higgs mass can be schematically broken into two parts, corresponding to the bare mass and the radiative corrections due to the higher-order diagrams [49]:

$$m_{\mathrm{h}}^2 = m_{\mathrm{h},0}^2 + \kappa \delta m_{\mathrm{h}}^2 \tag{1.52}$$

At one-loop level, the radiative contribution to the Higgs boson mass from a massive fermion has the following form [48]:

$$\delta m_{\mathrm{h}}^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{UV}^2 + O\left(m_f^2 \log \frac{\Lambda_{\mathrm{UV}}}{m_f}\right), \tag{1.53}$$

where λ_f is the strength of the Yukawa coupling between the fermion and the Higgs.

Gravity, as described rather successfully by general relativity [50], is not compatible with the Standard Model [51]. This means that Λ_{UV} is at most in the vicinity of the reduced Planck mass, $M_p \approx 10^{18}$ GeV[5],at which gravitational effects become non-negligible. Consequently, based solely on Eq. 1.53 we should *a priori* expect the Higgs mass to be

[5]In natural units, $c = \hbar = 1$, and therefore masses are quantified in units of energy.

proportional to the reduced Planck mass:

$$m_H \propto M_p. \tag{1.54}$$

However, the observed Higgs boson mass is at the level of the weak scale, *a posteriori* leading us to posit:

$$m_H \propto G_F^{-1/2}, \tag{1.55}$$

where G_F is the Fermi constant, which quantifies the electroweak scale. If equation 1.54 is the right relation, the constant of proportionality (κ in Eq. 1.52) has to be incredibly small, in the order of $(\sqrt{G_F}M_P)^{-1} \approx 10^{-16}$! *Why this one, instead of another*[6]?.

There are a few proposed solutions to this problem, the first of which is basically to ignore it. In that case, we simply accept that the low- and high-energy sectors of the SM do not decouple when computing the Higgs mass, in contrast to the rest of the model, which is unaffected by ultraviolet parameters [52, 53]. However, a successful theory would still need to explain why that is, or equivalently why is the $G_F^{-1/2}/M_P$ ratio so large [15]. For example, the *anthropic principle* states that if the ratio was anything other than something close to 10^{-16}, it would be very unlikely for the universe to sustain life, at least in a form that we can recognize; the equation has to be exactly this way for us to even exist and write it down. To side-step problem of naturalness, we can posit the existence of many different parallel universe, contained in a *multiverse*. Each of these universes is allowed to have different values of fundamental constants, and by definition we happen to live in a universe that has the right conditions for emergence of large-scale structures and complex life [49].

Actually fixing the problem comes a hefty price: new physics has to be introduced somewhere between the weak and the Planck scale in order to somehow tame the Λ_{UV}^2 term. Let's make this explicit and assume that there exists a mechanism to cancel it out. What about the remaining $\mathcal{O}\left(m_f^2 \log(\Lambda_{\mathrm{UV}}/m_f)\right)$ contribution from Equation 1.53? Because of the m_f^2 dependency, the biggest contribution of this form comes from the top quark whose mass, like the Higgs, is at the level of the weak scale. The induced correction

[6]In the words of Gian Giudice: "*Just to get a feeling of the level of parameter tuning required, let me make a simple analogy. Balancing on a table a pencil on its tip is a subtle art that requires patience and a steady hand. It is a matter of fine tuning the position of the pencil such that its center of mass falls within the surface of its tip. If R is the length of the pencil and r the radius of the tip surface, the needed accuracy is of the order of r^2/R^2 . Let us now compare this with the fine tuning in κ. The necessary accuracy to reproduce $[(\sqrt{G_F}M_P)^{-1}]$ is equal to the accuracy needed to balance a pencil as long as the solar system on a tip a millimeter wide! (. . .) This has led to a widespread belief among particle physicists that such an apparently fantastic coincidence must have some hidden reason.*" [49].

is, roughly,

$$\frac{\delta m_h^2}{m_h} \propto \frac{m_t^2}{m_h} \log \frac{\Lambda_{\mathrm{UV}}}{m_t}. \tag{1.56}$$

With m_t and m_h at the weak scale $\approx 10^2$ GeV and $\Lambda_{\mathrm{UV}} \approx 10^{18}$ GeV, a simple Fermi estimate[7] yields:

$$\frac{\delta m_h}{m_h} \approx 10. \tag{1.57}$$

In other words, *if we can make the* $O(\Lambda_{UV}^2)$ *term disappear,* the observed mass of the Higgs boson is approximately within an order of magnitude from what is expected from the mass computation, which would mean that the observed relationship equation 1.55 is the right one.

What can be said about the mechanisms that could solves the Λ_{UV}^2 problem? One solution would be to push down Λ_{UV} nearer to the weak scale; this is what happens in extra-dimensions models, in which the characteristic energy scale of gravity is O(TeV) but appears much larger because it is diffused in these extra dimensions, which are invisible to the other forces [54]. There are also hypothetical models of *composite Higgs* [55] in which the Higgs mass is dynamically generated by its constituents and no fine-tuning problem arises.

Another, perhaps more elegant solution is to protect the Higgs mass from large higher-order corrections by introducing a new symmetry. The idea is not far-fetched; for instance, large corrections to fermion masses via self-energy diagrams involving photons are forbidden by the chiral symmetry of QED [52]. For now, let's just note that the contribution to the Higgs mass from a new heavy scalar particle S would take the form [48]:

$$\delta m_h^2 = \frac{\lambda_S}{16\pi^2} \Lambda_{UV}^2 + O\left(m_S^2 \log \frac{\Lambda_{\mathrm{UV}}}{m_S} \right). \tag{1.58}$$

Notice the sign difference relative to equation 1.53:

$$\delta m_h^2 = -\frac{|\lambda_f|^2}{8\pi^2} \Lambda_{UV}^2 + O\left(m_f^2 \log \frac{\Lambda_{\mathrm{UV}}}{m_f} \right), \tag{1.59}$$

As we shall see in Chapter 2, such expressions arise in the context of supersymmetric extensions of the Standard Model.

[7] https://en.wikipedia.org/wiki/Fermi_problem

1.2.2 Dark Matter

Dark Matter is a term encompassing a family of hypothetical massive particles whose presence in the universe is hinted at by many supporting observations that show a large amount of massive, non-luminous matter that is unaccounted for in the Standard Model.

Galactic rotation curves

One of the better known supporting evidence for this "missing" matter is found in the study of the orbital velocity of light-emitting bodies in spiral galaxies. Indeed, there is a relationship between the rotational speed of an object at a distance r from the galactic center and the galaxy's mass distribution $m(r)$ due to Kepler's third law:

$$v(r) \propto \sqrt{\frac{m(r)}{r}}. \tag{1.60}$$

From observation of light emitted from galaxies and through the use of mass-to-luminosity ratios, it was long thought that most of the spiral galaxies's mass is distributed near their center. Within this region, the total mass enclosed within an orbit grows with the volume it encloses while outside of this massive center, the remaining mass due to luminous matter can be neglected and the enclosed mass is approximately constant [56]. Thus:

$$v(r) \propto \begin{cases} r & \text{near the galaxy's center,} \\ 1/\sqrt{r} & \text{out of the center.} \end{cases} \tag{1.61}$$

The advent of radio-wave astronomy in the second part of the 20th century allowed scientists to measure these rotational curves by studying the Doppler shifts of electromagnetic radiation emitted by stars and gas at well-defined wavelengths. The results from such observations are at odds with the behavior prescribed by Eq. (1.61): the measured rotation curves show a roughly constant velocity out of the galaxy's center as seen in Figure 1.10. This implies that a large portion of the mass of spiral galaxies is due to non-luminous, "dark" matter [57, 58].

Gravitational lensing and the bullet cluster

The phenomenon of gravitational lensing is without a doubt one of the most startling implications of general relativity; it states that even though light is massless, its path is

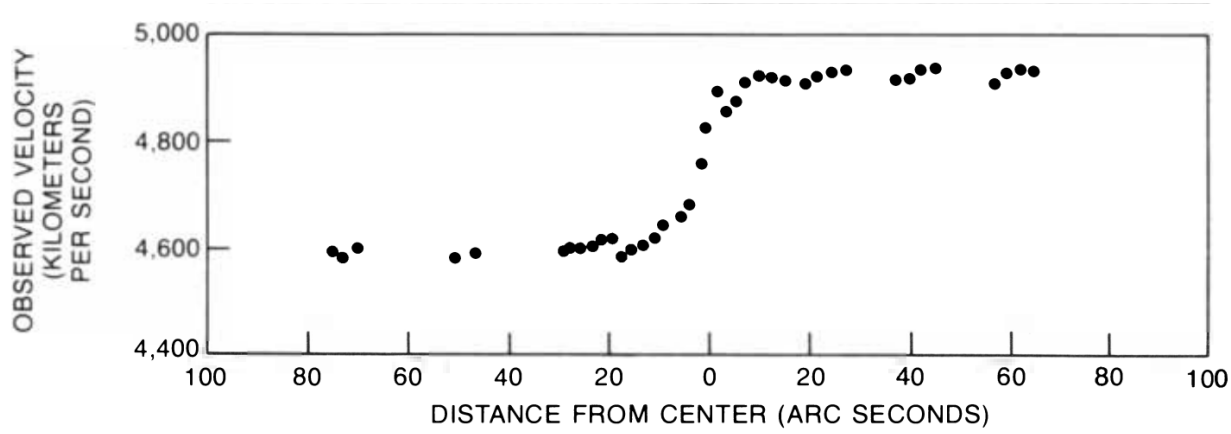


Figure 1.10 – Rotational curve for the NGC 2998 galaxy in Ursa Major, measured from its hydrogen-alpha emission spectrum [59].

still affected and can be bent by gravity. One important implication is that by analyzing data from distant light sources from the universe, one can infer the mass distribution in between the sources and the observation point. Such techniques have been used to study the *Bullet cluster*, which consists of two merging (or colliding) clusters of galaxies. As seen in Figure 1.11, the luminous matter distribution is peaked at the center of the system, roughly at the collision point, while the mass distribution inferred from gravitational lensing has two peaks on either sides of the collision area. These non-luminous matter peaks are thought to be a strong hint of presence of the dark matter in the colliding clusters; since dark matter does not interact strongly or electromagnetically, it is more or less free to pass straight through the collision center, as opposed to luminous matter that is seen to stay in the center of the system [60].

The Cosmic Microwave Background and the The Λ_{CDM} model

Another piece of evidence in favor of dark matter comes from measurements of the anisotropy of the cosmic microwave background (CMB). This microwave background originates from the the early universe, at a time at which its density had decreased enough through expansion that photons started escaping from the plasma originating from the Big Bang – the era of *last scattering*. Since photons were previously confined to areas of high density, by measuring temperature fluctuations in the CMB we obtain a map of the baryonic structures present in the early universe. In practice, the scales of these structures is inferred by measuring the temperature of the CMB in a multipole expansion. In a universe only containing baryonic matter, the energy at higher multipole moments tends to be diffused, since electromagnetic interactions before the recombination of baryonic matter

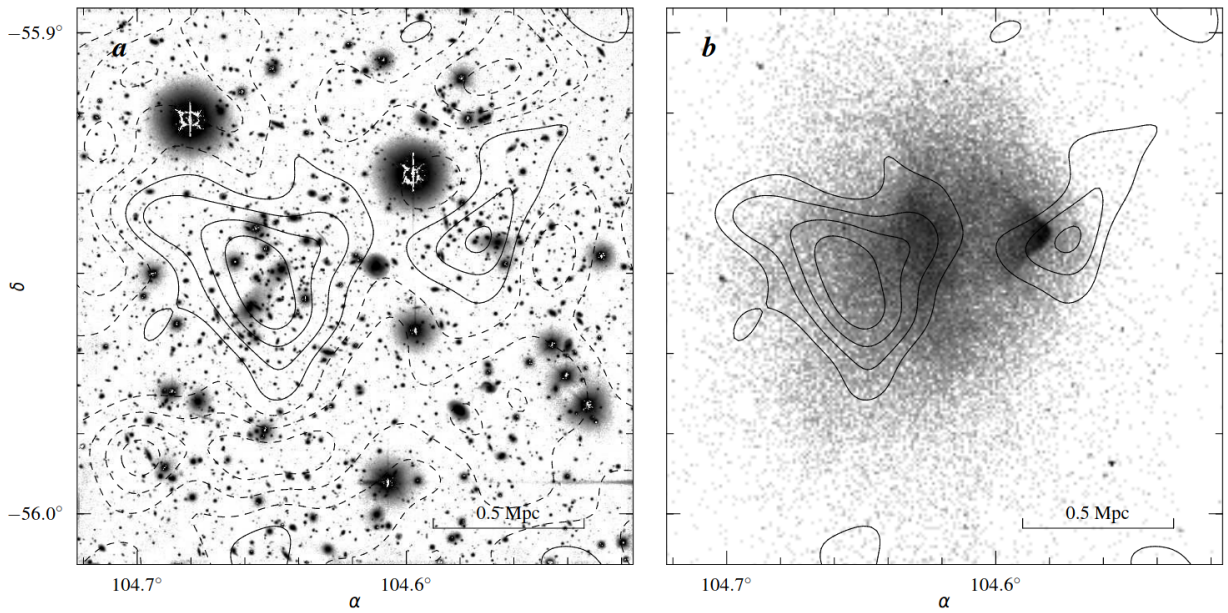


Figure 1.11 – Overlay of the mass contours inferred from graviational lensing and (a) a map of visible light and (b) a map of x-ray spectrum radiation from the Bullet cluster, showing that the mass distribution has peaks on either sides of the center [60].

in neutral atoms exerts too much outward pressure at small length scales. However, measures of the CMB by various satellites show clear peaks at these higher-order moments, which indicates presence of neutral ("dark") massive particles that allowed formation of such structures by gravitational pull [56], as seen in Fig. 1.12a.

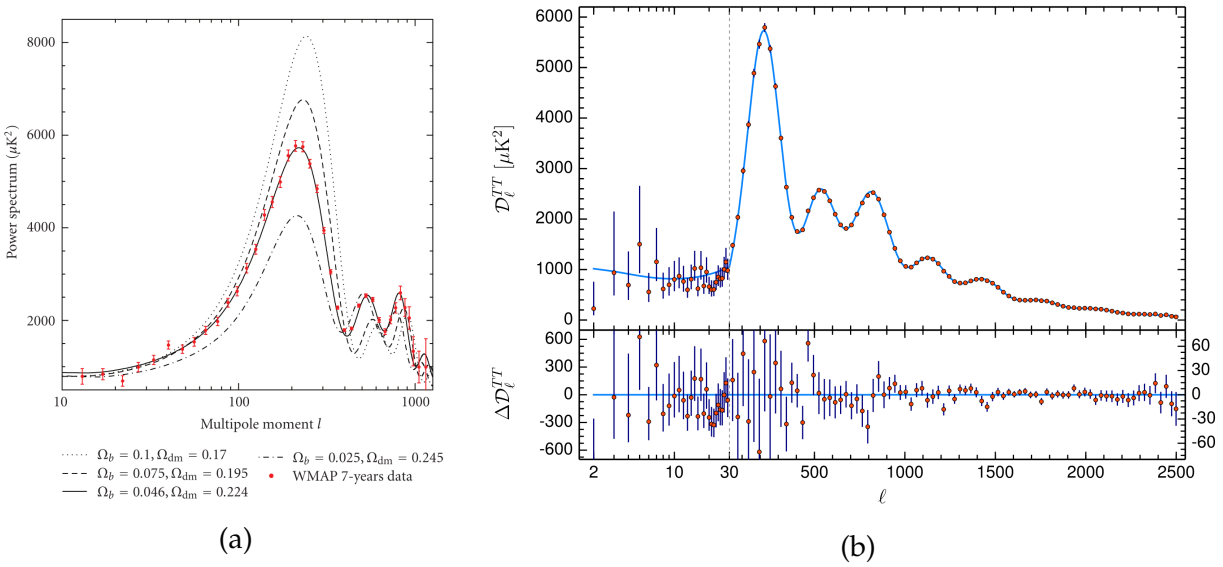
This is epitomized in the Λ_{CDM} model, sometimes called the "Standard Model of cosmology". This model describes very successfully the structure of the CMB by including non-charged matter with velocity $\ll c$; hence the name of "Cold Dark Matter"[8]. Surprisingly, according to Λ_{CDM}, baryonic matter accounts for about only 26.6% of all matter in the universe, the rest being made up of cold dark matter for which science has yet to experimentally observe a live specimen [63].

WIMP dark matter candidates

The prime candidate for particle dark matter[9] is a "WIMP", or "Weakly Interacting Massive Particle" [66]. Naively, it may seem that the SM neutrinos would be good such candidates:

[8]Models allowing relativistic dark matter generally fail to correctly describe how matter clusters at different scales in the universe.

[9]There are other possible ways of accounting for observations of dark matter. An obvious solution would be that of modifying the mathematical description gravity; this is the "MOdified Newtonian Dynamics" (MOND) solution [64]. However, such models have historically been much less experimentally successful than the Λ_{CDM} model, which assumes existence of particle dark matter [65].



(a) (b)

Figure 1.12 – (a) WMAP seven-year data [61] overlaid with various predictions showing how the baryon and dark matter fractions, Ω_b and Ω_{dm}, affect the CMB anisotropy [56]. (b) Power spectrum measured from the Planck 2018 results, which are used to compute the current values of Ω_b and Ω_{dm} [62].

they only interact weakly, and their masses, while still under-constrained by experiments, are known to be non-zero. However the very small upper bound on the neutrino masses implies that they are always relativistic and thus not viable *cold* dark matter candidates. Thus, a WIMP solution to the dark matter problem implies beyond-the-Standard-Model (BSM) physics. As we will see in Section 2.1.1, such particles naturally arise in just the right amount in many R-parity conserving supersymmetric extensions of the SM.

1.3 Conclusion

In this foundational chapter, we have first seen how the Standard Model arises in a surprisingly simple and elegant manner from its $SU(3) \times SU(2) \times U(1)$ local gauge structure. Conceptually, it contains two sub-theories, Chromodynamics and the Electroweak model, which have both been extensively validated by experiments. Nonetheless, we've also seen that some important problems remain, and we've discussed in some details two of them: the hierarchy and the dark matter problems. In the next chapter, we'll review a well-known framework that could be used to extend the Standard Model and solve the two aforementioned problems: supersymmetry.

Chapter 2

Supersymmetry

Supersymmetry (SUSY) [67–72] is an extension of the space-time symmetries that transforms bosons into fermions, and vice-versa. The idea of such an operation originated in the late 60s and early 70s independently in the Western world and in the Soviet Union [73] but, as seen in Figure 2.1, the idea really took off in the early 80s, concurrently with the realization that this new kind of symmetry could solve the naturalness problem of the Higgs sector [74, 75], discussed in Section 1.2.1.

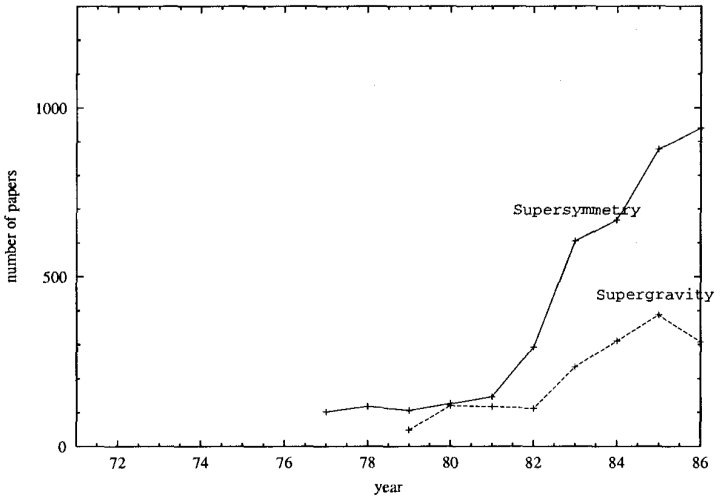


Figure 2.1 – Number of published papers about supersymmetry and supergravity (a locally-invariant version of SUSY) per year in the 70s and 80s [73].

We start with a theoretical overview in Sec. 2.1, motivating the search for supersymmetric partners of the gluon and third generation quarks, followed by a status review of the experimental search for such particles in Sec. 2.2.

2.1 Theoretical overview

Note: unless otherwise noted, this section is based on References [48, 52, 76–78]

Supersymmetry is not an actual theory, but simply the idea of a symmetry operation relating bosons and fermions. To obtain a properly supersymmetric theory, we must construct a Lagrangian in which this symmetry is manifest.

Analogously to the chiral multiplets introduced when discussing the electroweak model (Section 1.1.3), particles related by a supersymmetric transformation are arranged in *supermultiplets*. These must possess the following properties, among others:

1. Within a supermultiplet, particles must have the same basic properties except for spins, which must differ by 1/2 between the fermions and the bosons;

2. When a spin-1 boson and a spin-1/2 fermion are together in a supermultiplet, both chiral components of the fermion transform under the same gauge group as the boson;

3. Within a supermultiplet, the number of fermionic and bosonic degrees of freedom must be equal.

The first property has an important consequence: it effectively forbids supermultiplets populated by SM particles only. Therefore, SUSY requires BSM physics. The second and third properties allow us to decide how to place SM particles in supermultiplets – let's begin with the case of SM fermions. Since they exhibit chiral behavior, by property two they can only form multiplets with scalar bosons. Again, since they are chiral objects, by property three the simplest possible supermultiplet involving an SM fermion additionally contains two scalar fields (or a single complex field) and is called a *chiral supermultiplet*.

A single free chiral supermultiplet consisting of a fermion, ψ, and two scalar bosons, ϕ_1 and ϕ_2, is already enough to build a basic supersymmetric theory, called the Wess-Zumino model [69][1]:

$$\mathcal{L}_{\mathrm{WZ}} = \frac{1}{2}\left(i\overline{\psi}\partial\!\!\!/\psi + (\partial_\mu\phi_1)^2 + (\partial_\mu\phi_2)^2\right). \tag{2.1}$$

A supersymmetric transformation mixes the scalar bosons and fermions:

[1] To simplify, here we are ignoring the complications due to "auxiliary fields", which do not correspond to physical degrees of freedom.

$$\psi \rightarrow \psi + \partial\!\!\!/\phi_1\alpha + i\partial\!\!\!/\phi_2\alpha, \tag{2.2}$$

$$\phi_1 \rightarrow \phi_1 + i\overline{\alpha}\gamma^5\psi, \tag{2.3}$$

$$\phi_2 \rightarrow \phi_2 - \overline{\alpha}\psi, \tag{2.4}$$

where α is a spacetime-independent anti-commuting spinor that ensures that various terms have consistent dimensionality and which parameterizes the transformation. The Lagrangian of Eq. 2.1 gains an extra term, which fortunately disappears after applying the Euler-Lagrange equation. Therefore, the theory is invariant under this specific supersymmetric transformation.

Standard Model vector bosons are placed in supermultiplets with their spin-$1/2$ superpartners[2], called *gauginos*, to obtain *vector supermultiplets*. An example of a supersymmetric Lagrangian involving a such a supermultiplet would be

$$\mathcal{L}_v = -\frac{1}{4}F^a_{\mu\nu}F^{a,\mu\nu} + i\lambda^{a,\dagger}\overline{\sigma}^\mu(\partial_\mu\lambda^a + g\sum_{b,c} f^{abc}A^b_\mu\lambda^c), \tag{2.5}$$

where A_μ is the vector boson with kinetic term $F_{\mu\nu}$, λ is the gaugino, σ^μ is an anti-commuting spinor (again to ensure the right dimensions between the different terms) and the a, b and c indices run over all the generators of the appropriate gauge group. A supersymmetric transformation for this supermultiplet mixes the vector bosons and the fermions:

$$A^a_\mu \rightarrow A^a_\mu - \frac{1}{\sqrt{2}}(\epsilon^\dagger\overline{\sigma}_\mu\lambda^a + \lambda^{a,\dagger}\overline{\sigma}_\mu\epsilon), \tag{2.6}$$

$$\lambda^a \rightarrow \lambda^a + \frac{i}{2\sqrt{2}}(\sigma^\mu\overline{\sigma}^\nu\epsilon)F^a_{\mu\nu}. \tag{2.7}$$

The ϵ factor parametrizes the transformation. Both transformation "mix" the SM bosons and the gauginos, and again we have a supersymmetric Lagrangian.

[2]The spin=$3/2$ case makes the theory non-renormalizable

2.1.1 The Minimal Supersymmetric Standard Model

We've just seen how it's possible to create supersymmetric Lagrangians by placing an SM fermion in a chiral supermultiplet with two new scalar fields, and by placing an SM boson into a vector supermultiplet with a fermionic gaugino. Moreover, the supersymmetric Lagrangians of Equations 2.1 and 2.5 are built from the same components as the full Standard Model (Section 1.1); therefore, we only need to construct chiral and vector supermultiplets involving SM particles to obtain the Minimal Supersymmetric Standard Model, or MSSM [79, 80].

As explained in the previous section, each Standard Model fermion is put into a supermultiplet with two scalar fields. These supersymmetric particles are named by prefixing the name of their SM partner with the letter "s" (for "scalar"), and are symbolized by a tilde. For examples, the superpartners of the top (t) and bottom (b) quarks are the stops ($\tilde{t}_R$, $\tilde{t}_L$) and the sbottoms ($\tilde{b}_R$, $\tilde{b}_L$), respectively. Each chiral component gets its own superpartner, but in general the mass eigenstates will be linear combinations; these are denoted by a numeric index ordered by mass (for example, the $\tilde{t}_R$ and $\tilde{t}_L$ can mix to form $\tilde{t}_1$ and $\tilde{t}_2$, with the first one being lightest).

The supersymmetric partner of an SM boson is named by appending the suffix "-ino" to their partner's name. Therefore, the gluon is associated to the *gluino* ($\tilde{g}$). The situation is a little more complicated for the electroweak sector: the gauge bosons are associated to the *gauginos*, namely, the Winos ($\tilde{W}_{1,2,3}$) and the Bino ($\tilde{B}$); the Higgs sector has to be enlarged to avoid gauge anomalies [81], and consist of two complex Higgs doublets and their superpartners, the *Higgsinos*. As a consequence of $SU(2) \times U(1)$ symmetry breaking, the Higgsinos and gauginos mix into charged and neutral states, the *charginos* $\tilde{\chi}_i^{\pm}$ and *neutralinos* $\tilde{\chi}_i^0$, whose indices are also mass-ordered.

In total, the MSSM adds a whopping 105 free parameters to the 18 parameters of the Standard Model: 36 parameters expressing the mixing of supersymmetric particles (often called "sparticles"), into mass eigenstates; seven parameters governing CP-violation in the extended electroweak sector; 40 parameters governing CP-violation in sfermionic interactions; and 21 masses. All the SM particles along with their respective superpartners are shown in Figure 2.2

Supersymmetry breaking

Earlier in this chapter, we mentioned how all particles in a supermultiplet must have the same properties save for their spin. This effectively constrains supersymmetric particles to have the same masses as their SM partners. This is a problem, since if that were the

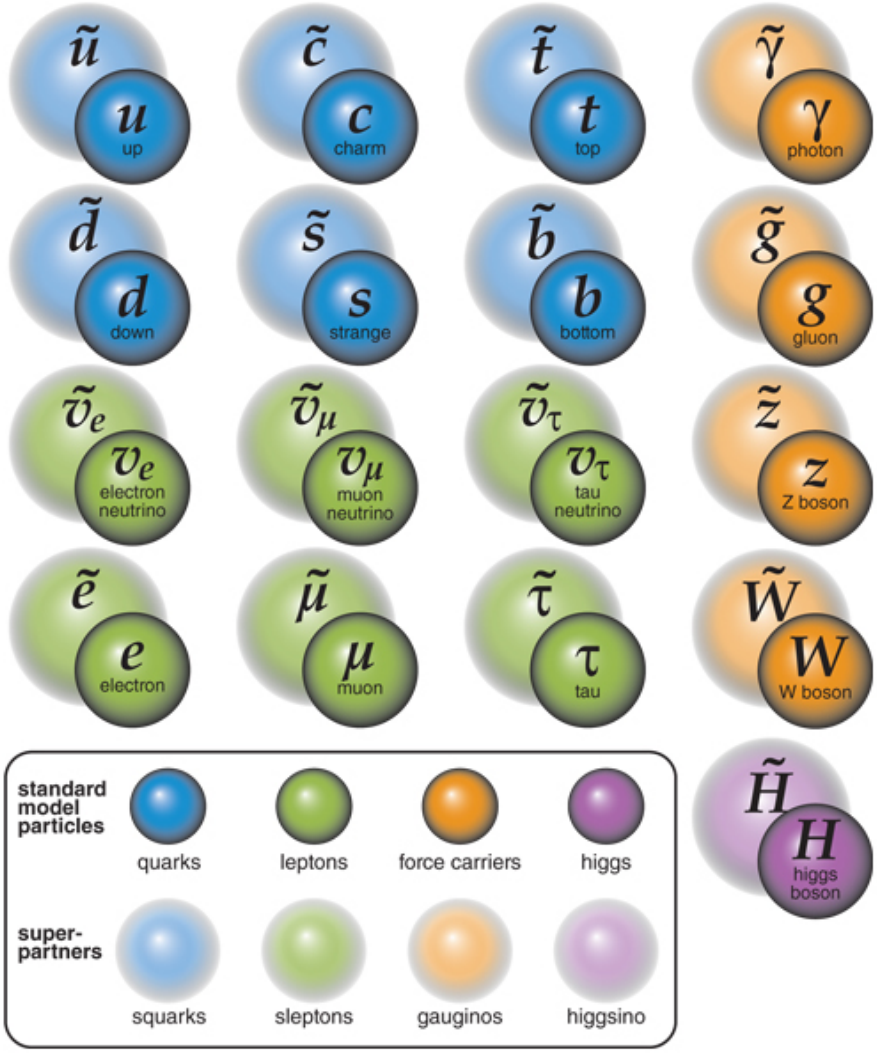


Figure 2.2 – Standard model particles along with their supersymmetric partners [82].

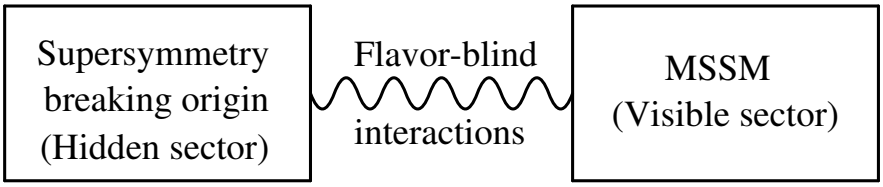


Figure 2.3 – Hypothesized Supersymmetry breaking mechanisms usually involve a hidden sector, which couples only to the supersymmetric sector of the MSSM [48].

case the universe would probably be a very different place than it appears now; at the very least, superpartners of stable SM particles would have been observed already, which is manifestly not the case. The implication is clear: either SUSY is not realized in nature, or it is a broken symmetry.

Since the exact breaking mechanism is unknown, the masses and the mixing parameters of supersymmetric particles cannot be predicted out-of-the-box in the MSSM, which explains why we need what seems like an absurd amount of additional free parameters. There are a few hypothesis as to what a viable SUSY breaking mechanism might look like, and they all reduce the number of additional parameters to less than a dozen. Most breaking mechanism split the MSSM into a visible sector, which includes the SM, and a hidden sector, which is completely decoupled from the SM but not from the supersymmetric part of the visible sector, as can be seen in Figure 2.3; these hidden–visible couplings effectively raise the sparticle masses beyond that of their SM partners. For example, the hidden sector might be an additional gauge sector, as in the "GMSB" model [83–85], or can be related to gravity, such as in the minimal SuperGravity (mSUGRA) model [86–88].

The Higgs mass in the MSSM

We've seen in section 1.2.1 how in the Standard Model the Higgs boson mass receives large contributions at one-loop level proportional to the square of the ultraviolet cutoff from massive fermions:

$$\delta m_{\mathrm{h},f}^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{\mathrm{UV}}^2 + O\left(m_f^2 \log\frac{\Lambda_{UV}}{m_f}\right). \tag{2.8}$$

In the MSSM, each of the associated scalars in the chiral supermultiplet containing the fermion also contributes:

$$\delta m_{\mathrm{h},s}^2 = \frac{\lambda_S}{16\pi^2}\Lambda_{\mathrm{UV}}^2 + O\left(m_S^2 \log\frac{\Lambda_{UV}}{m_S}\right). \tag{2.9}$$

Two things are worth noting about the $O(\Lambda_{UV}^2)$ terms of these two equations: their signs are opposite, and they differ by a factor 2 if $\lambda_s = |\lambda_f^2|$[3]. Each chiral supermultiplet will contribute two additional diagrams beyond the top contributions, as seen in Figure 2.4, and the three $O(\Lambda_{UV}^2)$ terms will sum to zero. This procedure is not ad-hoc; it's a direct consequence of the new symmetry introduced in the MSSM and therefore the result holds for all fermions and at all perturbative orders. Because of supersymmetry breaking,

[3]This relationship always holds in unbroken SUSY, and is also valid when SUSY is broken by a hidden sector.

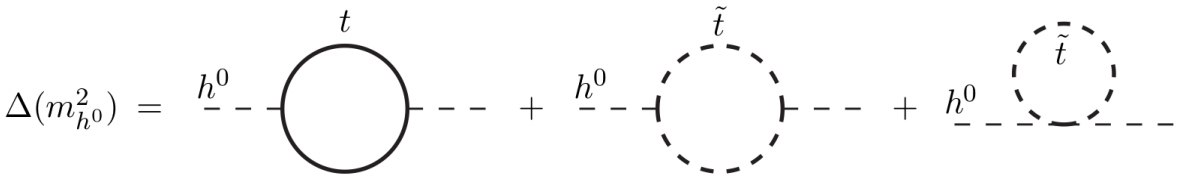


Figure 2.4 – One-loop contributions to the Higgs mass from the top quark supermultiplet [48].

Figure 2.5 – Gluino contribution to the Higgs mass at two-loop level [89].

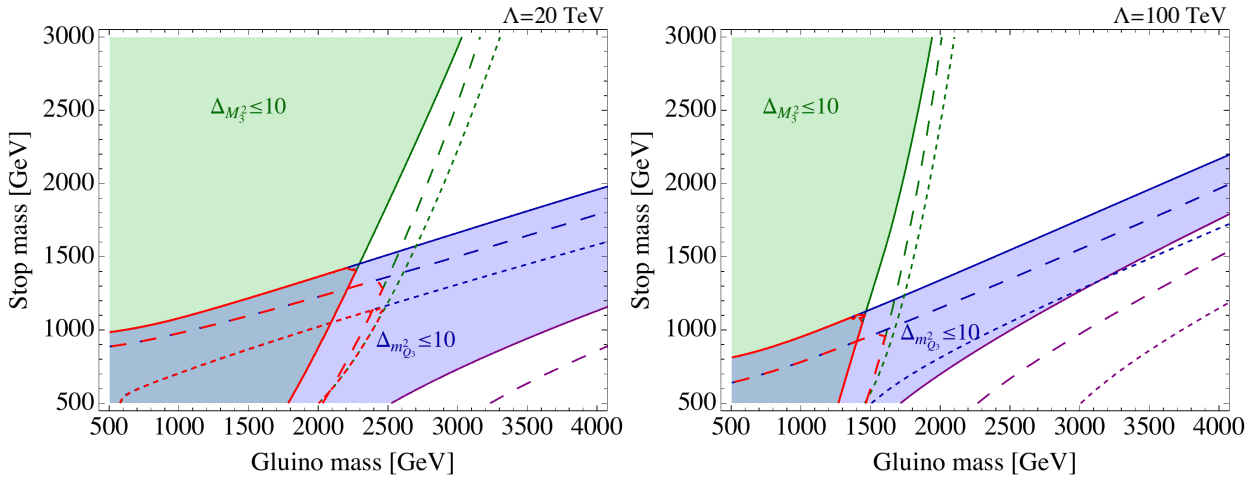
the magnitude of the remaining correction is then dominated by the scalar's mass, which shouldn't be too far above the electroweak scale if the fine-tuning is to be kept at a minimum.

The highest mass fermions being the top and bottom quarks at ≈ 175 GeV and ≈ 4 GeV, respectively, the stops and the sbottoms cause the largest remaining corrections to the Higgs mass. It is also worth noting that the stop and sbottom masses themselves get important corrections from the gluino and consequently so does the Higgs, once the computation accounts for two-loop effects as seen in Figure 2.5. Therefore, if the fine-tuning problem is to be solved in the MSSM, these three supersymmetric particles must have masses not too far above the electroweak scale and as a consequence would be likely to be produced at TeV-scale colliders such as the LHC.

It is possible to quantify the amount of fine-tuning introduced by a particular realization of the MSSM mass parameters, by using the Barbieri-Giudice measure [90, 91]:

$$\Delta \approx \max_{M \in \mathrm{MSSM}} \left| \frac{M^2}{m_{\mathrm{h}}^2} \frac{\partial m_{\mathrm{h}}^2}{\partial M^2} \right|, \tag{2.10}$$

where M stands for any MSSM mass parameters. This measure probe the sensitivity of the Higgs mass to perturbation of M. Figure 2.6 shows an example of the use of this measure in the stop–gluino mass plane, using precision computations for the sparticles contributions to the Higgs mass; it shows that a sizable part of this phase-space can lead



(a) SUSY breaking scale Λ = 20 TeV (b) SUSY breaking scale Λ = 100 TeV

Figure 2.6 – Higgs mass fine-tuning introduced by the gluino and stop masses (M_3 and M_{Q3}) estimated with the Barbieri-Giudice measure (Eq. 2.10) with two different values of the SUSY breaking scale [91].

to minimally fine-tuned models even with TeV-scale masses.

The "Electroweak measure", Δ_{EW}, is an alternative way of quantifying the amount of fine tuning which defines a minimally fine-tuned model as one where all SUSY contributions to the Higgs scalar potential $V(\phi)$ are of order $m_Z^2/2$ [92–94]:

$$\Delta_{\text{EW}} = \max_{C} \left| \frac{C}{m_Z^2/2} \right|. \tag{2.11}$$

Here, the C terms are individual SUSY contribution terms involved in the $m_Z^2/2$ computation. If some terms contribute substantially more than $m_Z^2/2$, then the Z mass becomes fine-tuned. The study performed in Ref.[94] attempts to probe the fine-tuning introduced by many different realizations of SUSY breaking models using both measures. As seen in Figure 2.7, there are still many unexcluded models with low fine-tuning and which are consistent with the currently observed bounds on the dark matter relic density.

R-parity conservation and dark matter

The lightest neutralino, $\tilde{\chi}_1^0$, is of particular interest because of the principle of *R-parity conservation* [96]. Every particles of the MSSM has an R-parity number constructed from its baryon and lepton numbers, B and L, and its spin S:

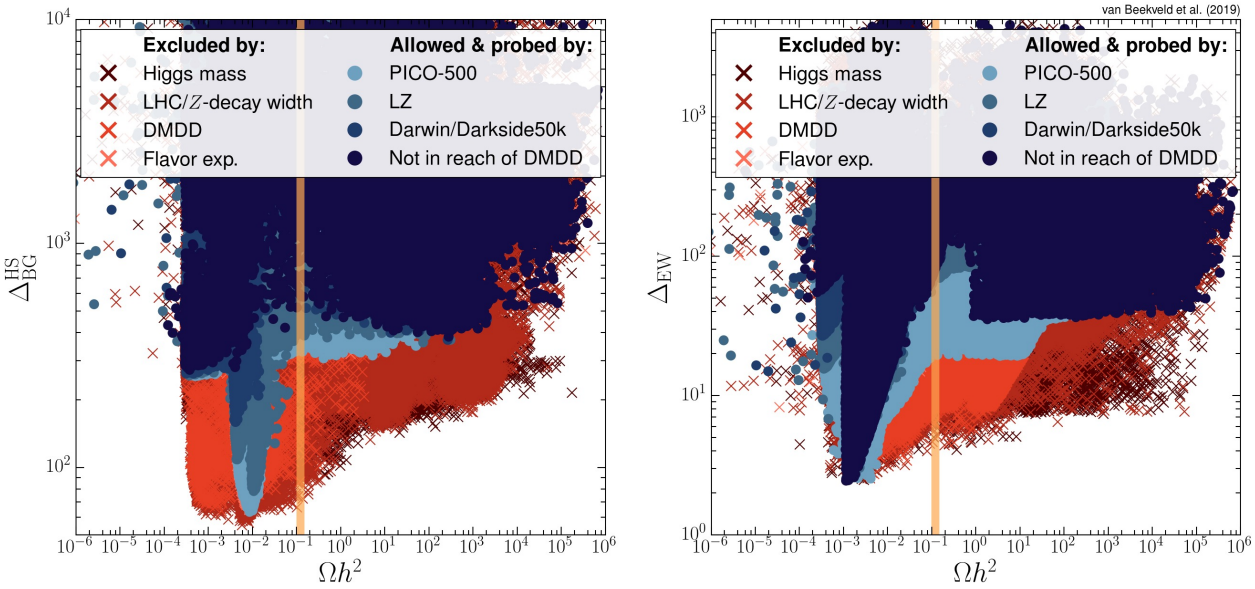


Figure 2.7 – Fine-tuning versus the physical dark matter density parameter Ωh^2 for realizations of the pMSSM-GUT [95] SUSY breaking scenario. Red crosses indicate models already excluded by Higgs mass considerations, LHC experiments, dark matter direct detection (DMDD) experiments, and flavor physics measurements. The circles represent not-yet-excluded models that could be constrained by various DMDD experiments. The shaded band corresponds to the Λ_{CDM} value of Ωh^2. [94]

$$R = (-1)^{3(B-L)+2S}. \tag{2.12}$$

Requiring that R-parity be multiplicatively conserved at each interaction vertex effectively forbids proton decay. In the SM, there is no renormalizable interaction that could lead to such an event, but the extended Lagrangian of the MSSM has no such constraint; The proton is known to be extremely stable, with a lifetime of at least 2.1×10^{29} years[10], and so requiring that it be absolutely stable in the MSSM is likely to be at least a good approximation. R-parity has at two very important consequences beyond proton stability, both owing to the fact that SM particles all have an R-parity of $+1$ while their superpartners have an R-parity of -1:

- Supersymmetric particles are always pair-produced if there are only SM particles in the initial state, as is the case in particle colliders;
- Supersymmetric decay chains must end with an even number of the lightest supersymmetric particle, or LSP.

If the neutralino is the lightest supersymmetric particle, then by the latter statement it will

be stable. Moreover, since it is formed by mixing superpartners of massive gauge bosons, it also has a mass; and by construction it only interacts via the weak interaction and is electrically neutral. Therefore, if R-parity is conserved then *the lightest neutralino, $\tilde{\chi}_1^0$, is a perfect WIMP dark matter candidate* [97, 98]. Indeed, a large part of phase-space of the MSSM parameters lead to $\tilde{\chi}_1^0$ cross-sections in the early universe that are compatible with the observed dark matter relic density.

2.2 Experimental status

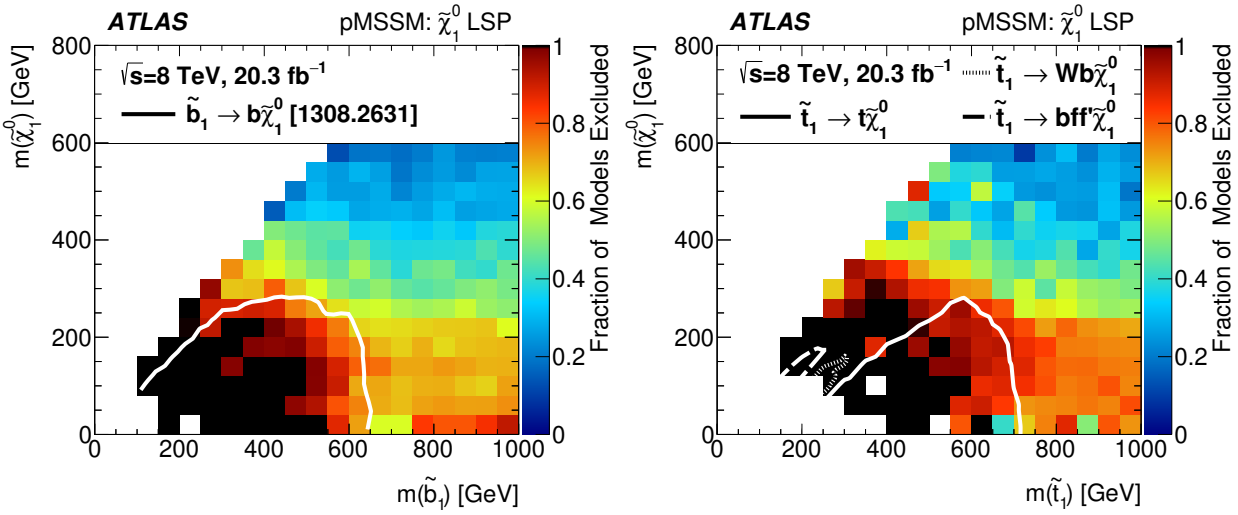
With 105 free parameters beyond that of the SM, the MSSM phase-space is simply too large to be considered in its entirety[4], and some cleverness is needed to organize the searches and interpret the constraints that they set. In this section, we consider two different paradigms: the pMSSM (Section 2.2.1), and simplified models (Section 2.2.2).

2.2.1 The pMSSM

As just mentioned, the MSSM phase-space is extremely huge. Fortunately, its overwhelming majority is already ruled-out by experiment. For example, many models arising from the MSSM lead to an excess of dark matter and/or CP-violation with respect to current observations. Using experimental considerations such as these, it is possible to project the 105 parameter MSSM into 19 or so parameters; the resulting model is referred to as the *phenomenological MSSM*, or pMSSM [99, 100]. The resulting phase-space is still enormous; however, since all model points in the pMSSM are, by design, phenomenologically viable, it's possible to estimate the constraints imposed by searches by sampling from the allowed phase-space and performing statistical tests on the resulting models to gauge how many of them are conclusively excluded. The results can then be visualized in many different planes, such as in Figure 2.8, which show the impact of ATLAS $\sqrt{s} = 8$ TeV SUSY searches in the sbottom/stop–neutralino mass plane. Comparing this with Figure 2.6, we see that the first ATLAS data-taking run has not ruled out natural supersymmetry.

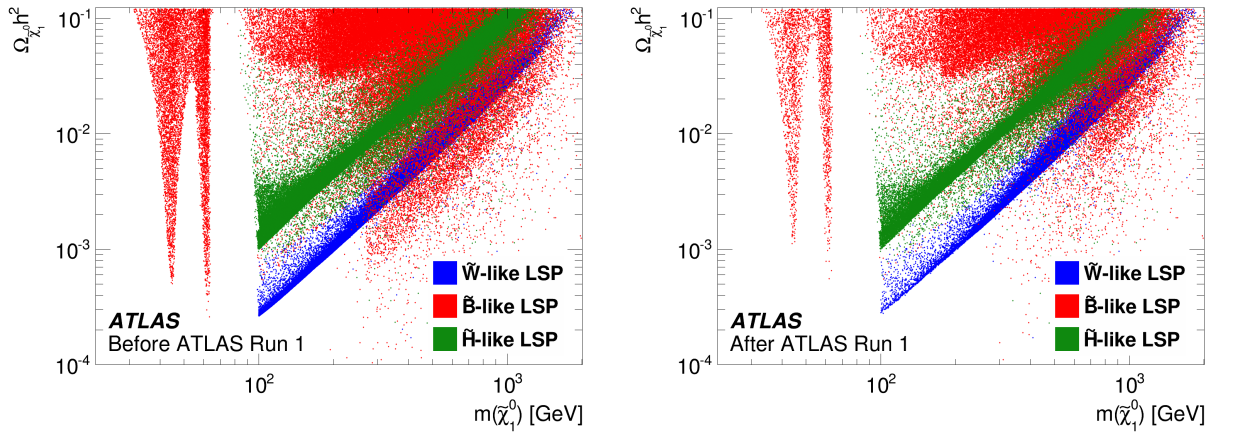
Since the pMSSM takes the dark matter relic density into account, it provides a natural framework in which to assess the impact of searches on the viability of SUSY as a solution to the dark matter problem. In Figure 2.9, the conclusion is similar as before: the first ATLAS data-taking run has not ruled out this particular motivation.

[4]To give an idea of just how so, discretizing each parameter axis into two halves each would still leave us with $2^{105} \approx 4 \times 10^{31}$ combinations to consider. Assuming you could do an hypothesis test in just a single nanosecond, it would take more than one million years to fully cover this crude binning using a computer with a *million* cores.



(a) sbottom (b) stop

Figure 2.8 – Impact of ATLAS search on the pMSSM in the (a)sbottom–neutralino and (b)stop–neutralino planes, quantified by the fraction of pMSSM models with the required masses that are excluded by ATLAS analyses. [101].



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

Figure 2.9 – Impact of ATLAS Run 1 searches on the amount of dark matter generated in the pMSSM, for three different LSP mixing scenarios (Wino-, Bino-, or Higgsino-dominated) [101].

2.2.2 Simplified SUSY models

The pMSSM approach outlined in the last section is very useful, but the process of sampling parameters, generating datasets and performing statistical tests is quite resource-intensive[5]. Therefore, a simpler approach is usually employed to guide individual searches: the *simplified model* paradigm [102–104]. In such models, only a few parameters of interest are considered – usually the masses of the sparticles being searched for – and the rest are decoupled. This allows setting limits based on specific final states, without reference to the underlying MSSM parameters that might have produced them. In this way, the phase space is reduced to a very manageable set of parameters (usually two or three) that can be comprehensively covered. For example, the model in Figure 2.10a has three free parameters corresponding to the gluino, stop, and neutralino masses. Since limits are better visualized in less than three dimensions, some parameters can be removed by assuming that some of the intervening particles are off-shell, such as in Figures 2.10b and 2.10c in which the gluino decays are modeled as three body decays, yielding the *Gtt* and *Gbb* models; these are the main benchmark models considered in this thesis and searched for in Chapter 5. Such models, which have two free parameters, are sometimes called "grids" since the parameter space can be visualized in a two-dimensional plane.

Events from these models have very striking topologies, and thus contain a lot of information that can be exploited to recognize them among the SM background. Both models start with each gluino of the pair decaying to pairs of third-generation quarks and a neutralino:

$$\tilde{g} \rightarrow b + \bar{b} + \tilde{\chi}_1^0 \quad (\mathrm{Gbb}), \tag{2.13}$$

$$\tilde{g} \rightarrow t + \bar{t} + \tilde{\chi}_1^0 \quad (\mathrm{Gtt}). \tag{2.14}$$

At this stage, the four b-quarks from a Gbb event will form jets that will contain b-hadrons and thus can be recognized by b-tagging algorithms (Section 5.2.3). In Gtt events, the four top quarks overwhelmingly decay to a b-quark and a W boson, which itself can decay to quarks or leptons:

$$t \rightarrow b + W, \tag{2.15}$$

[5]For this reason, the pMSSM sampling procedure outlined above is only performed by the collaboration after data taking runs of several years.

followed by:

$$W \to q + \overline{q}' \quad \text{or} \quad W \to \ell + \overline{\nu}. \tag{2.16}$$

Therefore, a fully-hadronic Gtt event will have twelve jets originating from the interaction, four of which are b-jets, while maximally leptonic events have four leptons and four associated neutrinos as well as four b-jets. A large amount of missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$ (Section 5.2.6), is also expected since in both grids the two neutralinos and the potential neutrinos deposit no energy in the detector.

The mass splitting, defined as the difference between the gluino and neutralino mass, has a large impact on the kinematics of the final states. Schematically, we can define two important regions in this phase-space:

- The boosted region: $m_{\tilde{g}} \gg m_{\tilde{\chi}_1^0}$;
- The compressed region: $m_{\tilde{g}} \approx m_{\tilde{\chi}_1^0}$.

In the boosted region, final state particles typically have more momentum than events from the compressed phase-space and are thus easier to distinguish from the SM background. Figure 2.11 shows the leading small-radius jet p_{T} (Section 5.2.2) and $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions in different mass splitting regimes of the Gbb and Gtt models.

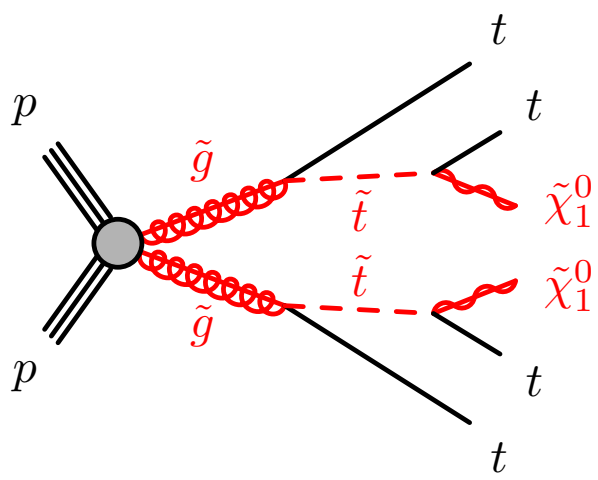
Figure 2.12 compares the bounds on these models obtained by ATLAS and CMS run 2 searches. For an approximately massless neutralino, gluino masses of $\gtrapprox 2.2$ TeV are not excluded and so natural SUSY is not ruled out by these searches.

In these models, it has been tacitly assumed that the stop and sbottom masses plays no large part in the resulting phenomenology. Figure 2.13 shows that this is a good approximation in the Gtt model; the limits are only significantly different when the stop is approximately mass degenerate with the gluino or the neutralino, in which case the available kinematic phase-space is reduced.

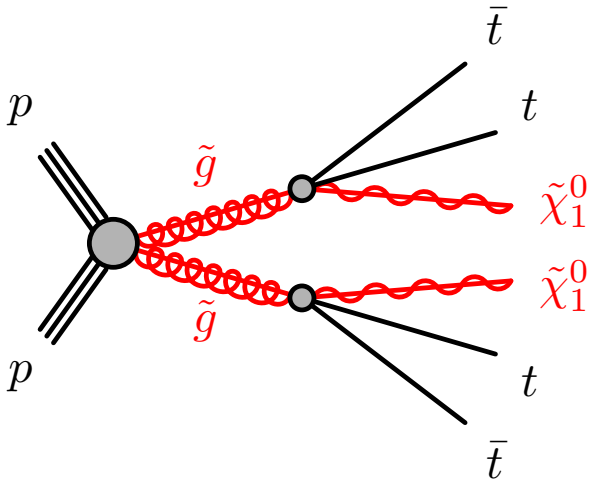
Stop and sbottom masses are somewhat less constrained by direct searches, as seen in Figure 2.14, with the most stringent lower bounds being approximately 1.2 TeV in both cases; natural SUSY is not ruled out by these searches either.

2.3 Conclusion

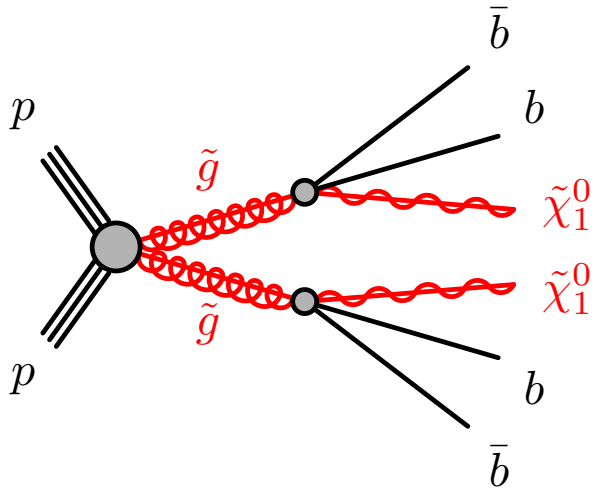
In this second and last theoretical chapter, we have learned about a family of theories that extend the usual spacetime symmetries to include exchange of bosons and fermions, called supersymmetry. By considering the basic Wess-Zumino model, we've seen how



(a) $\tilde{g} \to t + \tilde{t}/\tilde{t} \to t + \tilde{\chi}_1^0$



(b) $\tilde{g} \to t + t + \tilde{\chi}_1^0$



(c) $\tilde{g} \to b + b + \tilde{\chi}_1^0$

Figure 2.10 – Simplified models of pair-produced gluinos decaying to (a)top quarks via stop quarks and (b,c)top or bottom quarks and neutralinos via off-shell squarks.

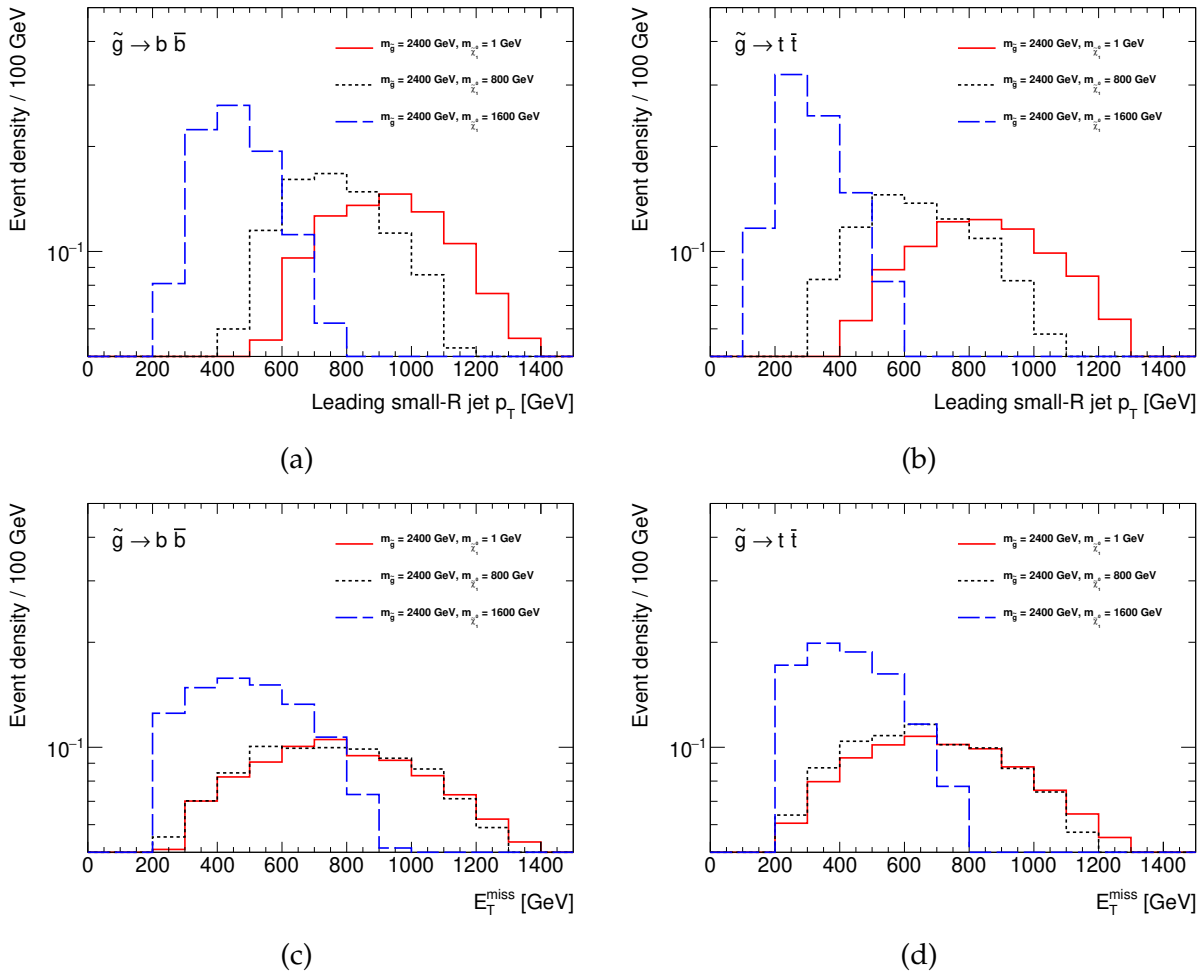
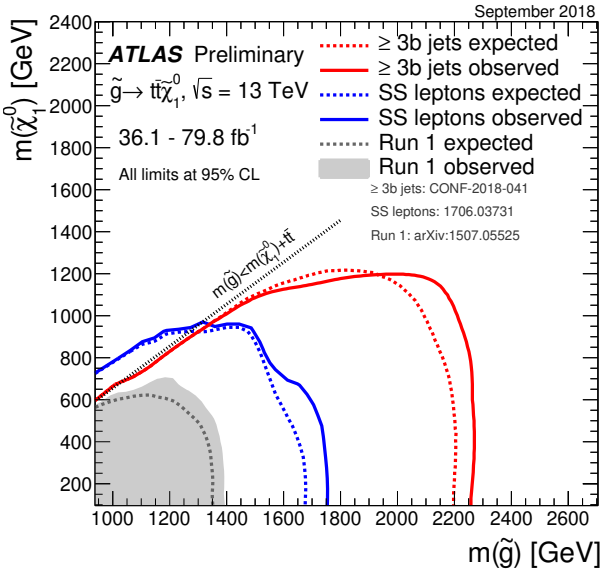
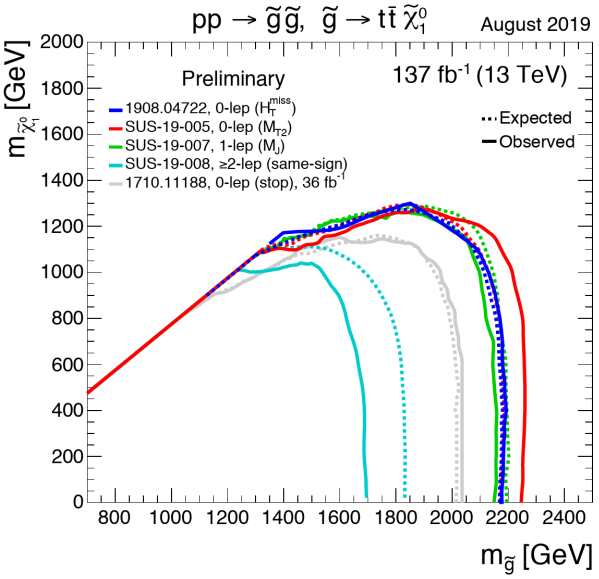


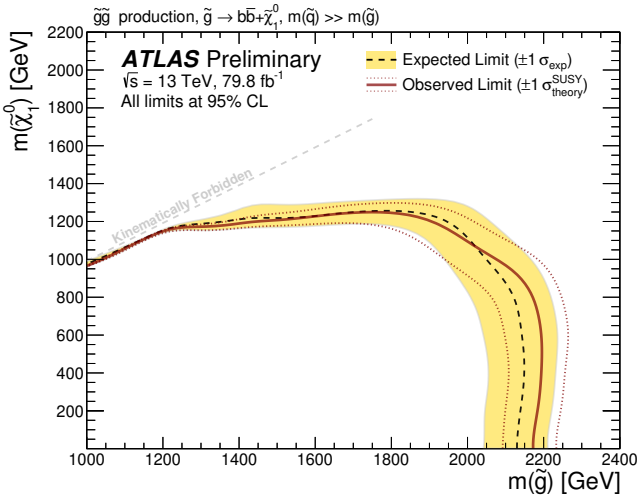
Figure 2.11 – (a,b) Leading jet p_{T} and (c,d) $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions for different mass points of the (a,c) Gbb and (b,d) Gtt models, showing how the events become less energetic as the mass splitting decreases.



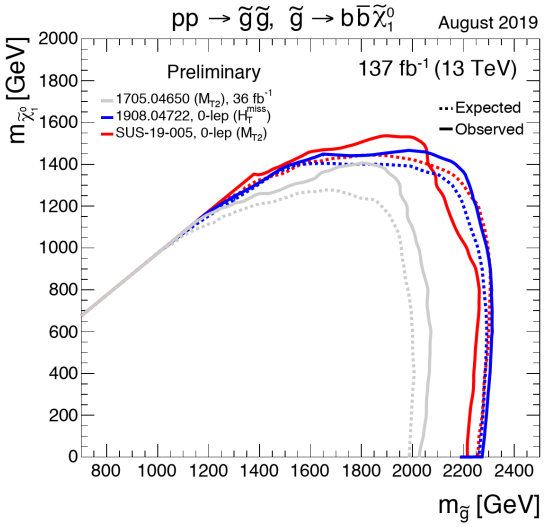
(a) Gtt, ATLAS



(b) Gtt, CMS



(c) Gbb, ATLAS [9]



(d) Gbb, CMS

Figure 2.12 – ATLAS and CMS Run 2 constraints on the (a,b) Gtt and (c,d) Gbb models with off-shell squarks. The total integrated luminosities are of 79.8 fb^{-1} and 137 fb^{-1} for the ATLAS and CMS results, respectively.

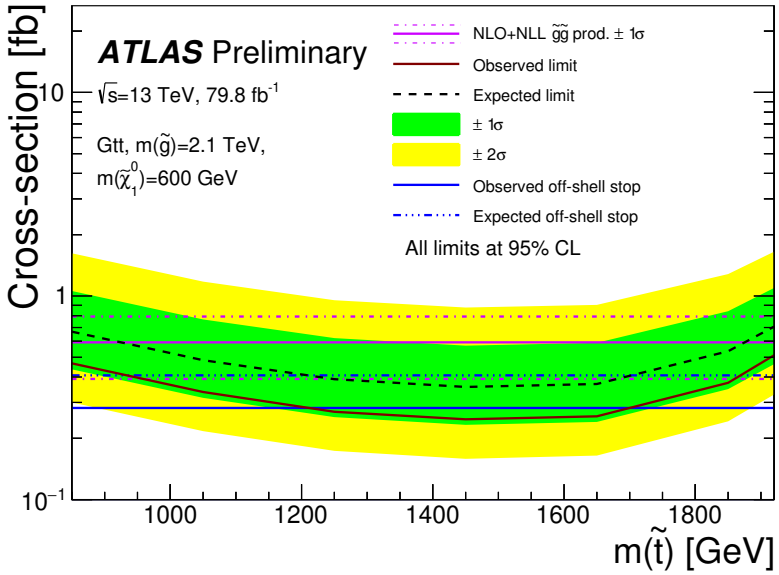
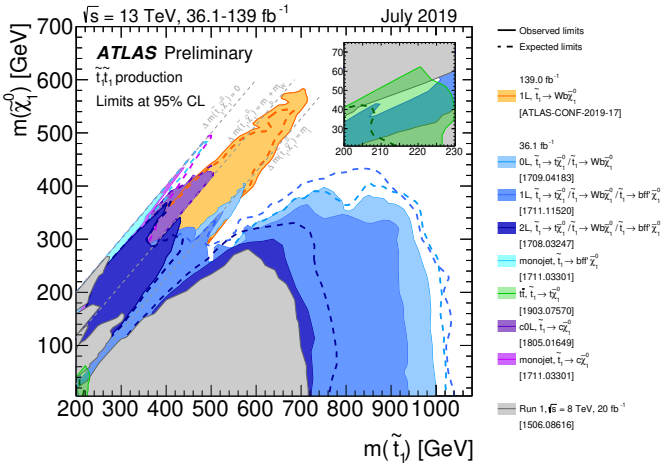


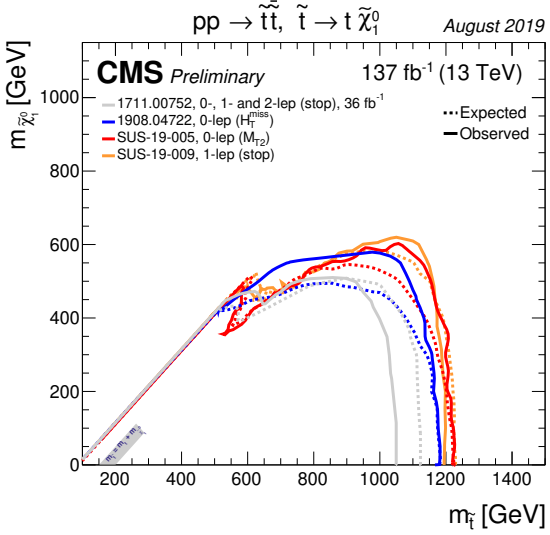
Figure 2.13 – Impact of the stop mass on the Gtt cross-section limit for the simplified model of Figure 2.10a. The limits are only significantly different when the stop is approximately mass degenerate with the gluino or the neutralino [9].

building supersymmetrical Lagrangian is quite simple, and that all Standard Model particles can be grouped with their beyond-the-standard-model superpartners into chiral or vector supermultiplets. We've then discussed the MSSM model, which makes the SM supersymmetric; in this extended theory, the hierarchy and dark matter problems that were discussed in the previous chapter are potentially solvable. Moreover, by considering the phenomenological MSSM (pMSSM) as well as simplified SUSY models, we've seen that viable supersymmetric extensions of the SM are not yet ruled out by experiment.

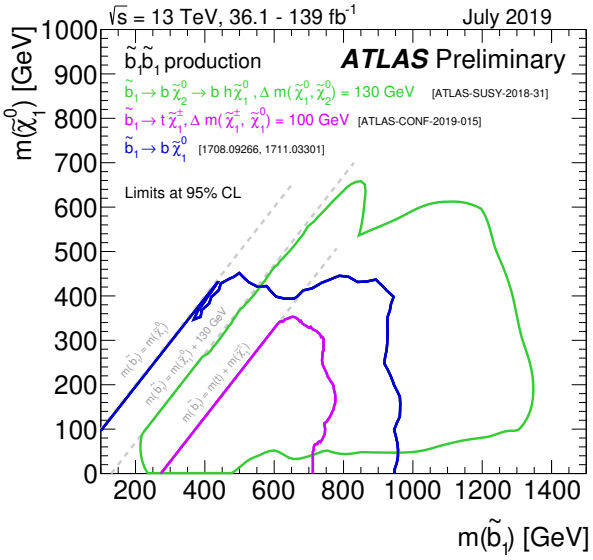
With all of this in mind, it is now time to switch our focus to experimental considerations, starting with a description of the current most powerful particle accelerator: the Large Hadron Collider, or LHC, a machine that could very well lead to laboratory observations of supersymmetric particles.



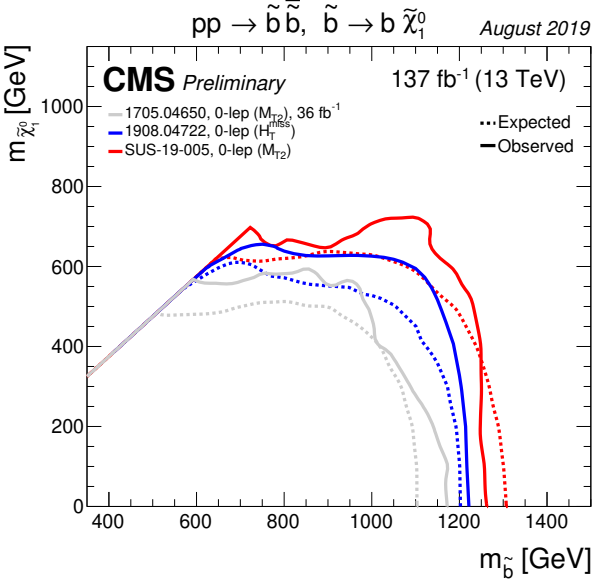
(a) ATLAS



(b) CMS



(c) ATLAS



(d) CMS

Figure 2.14 – ATLAS and CMS Run 2 limits in the (a,b) neutralino–stop and (c,d) –sbottom mass planes.

Chapter 3

The Large Hadron Collider and the ATLAS detector

3.1 The Large Hadron Collider

Note: unless otherwise noted, this section is based on Reference [6].

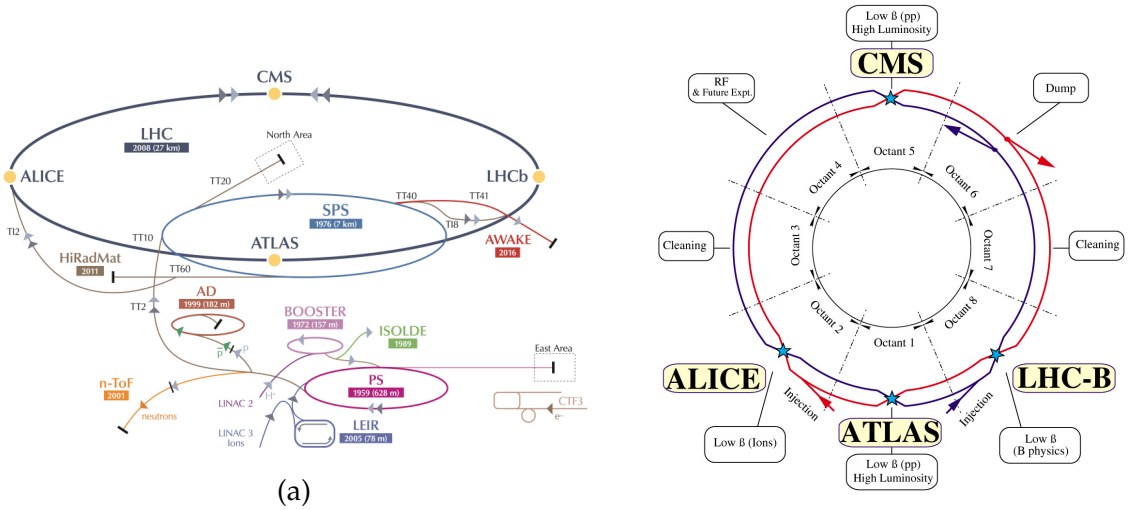
The Large Hadron Collider (LHC) [6], is a 27 km particle accelerator situated at the European Organization for Nuclear Research (CERN) along the France/Switzerland border near Geneva. Its tunnel was bored in the 80s to house the Large Electron-Positron Collider (LEP) [105], which finished operations in the year 2000 to allow construction of the LHC.

The LHC is a hadron collider primarily used to collide beams of protons. It has also seen special runs of proton–lead, lead–lead, and xenon–xenon collisions; this thesis is chiefly concerned with proton–proton (p–p) collisions. The LHC's second data taking run, colloquially identified as "Run 2", took place from 2015 to 2018 and collided two 6.5 TeV proton beams for a center-of-mass energy of 13 TeV, the largest ever achieved in laboratory.

The LHC is designed to accelerate protons from an initial energy of 450 GeV. Furthermore, single protons are found within hydrogen-1 atoms, which are electrically neutral and cannot be electromagnetically accelerated. Consequently, a few preparation stages are needed upstream of injection into the LHC's rings.

The adventure starts by opening the valve on a small gas bottle that feeds hydrogen-1 atoms into a *duoplasmatron*; this device subjects the hydrogen gas to an electrical field strong enough to produce a plasma, allowing the electrons to be collected by a cathode and the protons to undergo a small initial electrostatic acceleration in the direction of an anode with a small central aperture, thereby forming a beam.

This beam is then transferred to the first proper accelerator in the chain: Linac 2 [106],



(a) (b)

Figure 3.1 – (a)CERN accelerator complex [114] and (b)LHC layout [115].

a linear accelerator inaugurated in 1978, which raise the protons's energy up to 50 MeV, a regime suitable for injection into a circular accelerator[1]. The beam is then transferred to the Proton–Synchrotron (PS) complex [107, 108], comprising the Proton–Synchrotron Booster (PSB, 1972) and the PS itself (1959), which accelerates the protons up to 1.4 GeV and 25 GeV, respectively. Finally the beam is circulated in the Super Proton–Synchrotron [109] (SPS, 1976), the accelerator that famously enabled the discovery of the W and Z bosons (Section 1.1.3) in 1983 [110–113]. This last machine in the pre-LHC acceleration chain energizes the beam to 450 GeV before handing it over to the LHC where it will be collided. The whole accelerator chain, as well as a depiction of the LHC's ring, can be seen in Figure 3.1.

3.1.1 The recipe for high energy

Let's take a few moments and consider just how the LHC is able to achieve a record center-of-mass energy of 13 TeV. There are two basic ingredients to this feat: radio-frequency (RF) cavities and dipole magnets, both of which are enabled by superconducting technology.

Radio-frequency cavities

The accelerating sector of the LHC and its upstream chain are made of series of closely-spaced hollow resonant cavities, made of niobium sputtering on copper, joined by an insulating beam-pipe. Each cavity is subject to an electromagnetic wave to build up a resonant electric field on the inside, producing much larger gradients than is possible with

[1]A new machine, Linac 4, is slated to replace Linac 2 starting from the LHC's third data-taking run.

electrostatic or inductive acceleration. However, since the field is oscillating, the protons have to be arranged into bunches and timed such that they enter the cavity when the field points in their direction of motion. This requires the following relationship between the spacing between cavities L, the proton's velocity v and the standing wave frequency f:

$$f = \frac{v}{2L}. \tag{3.1}$$

At a fixed frequency, an in-phase proton will tend to settle at the velocity v since any further acceleration will make it fall out of phase and decelerate. Therefore, the beam energy is largely dependent on the frequency of the standing wave inside the cavities [116].

The field oscillation can be exploited to minimize the longitudinal spread of a bunch by having it enter the cavity slightly early. In such cases, protons lagging behind the bunch center will be nearer to the cavity's center when the field peaks and they get a correspondingly bigger acceleration, while protons going too fast are more de-phased than the rest of the bunch and are accelerated less.

The LHC has a total of 16 such cavities (eight per beam) housed in four separate cryomodules, which cool the apparatus to as low as 4.5 K. They are all situated in the same octant of the LHC ring (see Figure 3.1b).

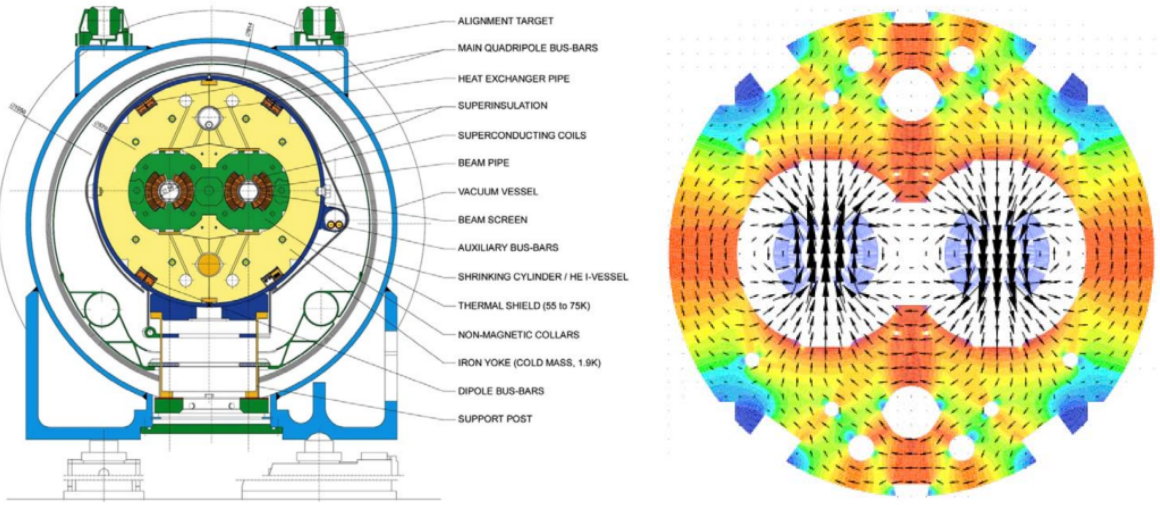
Dipole magnets

The dipole magnets' purpose is to keep both proton beams on a circular trajectory in the center of the LHC's beam pipes. There is a direct relationship between the beam energy E_{beam} (in GeV) and the required magnetic field B (in Teslas) to bend it around a given radius of curvature R (in meters) [117]:

$$B = \frac{E_{\text{beam}}}{0.3R},\ E_{\text{beam}} \gg m_{\text{proton}}. \tag{3.2}$$

For a beam energy of 6.5 TeV and a radius of 2.6 km[2], the required B-field is an extreme 8.33 T. This field is created by a correspondingly intense electric current of about 11.85 kA. The only viable way to operate in this regime is by employing superconducting technology; the LHC's dipole use niobium-tin (NbTi) coils cooled to 1.4 K by super-fluid helium.

[2]An astute reader might notice that $2\pi \times 2.6 \approx 16$ km; this is because the LHC ring has many straight sections, wherever there are no dipoles, which raise the effective radius for a total circumference of 27 km.



(a) [6] (b) [118]

Figure 3.2 – (a) Cross-section and (b) magnetic field of the LHC dipole system.

Owing to the two counter-rotating proton beams, the dipole system comprises two beam-pipes subject to opposite magnetic fields to steer the beams in the right directions, as can be visualized in Figure 3.2.

3.1.2 The recipe for high luminosity

By analogy with the brightness of a light source being defined by the amount of outgoing photons, the *luminosity* at one of the LHC's interaction point (IP) is a controllable parameter that defines the expected amount of outgoing particles. For two identical and symmetrical beams, the instantaneous luminosity is

$$L_{\text{inst}} = f_{\text{coll}} \frac{n^2}{4\pi\delta^2} \times \mathcal{F}, \tag{3.3}$$

where f_{coll} is the bunch collision frequency (25 ns^{-1} at the LHC), n is the number of protons per bunch, δ is the transverse beam spread and $\mathcal{F}$ is a correction factor accounting for various higher-order effects such the bunch-crossing angle and the longitudinal bunch length. For a physical process with cross-section σ, recorded for a length of time T:

$$N_{\text{exp}} = \sigma \int_0^T L_{\text{inst}}(t)dt. \tag{3.4}$$

The three most important ingredients required to achieve a high luminosity are a very high collision rate, a large bunch density, and a small transverse beam spread. These first two design goals ruled out from the outset the attractive idea of having one of the beam be composed of anti-protons, like at the Tevatron collider at Fermilab in the USA and at the S$p\overline{p}$S at CERN; such beams have several advantages such as allowing very precise measurements of CP-violating processes [119] and allowing a simpler dipole magnet design, but anti-protons must be produced by first colliding a proton beam with a fixed target, a process that is not fully efficient. This makes it more difficult to obtain enough particles to sustain a high collision rate and to raise the bunch density with a reasonable turnaround time[3]. In fact, the Tevatron operated with ≈ 6 bunches/km, while the LHC is nominally designed to have ≈ 70 bunches/km, an order of magnitude more [120].

The transverse beam spread is defined by two further parameters, the transverse *emittance*, ϵ, and the value of the β function at the interaction point, β^*. Along one of the transverse axis:

$$\delta = \sqrt{\epsilon\beta^*}. \tag{3.5}$$

The emittance is the primary beam quality parameter; it measures the spread of particles in the x–Δ_x phase space, where x is the particle position and $\Delta_x \equiv \frac{dx}{ds}$ is the particle deviation from the nominal path s. It can be numerically defined as the area of the ellipse encompassing all protons in this plane, as seen in Figure 3.3.

In ideal conditions, the emittance is a conserved quantity for the entirety of the beam's lifetime. However, this, assumes a perfectly linear magnetic field that is perfectly known all along the trajectory, a condition not realized in the real world. For example, mismatch between the optics at injection lead to emittance growth, and non-linearity in the bending and focusing magnetic fields can lead to fluctuations in Δ_x known as *betatron oscillations*, causing the emittance to grow over time. To counteract these effects, the LHC has a transverse feedback system known as the ADT, which corrects the beam orbit by recording its deviation and applying a correcting electric field at a latter point along the trajectory [122], as seen in Figure 3.4.

Emittance growth is synonymous with a loss of beam quality, since it comes with a corresponding drop in luminosity beyond the inevitable losses due to p–p interactions. Since the turnaround time of the LHC is not prohibitively long, the beam is usually dumped

[3]A viable turnaround time was achieved at the Tevatron by the usage of a storage ring to build-up the $\overline{p}$ beam ahead of time.

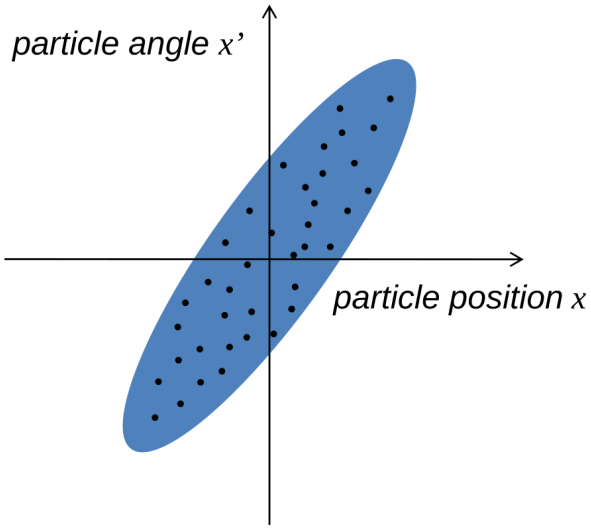


Figure 3.3 – The emittance ϵ is defined as the area of the ellipse encompassing all of the beam's constituent particles in the x–Δ_x plane, divided by π [121].

after about 12 hours solely based on luminosity considerations[4].

The β function value is related to the width of the emittance ellipse along the positional axis, as visualized in Figure 3.5. The use of quadruple magnets for focusing do not reduce the emittance, as stated earlier, but rotates the beam in the x–Δ_x phase space. The β value at the interaction point (IP), β^*, has to be very low to obtain a high luminosity; this is achieved at the expense of a temporarily higher spread in Δ_x. Out of the IP, the beam is de-focused to keep the orbit as stable as possible.

3.1.3 The LHC experiments

There are currently eight experiments installed along the LHC ring. Four of these are situated at the interaction points (see Figure 3.1b):

- **ATLAS** [4], a general-purpose detector used for this thesis, described in detail in the next section;
- **CMS** [5], a general-purpose detector similar to ATLAS;
- **LHCb** [124]: a special-purpose detector optimized for B-meson physics;
- **ALICE** [125]: a special-purpose detector optimized for heavy ion physics.

[4]The maximum beam lifetime is, however, much longer. For instance, due to a scheduled maintenance stop of the PS accelerator the LHC fill no. 4947 on 2016/05/21 was kept going for 35h28m [123].

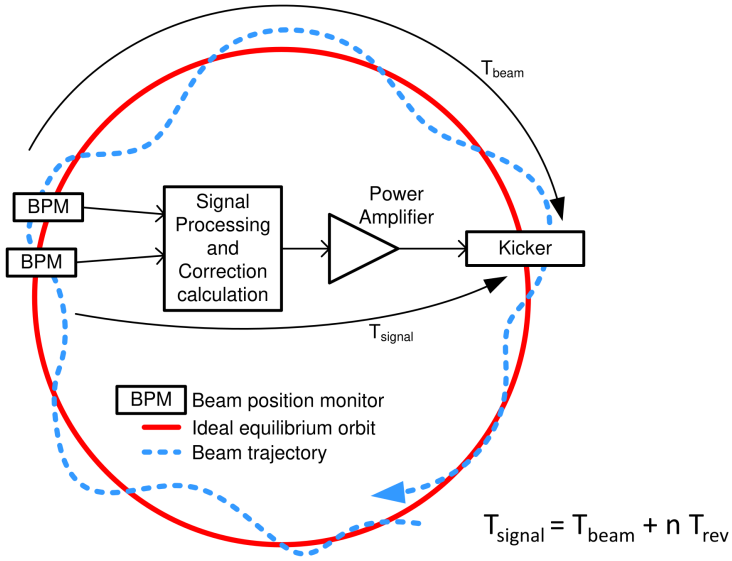


Figure 3.4 – Overview of the LHC's transverse feedback system (ADT) [122].

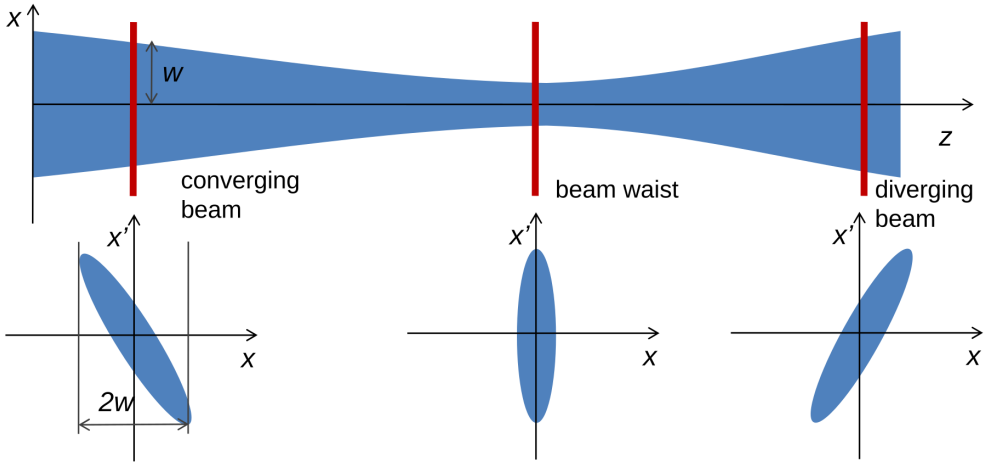


Figure 3.5 – β is related to the width of the emittance ellipse along the positional axis: $\beta = W^2/\epsilon$ [121].

There are also three smaller detectors installed on the beam-line near some of these detectors and their interaction points:

- **LHCf** [126], installed near the ATLAS detector cavern to measure the energy spectrum of particles produced at very small opening angles;
- **TOTEM** [127], installed on the beamline near the CMS detector cavern to study the total p–p cross-section as well as the proton structure;
- **MoEDAL** [128], installed near LHCb to look for magnetic monopoles;
- **FASER** [129], installed near ATLAS to search for long-lived exotic particles.

3.2 The ATLAS detector

Note: unless otherwise noted, this section is based on Reference [4].

The ATLAS[5] detector (Figure 3.6) is a huge general-purpose particle detector with a 25×44 m cylindrical shape, installed around LHC interaction point no. 1 (IP1). It can be decomposed in three main subsystems: the inner detector (ID, Section 3.2.2), the calorimeters (Section 3.2.3), and the muon spectrometer (Section 3.2.4). These subsystems produce a staggering amount of information about each collisions, and therefore a sophisticated data acquisition system is implemented (Section 3.2.5).

3.2.1 Standard ATLAS coordinate system

The coordinate system commonly used to describe the ATLAS detector has its origin at the interaction point. The z axis is longitudinal to the beam, while the x and y axes point to the center of the LHC ring and up towards the sky, respectively. The transverse plane, x–y, is particularly useful since the initial momentum around the z axis is approximately null. The azimuthal angle in this plane (around the z axis) is denoted by ϕ.

Massive objects can be localized in the longitudinal direction by their *rapidity*, Υ:

$$\Upsilon = \frac{1}{2} \log \left(\frac{E + p_z}{E - p_z} \right). \tag{3.6}$$

When the object's mass is small compared to it's energy,

[5] ATLAS ≡ A Toroidal LHC ApparatuS

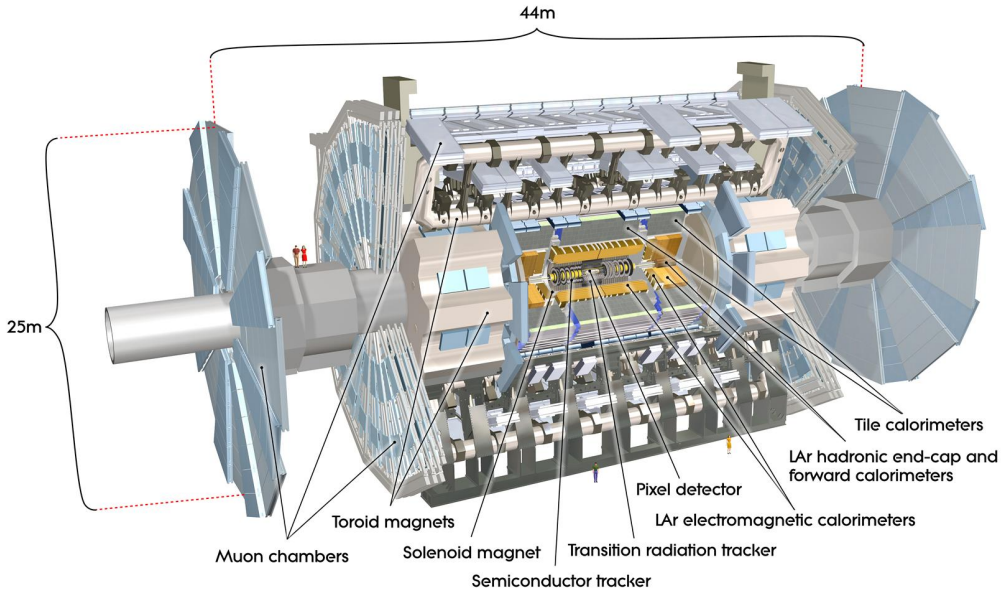


Figure 3.6 – The ATLAS detector [130].

$$E \gg m \implies E \approx |\overrightarrow{p}| \implies p_z \approx E \cos\theta, \tag{3.7}$$

where θ is the angle of the three-momentum vector with respect the z axis. In such cases, Υ can be approximated by considering the θ angle only:

$$\Upsilon \approx \eta = -\log\left(\tan(0.5\theta)\right), \tag{3.8}$$

The η variable is known as the *pseudo-rapidity*; it is visualized in the y–z plane in Figure 3.7. Distance between objects in the detector are often quantified in η–ϕ space by the dimensionless ΔR variable:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{3.9}$$

3.2.2 The inner detector

The ATLAS inner detector (Figure 3.8) is an array of three sub-detectors installed closest to the beam-line; its primary usage is to provide accurate charged particle track reconstruction

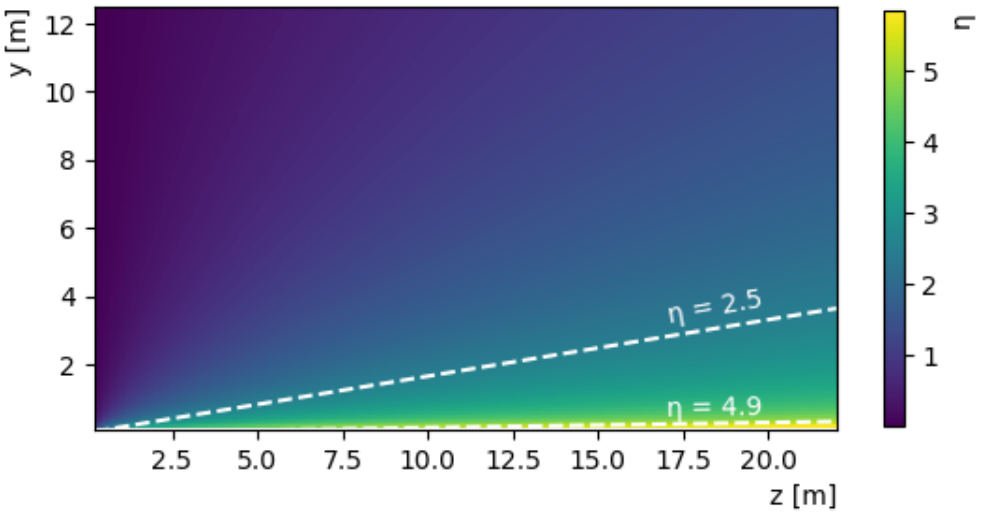
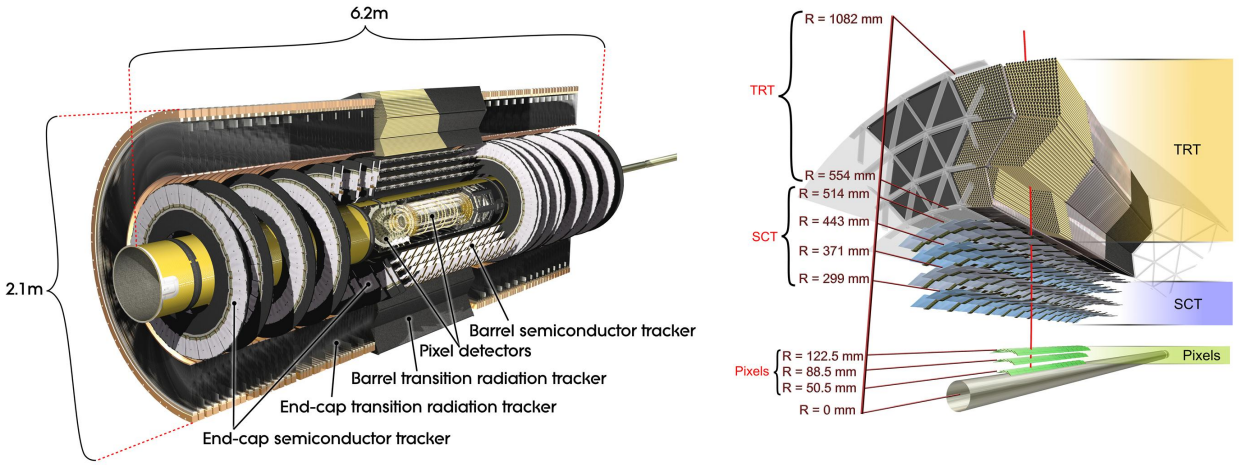


Figure 3.7 – Pseudorapidity (η) contours in the transverse–longitudinal (y–z) plane. The $\eta = 2.5$ and $\eta = 4.9$ lines define the limit of the inner detector and calorimeter acceptances, respectively.



(a) (b)

Figure 3.8 – (a) Longitudinal and (b) radial cross-sections of the ATLAS inner detector [131]. Note that there is now an additional pixel layer, the IBL, installed nearest to the beam in 2014 (this figure was made in 2008).

in the $|\eta| < 2.5$ region. The whole ID is enclosed in a solenoid providing a 2 T axial magnetic field and so charged particles moving away from the interaction point follow an helical trajectory with a radius of curvature in the plane transverse to the beam proportional to their momenta [117]:

$$R \propto \frac{p \cos \lambda}{B}, \tag{3.10}$$

where B is the magnetic field strength, R is the radius of curvature and λ is the angle of the helical trajectory, measured from the axis normal to the beam. Consequently, to obtain a measure of charged particle's momenta, an accurate measure of the track's curvature must be obtained. The error on the curvature measurement, δ_{curv} can be decomposed into two terms [117]:

$$\delta_{\text{curv}} = \sqrt{\delta_{\text{ms}}^2 + \delta_{\text{res}}^2}, \tag{3.11}$$

where δ_{ms} is the error due to multiple scattering of the particle inside the tracker's material, and δ_{res} is the error proportional to the finite precision ϵ_{meas} of the track position's sampling when it crosses an active layer of the tracker. The error due to multiple scattering is roughly proportional to the inverse of the momentum, therefore in high energy collisions the resolution error usually dominates the multiple scattering contribution. For an idealized tracker with many uniformly-spaced active layers, this error is estimated by:

$$\delta_{\text{res}} \approx \frac{\epsilon_{\text{meas}}}{L} \sqrt{\frac{720}{N+4}}, \tag{3.12}$$

where L is the track's length in the bending plane and N is the number of recorded hits along the trajectory.

Equation 3.12 implies that it's possible to obtain a good momentum resolution by maximizing the number of samplings, by obtaining very precise position measurements, or by maximizing the tracker's extent in the bending plane. The ATLAS ID implements all three of these solutions in order to achieve state-of-the-art momentum measurements:

- A four-layer silicon pixel detector (Section 3.2.2) is installed as close as possible to the beam-line, providing up to four very-high-precision measurements and maximizing the tracker's extent in the inward direction;

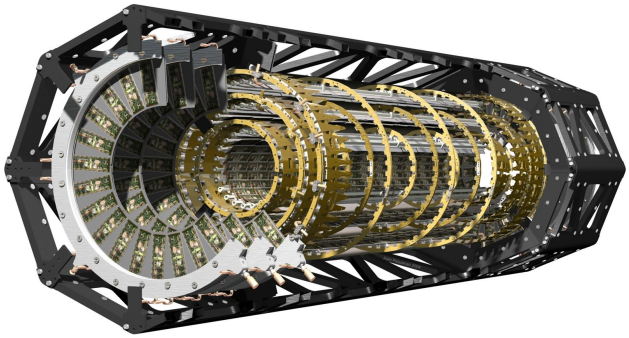


Figure 3.9 – Cut-out view of the ATLAS Pixel detector [133].

- A four-layer silicon strip detector (SCT, Section 3.2.2) is installed around the pixel detector, providing a further eight samplings, which are grouped into four high-precision measurements;
- An array of drift tubes, the transition radiation tracker (TRT, Section 3.2.2) is installed around the SCT, providing more than 30 less-precise measurements per tracks, maximizing both the number of measurements and the tracker's extent in the outward direction.

The Pixel Detector

The pixel detector (Figure 3.9) comprises four coaxial barrel-shaped layers with three disk-shaped layers on either end. Its silicon sensors are read out in 2-D regions called pixels (by analogy with a camera) that are defined by bump-bonds linking the silicon slab with dedicated channels of the readout electronics. In the three disks and the three outer barrel layers the nominal pixel sizes are of 50 and 400 μm in the bending and longitudinal planes while the first layer (the *Insertable B-Layer* [132], or IBL, whose installation in 2014 is pictured in Figure 3.10) has smaller 50×250 μm pixels. However, charged particles passing through a silicon sensor typically deposit energy in more than one pixel, and shape analysis techniques (Section 4) can achieve an intrinsic accuracy of 10 μm in the bending plane and of 115 μm in the longitudinal (barrels) or radial direction (disks).

The high luminosity attained in LHC collisions (Section 3.1.2) requires the pixel detector to be exceptionally radiation-hard, and since the performance of many ATLAS analyses is strongly dependent on track reconstruction efficiency and momentum resolution, it must remain so for the entirety of its useful life[6]. This motivates the use of non-conventional wafers made up of an n-type silicon bulk with p and $n+$ implants on either side. This

[6]The whole inner detector is slated to be replaced by a new all-silicon inner tracker, the ITk, in time for the fourth LHC run starting around 2027.

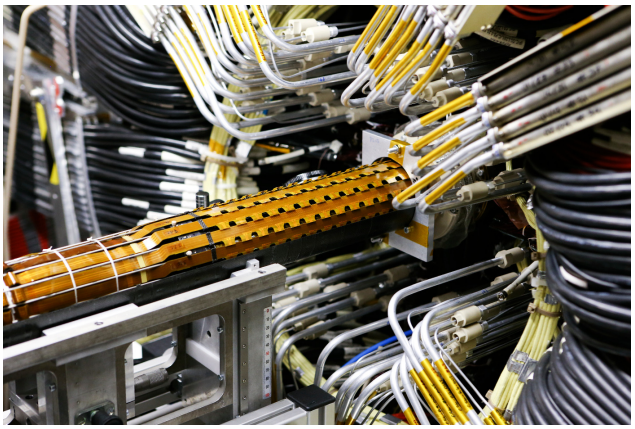


Figure 3.10 – Installation of the Insertable B-Layer in 2014 [134].

particular choice allows the sensors to remain operational even after type inversion of the bulk due to radiation damage. Furthermore, the silicon is highly oxygenated to delay the inevitable accumulation of radiation damage.

The Semi-Conducting Tracker

The semiconductor tracker, or SCT, is installed around the pixel detector and has a similar barrel–disk geometry with four layers. It employs more conventional and widely used silicon strip sensors, which are cheaper to produce[7]. Unlike pixel sensors, individual strips only produce a measurement in a single direction; to provide hit information in two dimensions, strip sensor modules are assembled in pairs laid out on top of each-other and misaligned by a 40 mrad angle as can be seen in Figure 3.11. Its intrinsic precision is of 17 μm in the bending plane and of 580 μm in the longitudinal direction. These measures are less precise than that of the pixel detector, but the four additional measurements at a much lower cost than that of a pixel detector of similar volume represent a good cost–benefit compromise.

The Transition Radiation Tracker

The inner detector is completed by an array of drift tubes know as the transition radiation tracker, or TRT. The tube structure is defined by a 4 mm polyimide wall whose aluminum-coated inner surface acts as a cathode, while the axial anode wire is made-up of gold-plated tungsten; a large voltage of 1.53 kV is applied between anode and cathode. The tubes are filled with a mixture of carbon-dioxide, oxygen, and either xenon or argon. Again, this

[7]While individual silicon strip sensors *are* cheaper to produce per unit area, the sheer volume of the SCT makes it one of the most expensive sub-detector within ATLAS.

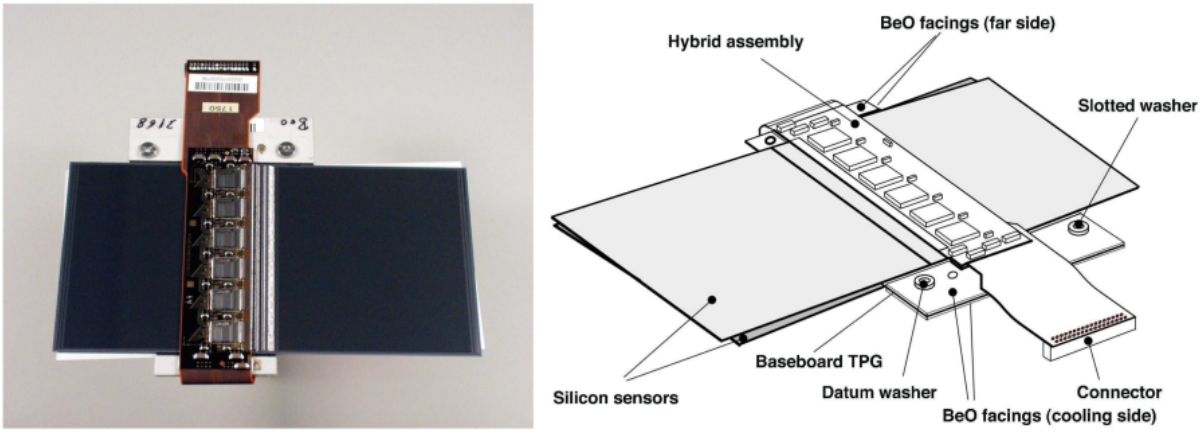


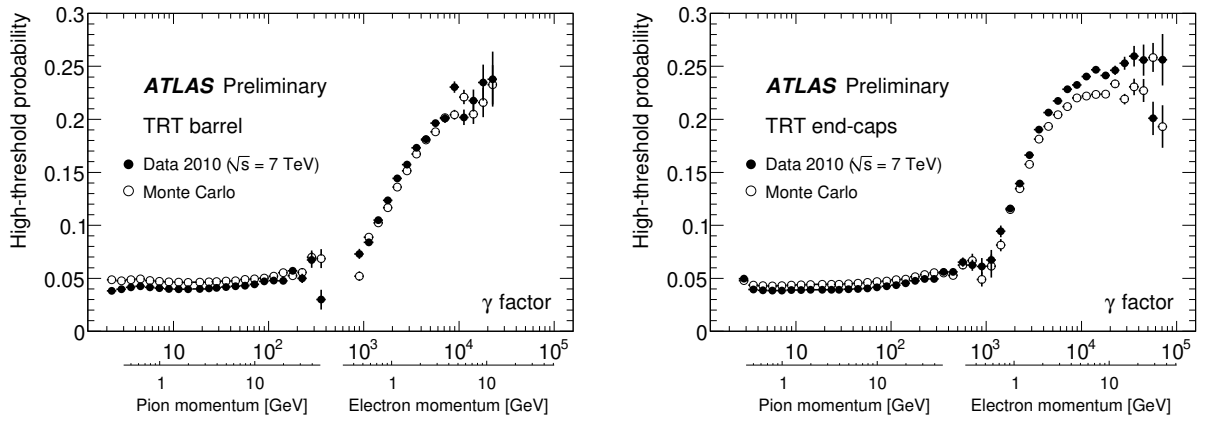
Figure 3.11 – (left) Image and (right) schematic of an SCT barrel module, showing the 40 mrad angle between both sensors [4].

sub-detector has a barrel-like part made up of tubes parallel to the beam, and disk sections with radial tubes. Charged particles intersecting a wire set off a cascade of electrons and ions that are picked up by the anode and the cathode. By analyzing the resulting signal's shape, the position of the cascade's origin between the anode and the cathode can be estimated (with a suitable calibration) with an intrinsic accuracy of 130 μm in the bending plane.

This sub-detector gets its name from the electromagnetic radiation produced by a charged particle crossing a boundary between materials with different dielectric constants. This effect is exploited to enhance the identification of electrons crossing a drift tube boundary, since the interaction of the transition radiation with the gas within the tube will set-off a potentially more energetic cascade than that due to the electron itself. Since the magnitude of this effect depends on the $\gamma = E/m$ factor of the incident particle, by tuning a high threshold, discrimination between electrons and, for example, charged pions can be achieved as can be seen in Figure 3.12. This technique works best when the tubes are filled with a xenon mixture. However, xenon is a rare gas and thus extremely costly to acquire in significant quantities, and high occurrences of leaks lead the collaboration to fill many sectors of the TRT with argon instead.

3.2.3 The Calorimeters

The ATLAS inner detector provides very precise estimations of charged particle's momenta, but it does not provide enough information for particle identification. Moreover, it is not sensitive to the passage of neutral hadronic bound states; a different paradigm



(a) TRT barrel

(b) TRT disks

Figure 3.12 – Different $\gamma = E/m$ factors for charged pions and electrons, due to the electron being much lighter, lead to different amounts of transition radiation when crossing a drift tube boundary, which can be exploited to enhance electron identification.

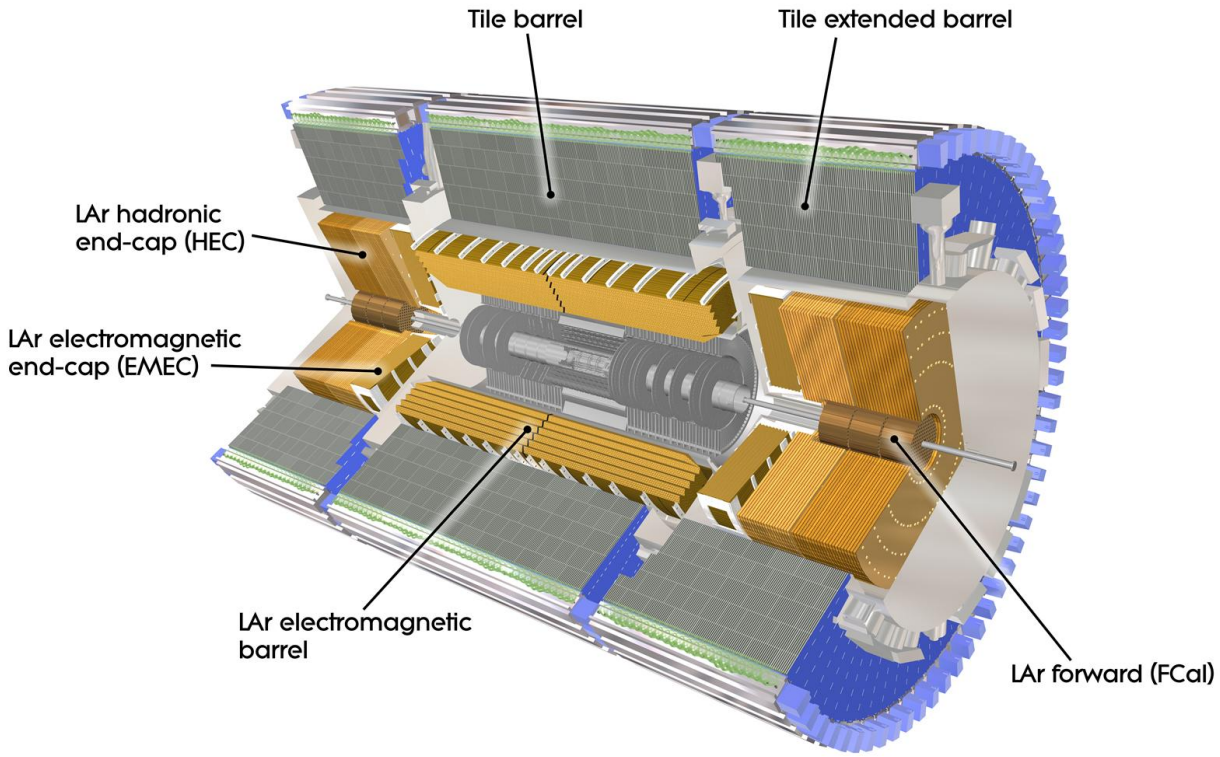


Figure 3.13 – Cut-away view of the ATLAS calorimetry system [135].

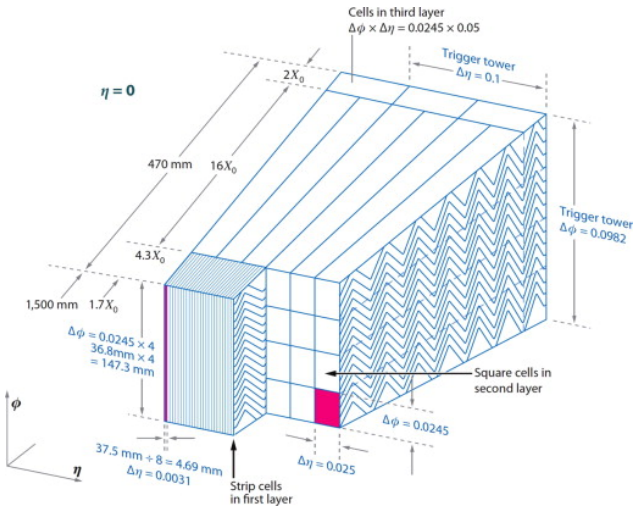


Figure 3.14 – Section of the ATLAS electromagnetic lead/liquid-argon calorimeter, showing its accordion-like geometry [136].

is needed to accurately measure the energy content of such particles, and to provide the needed information to perform accurate particle identification. To do so, a technique known as sampling calorimetry is employed, in which a decay chain is initiated by a particle that passes through dense, "passive" layers; the energy of this decay chain is then sampled in "active" layers. ATLAS employs two types of sampling calorimeters (see Figure 3.13), according to the nature of the particle or bound state being measured: a lead/liquid-argon electromagnetic calorimeter (Section 3.2.3), and a hadronic calorimeter comprising steel/scintillator, copper/liquid-argon, and copper-tungsten/liquid-argon subsystems (Section 3.2.3).

Electromagnetic Calorimeter

As in the case of the inner detector subsystems, the ATLAS electromagnetic calorimeter is arranged in a central barrel and two disk sections, providing energy measurement in the $|\eta| < 3.2$ region. It comprises alternating layers of lead (passive material) and liquid argon (active material) with kapton electrodes, arranged in a accordion-like geometry (as can be seen in Figure 3.14), which allows hermetic coverage around the azimuthal ϕ angle. The $|\eta| < 2.5$ region is more finely segmented than the outer regions, allowing precision measurement of high-p_{T} processes.

The energy resolution of a sampling calorimeter can be decomposed in a stochastic (σ_S) term due to unavoidable fluctuations in energy deposits and to the presence of passive layers and dead material, a noise term (σ_N) accounting for electronic and pile-up effects, and a constant term (σ_C) due to non-uniformities and finite performance of energy recon-

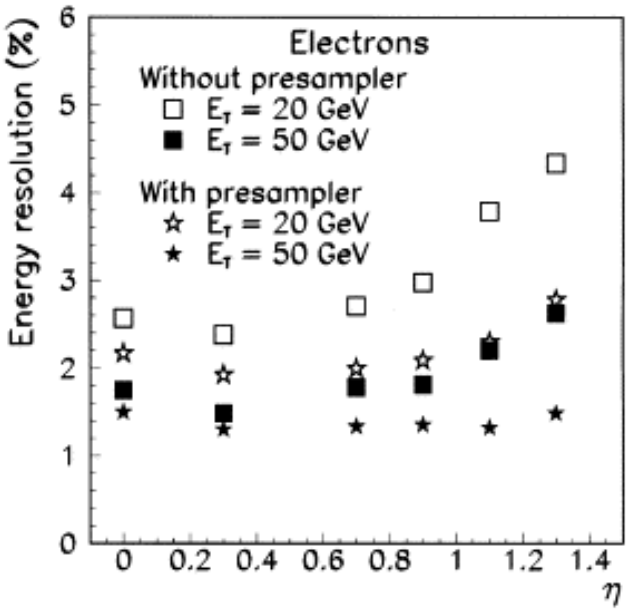


Figure 3.15 – Impact of the LAr pre-sampler on the electron energy resolution as a function of η, estimated by a `GEANT4` simulation. [138].

struction algorithms [137]:

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{\sigma_S}{\sqrt{E}}\right)^2 + \left(\frac{\sigma_N}{E}\right)^2 + \sigma_C^2}. \tag{3.13}$$

In order to minimize the stochastic error term, a thin pre-sampling liquid argon layer covering the $|\eta| < 1.8$ region is installed in front of the calorimeter and accounts for upstream energy losses; its effect on the energy resolution can be seen in Figure 3.15. The total energy resolution of the ATLAS electromagnetic calorimeter is of about $10\%/\sqrt{E} \oplus 0.3\%/E \oplus 0.4\%$ [117].

Hadronic Calorimeter

The hadronic calorimeter is again separated into different parts: a steel/scintillator barrel (known as the "Tile calorimeter", $0 \leq |\eta| < 1.7$), copper/liquid-argon end-caps ($1.5 \leq |\eta| < 3.2$), and copper-tungsten/liquid-argon forward calorimeters ($3.1 \leq |\eta| < 4.9$) installed at the center of the hadronic end-cap cylinders. The steel/scintillator barrel has an extended part that wraps around the end-caps. Together, these subsystems provide energy measurements of hadronic cascades in the $|\eta| < 4.9$ region.

The density and thickness of the passive layers are motivated by the fact that some

hadronic bound states, such as neutrons, are electrically neutral and must interact with the medium to develop a shower that can be measured. Hadronic calorimeters typically have worse energy resolution than electromagnetic calorimeters since hadronic cascades are more complex entities than single incident charged particles: for instance, part of the energy of a cascade, such as nuclear binding energy, cannot be measured; a cascade typically develops both an electromagnetic component in addition to the hadronic component, with a ratio that varies non-linearly and stochastically as a function of the initial particle's energy; and the sensitivity of a hadronic calorimeter to both of these component is not equal and also varies non-linearly as a function of energy. The resolution is limited to $\sigma/E \gtrapprox 50\%/\sqrt{E}$ for non-compensating[8] hadronic calorimeters in general [11].

3.2.4 The Muon spectrometer

The energy lost by a charged particle passing through a certain medium is estimated by the stopping power $\equiv \langle -dE/dx \rangle$ as a function of the medium's atomic density and the particle's $\beta\gamma = p/m$ factor. In a given medium, the stopping power has, roughly, three regimes: starting at a local maximum at low $\beta\gamma$, the stopping power falls with increasing momentum to reach the *minimum of ionization* before rising again at high $\beta\gamma$ due to relativistic effects. As can be seen in Figure 3.16, a muon passing through a dense medium such as copper will be at its minimum ionization when its energy is in the GeV range, which is typical of muons produced in the decay of massive particles such as the W/Z bosons and top quarks. Most other particles are stopped in the calorimeters and therefore almost only muons will make it through, losing a minimal amount of energy in the process. Taking advantage of this fact, the ATLAS detector is completed by a second tracking detector designed specifically to measure the muon momentum, the muon spectrometer, installed around the hadronic calorimeters. It is embedded in an array of three huge toroidal magnets (one barrel and two end-caps), which bend the muon trajectories in the longitudinal plane.

Monitored drift tubes (MDT) cover the $|\eta| < 2.7$ range. The ≈ 30 mm diameter aluminum tubes are filled with an argon/carbon-dioxide mixture and each have a gold-plated tungsten-rhenium axial wire serving as the anode. Each tube has an intrinsic resolution of 80 μm; they are arranged in chambers in groups of three to eight, for an intrinsic resolution of 35 μm per chamber. Additional requirement for a higher rate capacity in the forward ($2 < |\eta| < 2.7$) region lead to the use of Cathode Strip Chambers (CSC), which are multi-wire proportional chambers; these have a resolution of 40 μm in the bending plane

[8] A compensating calorimeter keeps the hadronic and electromagnetic responses ratio h/e equal to unity and will have a resulting better resolution.

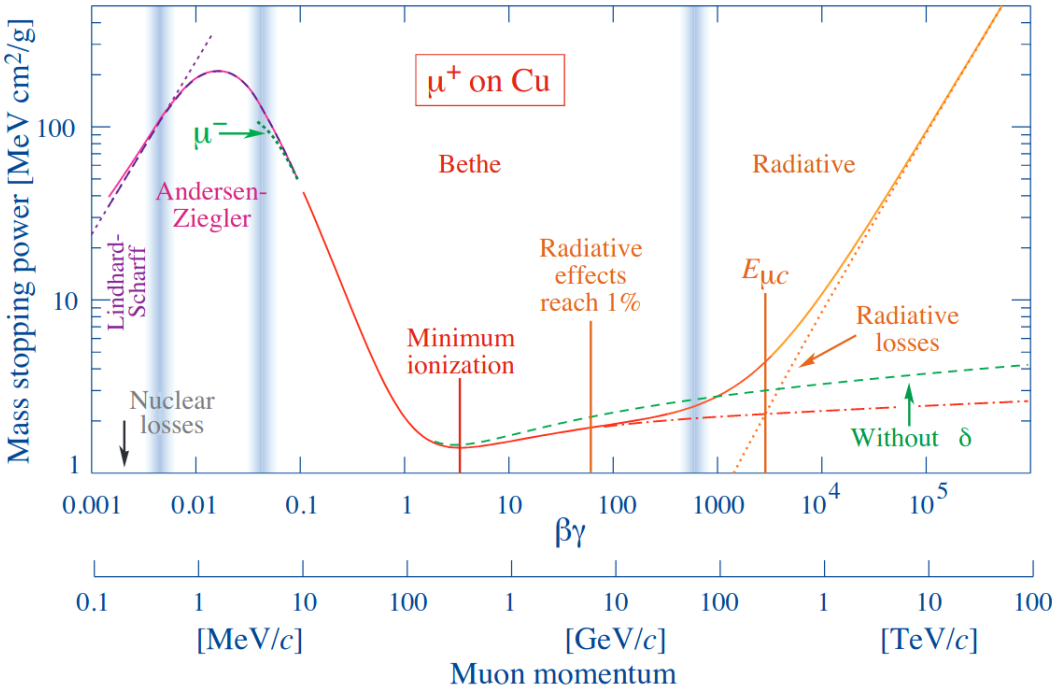


Figure 3.16 – Stopping power ≡ $\langle -dE/dx \rangle$ for a muon in copper [139]. Muons produced in the decay of massive particles such as W and Z bosons and top quarks typically have energy in the GeV range and thus are near the minimum of ionization, explaining why they can travel much farther than other ionizing particles in the detector.

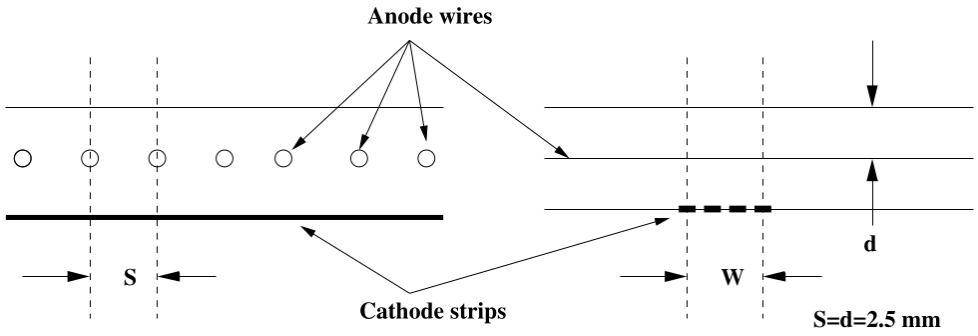


Figure 3.17 – Structure of the cathode strip chambers (CSC), (left) looking down and (right) across the wires, in the bending plane [4].

and of 5 mm in the transverse directions. Their basic layout is shown in Figure 3.17.

The muon spectrometer also includes two subsystems dedicated to triggering readout of events (Section 3.2.5) based on the presence of muons. Resistive plate chambers (RPC) cover the central $|\eta| < 2.4$ range, while thin gap chambers (TGC) cover the endcap region. The orientation of both of these two subsystems can be seen in Figure 3.18

3.2.5 The Trigger and Data Acquisition System

The LHC delivers one bunch crossing each 25 nanoseconds, corresponding to an event rate of 40 MHz and a data rate of about 60 TB/s. Such an astronomical readout rate is ruled out by bandwidth and storage considerations, and so a triggering system has to be implemented to bring down it down to an acceptable level.

The ATLAS trigger and data acquisition (TDAQ) system, shown in Figure 3.19, has two trigger levels: the level 1 (L1), implemented in hardware, identifies "Regions of Interest" (RoI) in η–ϕ spaces using limited information to bring down the event and data rates to about 100 kHz and ≈160 GB/s, respectively. The information from these RoIs are then sent on to the high-level trigger (HLT) farm, an array of computers situated near the detector that perform more complete event analysis and bring down the rates to acceptable levels for permanent readout: ≈ 1.5 kHz and 1.5 GB/s.

We've discussed in Section 3.1.2 how the instantaneous luminosity is not constant within a given LHC fill, but will gradually drop down due to collision losses and beam quality degradation. This is reflected in the ATLAS trigger and data rates, as seen in Figure 3.20. At the beginning of a fill, the raw rate for some triggers is often too high for the buffering system to deal with leading to "dead time", during which events are indiscriminately lost. To mitigate the problem, some triggers are initially *pre-scaled*; a

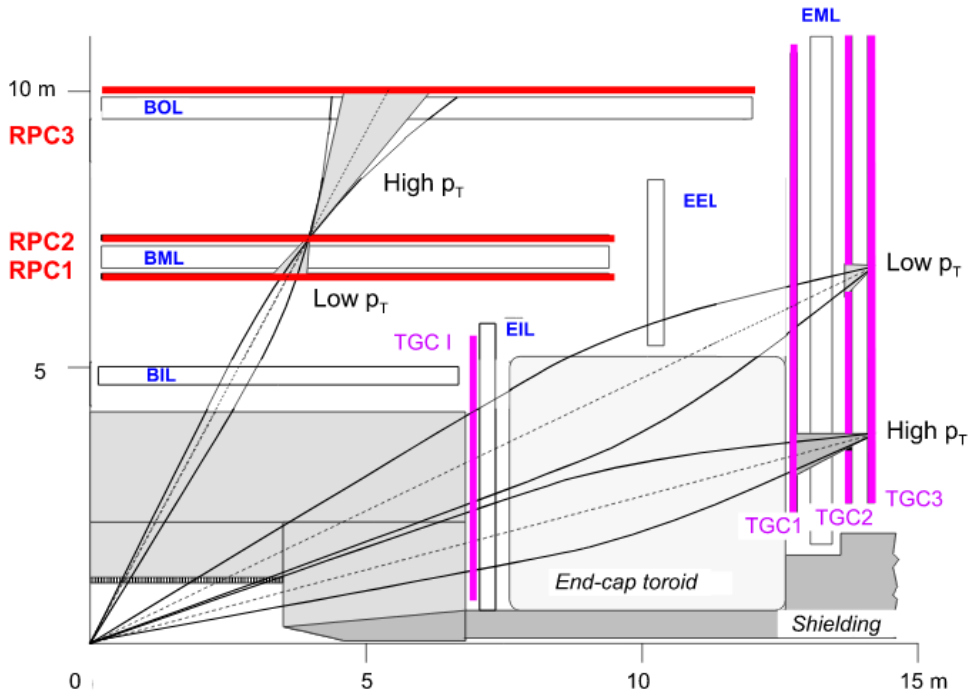


Figure 3.18 – Schematic of the muon trigger subsystem, showing the Resistive Plate Chambers (RPC) covering the central $|\eta| < 2.4$ region and the Thin Gap Chambers (TGC) covering the endcap region [4].

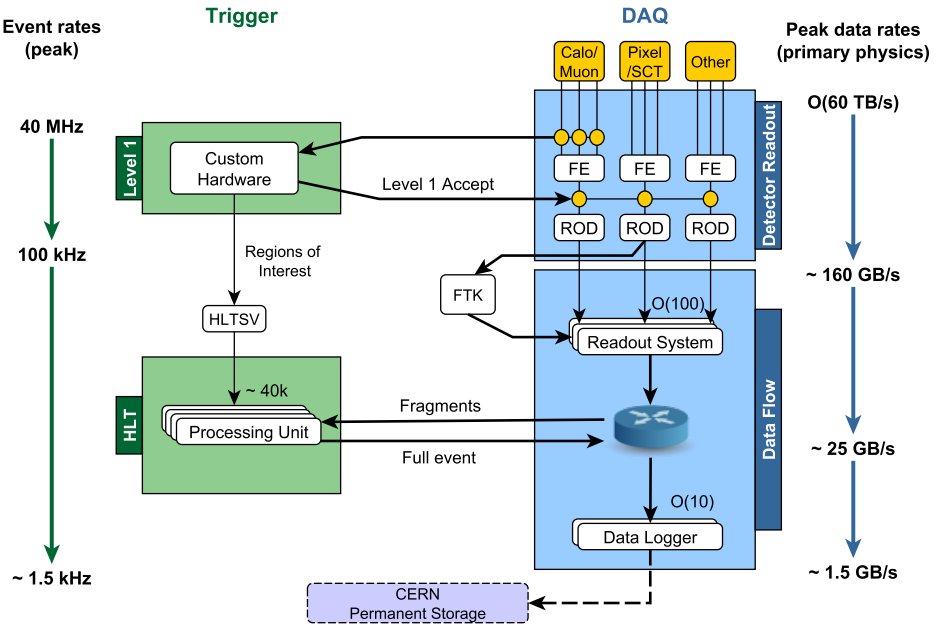
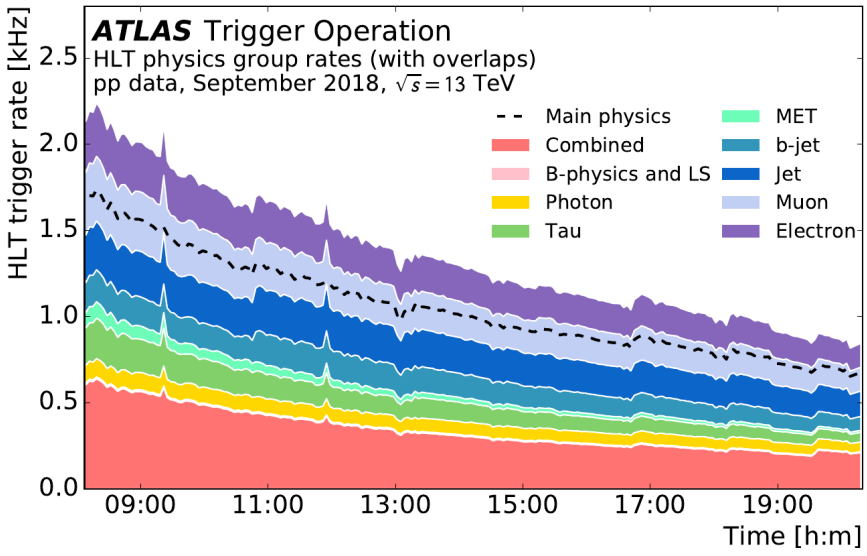


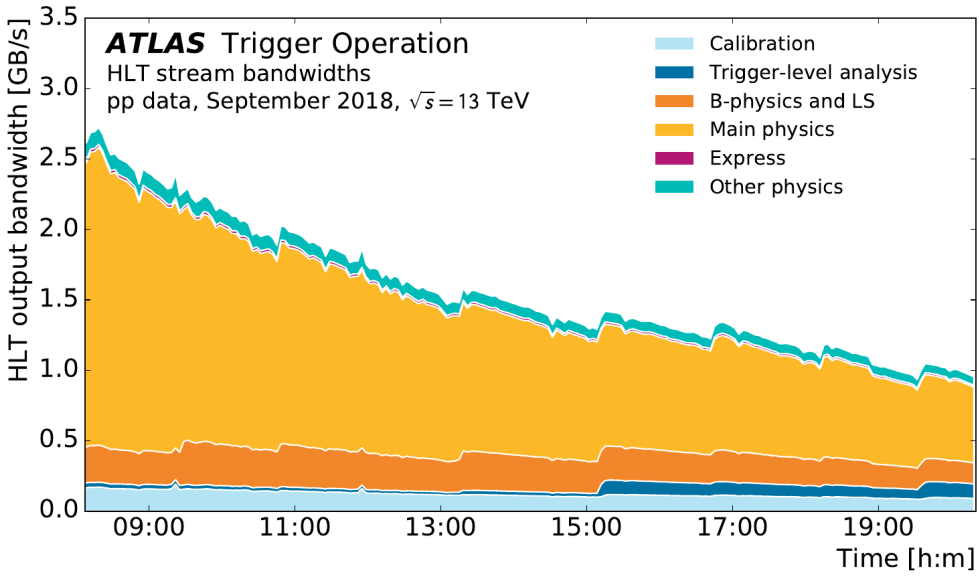
Figure 3.19 – Functional diagram of the ATLAS Trigger and Data Acquisition system in Run 2 showing expected peak event and data rates through each component [140].

pre-scaled trigger will drop a fraction of its accepted events at a rate corresponding to its pre-scale factor that gradually reduces down to zero during single fills.

Both L1- and HLT-level triggers select events based on signatures typical of physics processes of interests for ATLAS analyses. For example, there are triggers based on single, high-p_{T} leptons or jets, or energetic photons. Triggers based on missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) are of particular interest in the context of the search for R-parity conserving (RPC) supersymmetry. Since in LHC collisions the transverse momentum of incoming partons is initially approximately null, by conservation of energy the vector sum of momenta around the beam axis should sum to approximately zero and only gains a non-zero magnitude in the Standard Model due to neutrinos and experimental effects such as mismeasured jets. As discussed in Section 2.1.1, RPC SUSY cascades usually end with a pair of WIMP-like neutralinos, for which the detector is completely transparent. Since neutralinos are hypothesized to be much heavier than neutrinos, their presence can lead to unusually large amounts of E_T^{miss}, a fact that can be exploited to trigger the read-out of such events. Figure 3.21 shows typical efficiency curves for different E_T^{miss} trigger chains; they are all close to 100% efficient for events with more than 200 GeV of $E_{\mathrm{T}}^{\mathrm{miss}}$.



(a)



(b)

Figure 3.20 – HLT (a) trigger and (b) data rates within a single LHC fill. The luminosity drop due to beam losses and quality degradation can clearly be seen, as well as discontinuities corresponding to changes in trigger pre-scale factors.

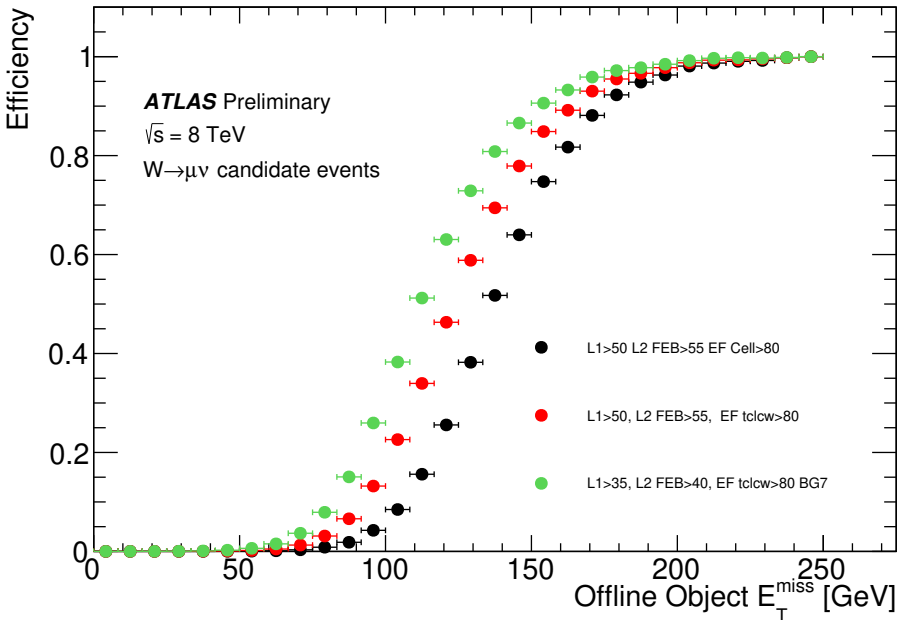


Figure 3.21 – $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency in 8 TeV data for three different trigger chains [141].

3.3 Conclusion

In this chapter, we have seen how the LHC achieves a very high instantaneous luminosity and a record center-of-mass energy of 13 TeV, enabled by the use of superconducting magnets and radio-frequency cavities. We've then moved on to discuss one of the eight LHC detectors, namely the ATLAS detector, used for this thesis: it is composed of a tracking detector (the inner detector) that records charged particle trajectories, enabling estimation of momenta and reconstruction of interaction vertices; an array of electromagnetic and hadronic calorimeters, which measures particle energies; a muon spectrometer, which provides dedicated triggering and estimation of muon momenta; and a sophisticated trigger and data acquisition system that reduces the data and event rates to manageable levels.

In the next chapter, we take a little detour to explore in more depth one aspect of track reconstruction in the ATLAS pixel detector, tracking in dense environments, where the average distance between particles become comparable to the pixel detector sensor resolution. In this context, we will discuss the theory of neural networks, which will be used to carry out a search for supersymmetric particles in Chapter 5.

Chapter 4

Tracking In Dense Environments

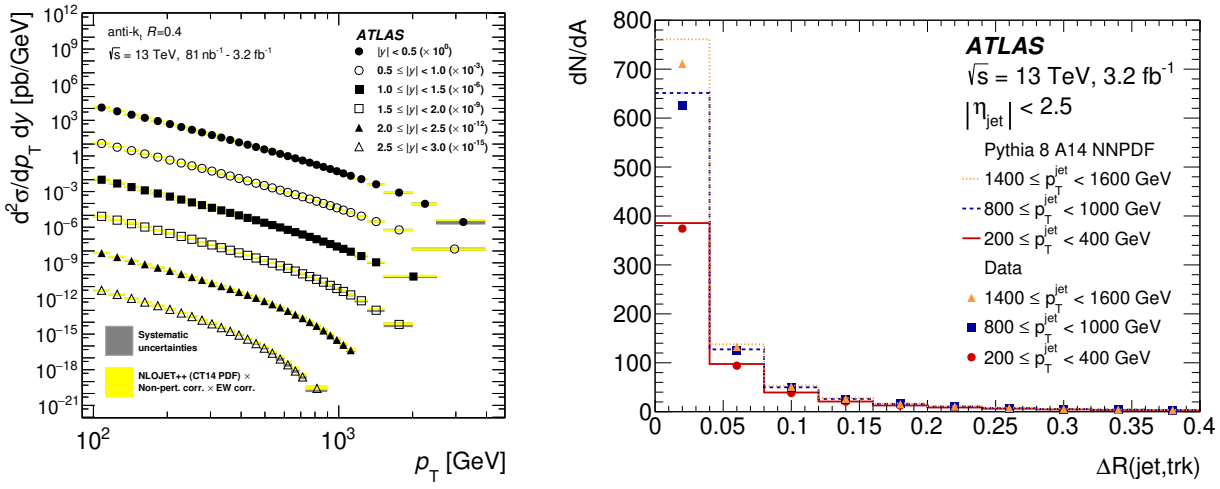
The unprecedented center-of-mass (CoM) energy of collisions at the LHC lead to significant production of high-p_{T} objects such as jets, as seen in Figure 4.1a. Due to the nature of Lorentz transformations, the decay products of highly boosted particles are very collimated, which creates areas of high occupancy in the inner detector (Section 3.2.2) as can be seen in Figure 4.1b. Such areas, referred to as *dense environments*, are especially challenging from a track reconstruction point of view. For instance, when charged particles pass through silicon pixel sensors, multiple pixels are usually illuminated due to the incident angle of the trajectory, charge diffusion, charge drift caused by the axial magnetic field of the ID, or δ-rays[1]. In dense environments, the resulting charge clusters can merge, as visualized in Figure 4.2, and the track reconstruction algorithms must implement additional machinery to recover optimal performance in such cases. This section starts with a brief review of track reconstruction in ATLAS (Section 4.1). In Section 4.2, the theory of neural networks is briefly reviewed before describing the ATLAS pixel clustering neural networks in Section 4.3.

4.1 Track Reconstruction in ATLAS

The task of track reconstruction consists of grouping charged particle energy deposits from the pixel, SCT and TRT subdetectors into proper trajectories from which the impact parameters[2] and angles with respect to the beamline and particle momenta (from the curvature) can be reconstructed. The end goal is to attain a maximum *recall*, that is, finding tracks accounting for all particles in an event, while also keeping the *precision*, or

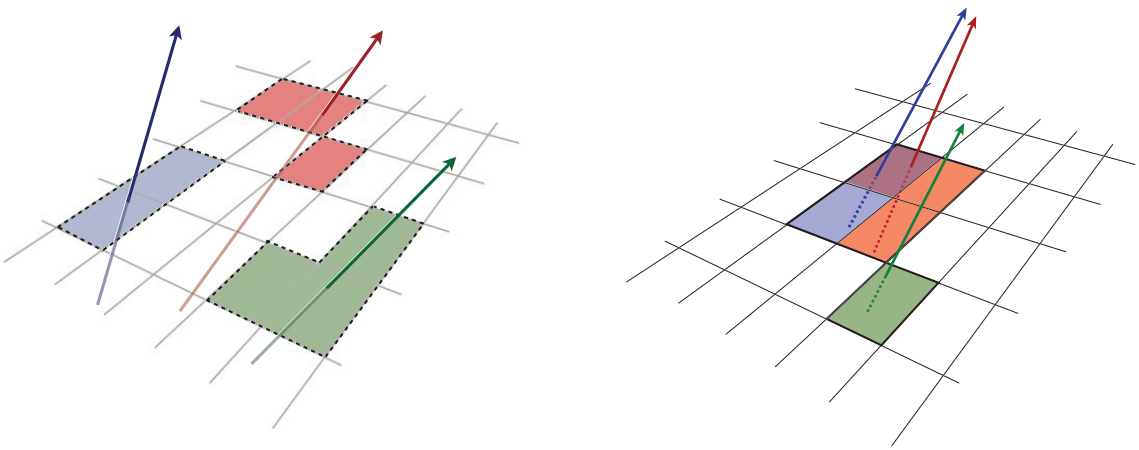
[1] δ-rays correspond to electrons ejected from their orbits by an incident charged particle.

[2] The transverse and longitudinal impact parameters define the perpendicular distance of the closest point between a track and the interaction point.



(a) (b)

Figure 4.1 – (a) Inclusive cross section for anti-k_t $R = 0.4$ jets (see Section 5.2.2) as a function of p_{T} in different absolute rapidity ($|y|$) ranges, showing non-negligible production of very boosted ($p_{\mathrm{T}} \gtrapprox 1$ TeV) jets [142]. (b) Track density per unit angular area as a function of the angular distance from the jet axis in different p_{T} ranges [143].



(a) (b)

Figure 4.2 – (a) Resolved and (b) merged charge clusters in a pixel sensor [143]. The particle trajectories are represented by arrows, while energy deposits from different particles are shown in different colors.

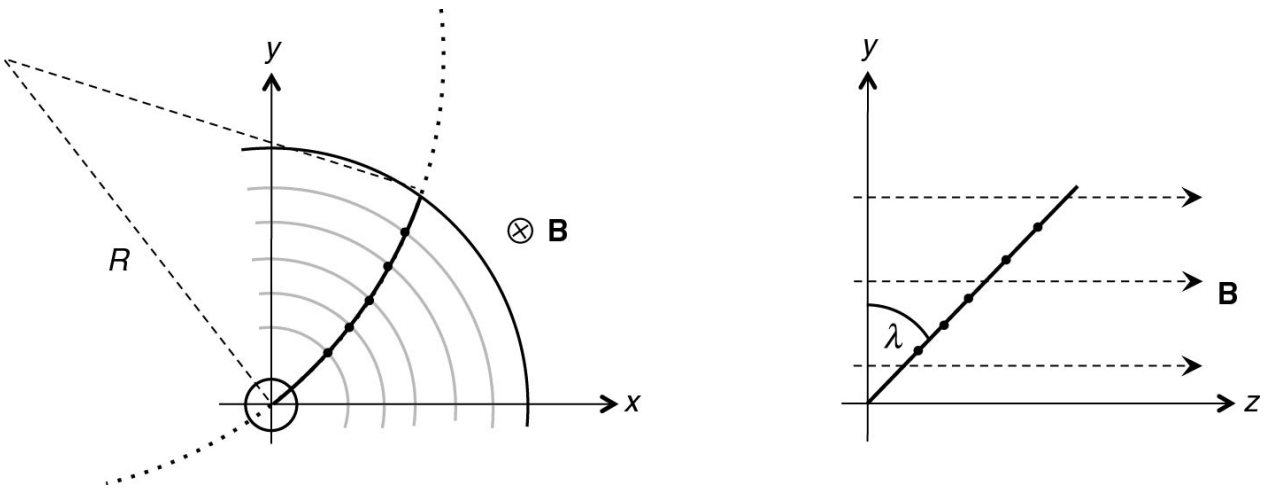


Figure 4.3 – Schematic view of track reconstruction in the (a) bending and (b) longitudinal planes [11].

the fraction of tracks really corresponding to a particle, very high. A schematic view of track reconstruction can be seen in Figure 4.3

The ATLAS track reconstruction chain can be separated into two main stages [143]. The first stage, that of "Track Finding", aims to have maximum recall at the expense of potentially low precision. The second stage, that of "ambiguity solving", aims to restore high precision while keeping a high recall by scoring all tracks and rejecting a low scoring subset that, in the optimal limit, consists only of fake tracks, that is, tracks not corresponding to a particle in the event.

Track finding starts by the identification of *space points*. These are simply 3D coordinates of all the recorded intersection of charge particle trajectories with ID sensor elements. Space points from the two innermost tracking detectors, the pixel and the SCT, are then arranged in sets of three, each defining a *track seed*. To minimize the amount of tracks to be processed downstream, a certain number of criteria are imposed on the seeds, such as p_{T} and impact parameter requirements. A combinatorial Kalman Filter [144] (KF) is then used to build a full set of track candidates. Starting from a seed, the KF iteratively estimates a probability distribution over the position of space points belonging to the track on the closest unincorporated layer, and adds matching hits to its current estimate. Whenever there are more than a single compatible space-point to be added, additional track candidates are created.

In the ambiguity solving stage [143], a score is assigned to each track according to a number of metrics, disfavoring low-p_T tracks as well as tracks with low fit quality. Notably, the ambiguity solver also penalizes tracks that share some clusters with other tracks in order to reduce the amount of duplicates, a strategy motivated by the usually

low probability of having intersecting trajectories. However, as previously discussed, this assumption breaks down in dense environments where the separation between particles is often comparable to the pixel dimensions.

In ATLAS, this problem is fixed with a neural network algorithm; before describing it in depth in Sections 4.3, the theory of neural networks is briefly reviewed in Section 4.2.

4.2 Neural Networks

The method of simplifying the expression of mathematical functions using an assembling of simpler terms is well known to physicists. For example, the k-th order Taylor expansion of a function $f(x)$ approximates it by using the function's derivatives around a known point a:

$$f(x) = \sum_{n=0}^{k} \frac{(x-a)^n}{n!} \frac{d^n}{dx} f(x)|_{x=a} + O(|x-a|^k). \tag{4.1}$$

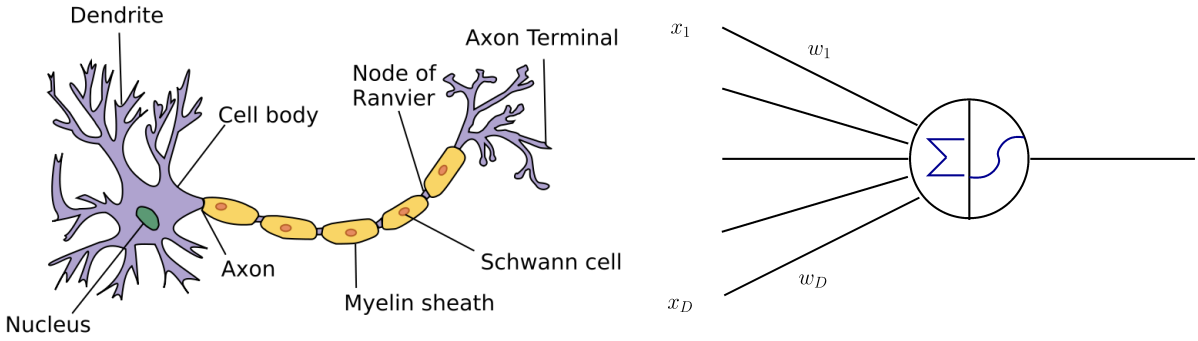
The error term has two important implications: the approximation is valid only in the local neighborhood of a, and therefore knowledge of the function at different points in its domain is needed to ensure good approximation everywhere; also, knowledge of a sufficient number of derivatives of the function is necessary to keep the error small. The latter implication can be particularly limiting: what if we only have access to input/output pairs, and have no knowledge of the derivatives?

Neural networks [7, 145, 146] are a popular class of approximation models that can lift this restriction by expressing the function in a generic form that assumes almost *no* knowledge of the analytical form. The suitability of such networks is guaranteed by the *universal approximation theorem* [147–149]:

Universal Approximation Theorem: Let $f : \mathbb{R}^m \mapsto \mathbb{R}^n$ be a continuous function defined on a closed and bounded subset of $\mathbb{R}$, σ be an element-wise non-polynomial non-constant function, and ϵ be a non-negative, non-zero real number. Then, there exists matrices $W^{(1)}_{m\times h}$ and $W^{(2)}_{h\times n}$ as well as vectors $b^{(1)}_{h\times 1}$ and $b^{(2)}_{n\times 1}$ such that:

$$\tilde{f}(x) = W^{(2)} \cdot \sigma(W^{(1)}x + b^{(1)}) + b^{(2)}, \tag{4.2}$$

$$|\tilde{f}(x) - f(x)| < \epsilon, \quad \forall x. \tag{4.3}$$



(a) Real neuron [150]

(b) Artificial neuron [151]

Figure 4.4 – Analogy between (a) real and (b) artificial neurons: the dendrites carry an electrical signal proportional to that received from upstream neurons, analogously to the artificial neuron's weighted inputs; the outgoing electrical current is a non-linear function of the total signal carried by all the dendrites, analogously to the artificial neuron's non-linear activation.

In plain English: there always exists a parametrization of the model that reaches the required precision level. The basic building block of this model is an artificial neuron, $\sigma(\Sigma_i w_i x_i + b)$, hence the name; the analogy with an anatomical neuron is shown in Figure 4.4.

Although in principle any non-polynomial function could be used, common choices for the σ function (usually called an activation, again by analogy with real neurons) are the sigmoid function, $\sigma(x) = 1/(1 + e^{-\alpha x})$, or the rectifying linear unit $\mathrm{ReLU}(x) = \max(0, x)$, both element-wise functions of the input vector x; they are shown in Figure 4.5. When a classification problem is being solved, it is customary to include a sigmoidal activation on the output layer since it maps all real numbers to the $(0, 1)$ interval as required for a probability. In multi-class situations, the softmax function, a generalization of the sigmoid[3], is often used:

$$S(x)_i = \frac{\sigma(x_i)}{\sum_j \sigma(x_j)}. \tag{4.4}$$

The universal approximation theorem stated above is only concerned with networks having a *single* hidden layer, that is, a single intermediate level of representation between the inputs and the outputs. However, it is possible to stack arbitrarily many hidden layers; the resulting models are often referred to as *deep* networks, which can be up to exponentially more efficient in term of information encoding than single-layer *shallow*

[3]While such output functions ensure the correct properties for probabilities, the network's output is *not* guaranteed to be a probability distribution in the Bayesian sense.

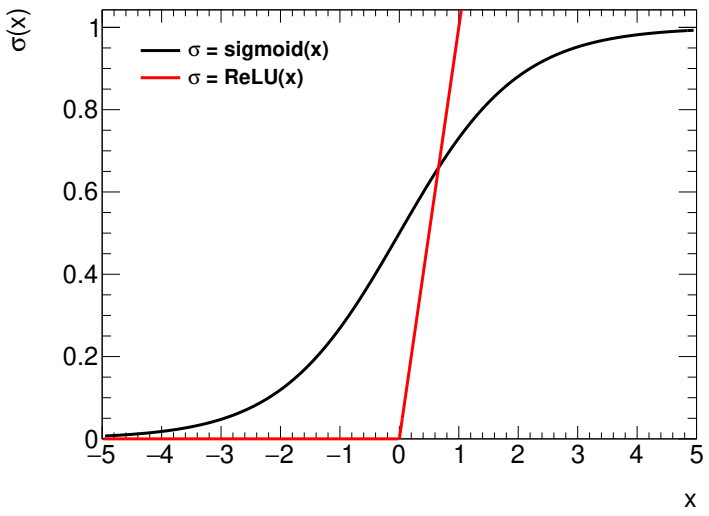


Figure 4.5 – Example of sigmoid and ReLU activations in the $(-5, 5)$ range.

networks [152–155]. Moreover, there is a growing body of evidence asserting that the efficiency gain stems in part from the resulting hierarchy of features, which also enables efficient representation of functions of lower-level (i.e., less feature-engineered) inputs[4]. The use of deep networks trained to construct a hierarchy of features starting from a low-level representation is a textbook example of the *deep learning* paradigm [146].

4.2.1 The training procedure

We have established that neural networks provide a way to efficiently express a large class of functions using very simple arithmetic operations. However, the universal approximation theorem only tells us that this is possible and does not provide a way to obtain the required parameters[5]. To overcome this difficulty, first notice that the network's equation (Eq 4.2) is fully differentiable. Given a network whose full set of parameters we represent by $\theta = (W^{(1)}, b^{(1)}, \ldots, W^{(N_h)}, b^{(N_h)})$ where N_h is the number of layers in the network, a set of input output pairs $X, Y = (x_1, y_1), \ldots, (x_N, y_N)$, and a differentiable *Loss* function $L(X, Y, \theta)$ that quantifies the approximation error, it is possible to iteratively update the parameters using gradient descent. The parameter update from a state τ to the next state $\tau + 1$ is performed using the following equation:

[4]The intuition is that since a shallow network cannot build a hierarchy of features, its single hidden layer would have to be very wide to account for all possibilities if it operates only on low-level features.

[5]In contrast to Taylor's theorem, the universal approximation theorem doesn't say anything about the size of the remaining error, just that it *can* be made arbitrarily small.

$$\theta_{\tau+1} = \theta_\tau - \eta \partial_{\theta_\tau} L(X, Y, \theta_\tau). \tag{4.5}$$

The initial parameter vector, θ_0, is usually randomly sampled [156]. The $\partial_{\theta_\tau} L$ term is computable using recursive application of the chain rule, a procedure known as *backpropagation* [157], and the η user-supplied parameter, called the learning rate, scales the updates. It is customary to estimate the gradient repeatedly on subsets of the datasets, usually called mini-batches (or simply batches) rather than on the whole, and to repeat in many passes (epochs) over the whole dataset. The resulting algorithm is known as *stochastic gradient descent* (SGD), which usually converges faster and produces better parametrizations than when computing the gradient on the whole dataset [156][6].

When the parameter set θ is very large, the gradient $\partial_{\theta_\tau} L$ has correspondingly many dimensions and therefore has many saddle points that can slow down or stall the gradient descent [158, 159]. To overcome this, it is customary to extend the SGD algorithm to include a momentum term:

$$\theta_{\tau+1} = \theta_\tau - \eta \partial_{\theta_\tau} L + \alpha \partial_{\theta_{\tau-1}} L. \tag{4.6}$$

Thus, the update equation now includes an exponentially decaying average of the past gradients, scaled by the α user-supplied momentum parameter, which allow the path in θ-space to avoid getting stuck[7].

Another, more ambitious variant of SGD is the Adaptive Moment (ADAM) algorithm [161]. ADAM also includes a momentum term (the first moment, $m^{(1)}$) but also accounts for the curvature (the second moment, $m^{(2)}$) of the path in parameter-space, the intuition being that the learning rate should be lower when in the vicinity of a local minimum to avoid overshooting it. The accumulated estimates of these moments are computed with exponentially-decaying averages scaled by the ρ_1 and ρ_2 constants, and are also corrected for potential bias in the early stages of training:

[6] This is because the gradients computed on partial datasets have more variance, which helps avoiding overfitting, and the network weights are updated more than once per pass over the whole dataset.

[7] The momentum term also influences the training dynamics in other, less obvious ways; see Reference [160].

$$m_\tau^{(1)} = \rho_1 m_{\tau-1}^{(1)} + (1-\rho_1)\partial_{\theta_\tau} L, \tag{4.7}$$

$$m_\tau^{(2)} = \rho_2 m_{\tau-1}^{(2)} + (1-\rho_2)(\partial_{\theta_\tau} L)^2, \tag{4.8}$$

$$\theta_{\tau+1} = \theta_\tau - \eta \frac{m^{(1)}}{1-\rho_1^\tau} \left(\frac{m_\tau^{(2)}}{1-\rho_2^\tau} \right)^{-1/2}, \tag{4.9}$$

A few more tricks are usually employed in order to speed up the convergence of the training procedure. For example, "standardizing" the input variables to zero mean and unit variance remove the need for the network to learn the characteristic scale of each variable [156]. A related, more sophisticated technique, *batch normalization* [162], re-scales all mini-batches:

$$x \to \gamma \frac{x - \mathrm{E}[x]}{\sqrt{\mathrm{Var}[x]}} + \beta. \tag{4.10}$$

In contrast to standardization, this procedure is usually implemented independently at activation level in each hidden layer of the network rather than only at input level. The neural network equation is still fully differentiable after such substitutions; thus, γ and β are learned at training-time and are part of the θ parameter vector.

While the universal approximation theorem guarantees that parametrizations *exist* for a wide range of functions, there is no guarantee that the SGD algorithm or any of its variant can find them. In fact, the aptly-named "No Free Lunch" theorem states that *all* optimization algorithms have the same performance when averaged over all possible problems [163]. In practice however, deep networks trained with SGD or related algorithms have historically been extremely successful in a wide range of applications [7].

There is a final caveat with respect to the training procedure: it only tries to find definite values for the trainable θ parameters; all parameters that are not included therein are called *hyperparameters* and are user-supplied. This includes, for instance, the dimensions of the parameter matrices in θ, and the scaling parameters $\eta, \alpha, \rho_1, \rho_2, \ldots$ of SGD and related algorithms. A simple solution to this conundrum is to perform a random search over all included hyperparameters, a procedure that can be exponentially more efficient than a grid scan over all of phase-space [164]. The resulting randomly-generated models must be compared on a held-out dataset or with cross-validation in order to avoid introducing bias at the model selection stage [165, 166]

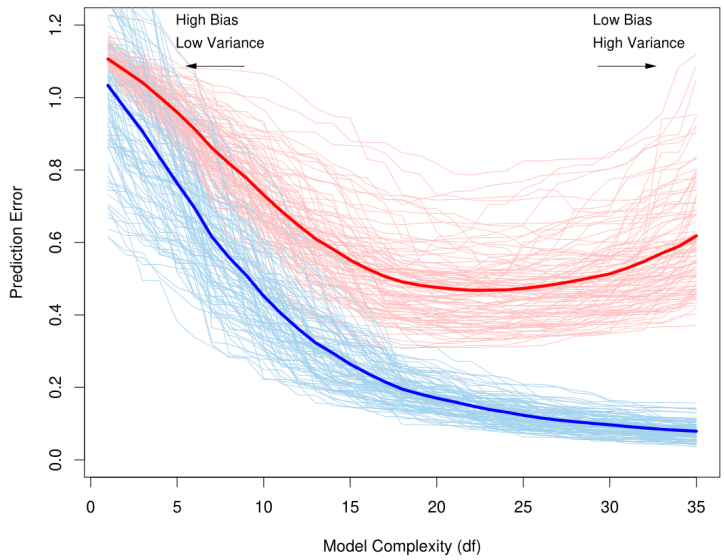


Figure 4.6 – Bias–variance decomposition as a function of model complexity [165]. The faint blue (red) curves show the training (test) set performance as a function of model complexity for many different samplings, and the bold curves show the averaged values. In this figure, the complexity is quantified by the number of degrees of freedom, but the same behavior would be seen if model capacity were considered instead. In the training set, both bias and variance drop significantly as a function of model complexity while in the test set, the variance grows with complexity but the bias drops and then rises again once the model enters the overfitting regime.

4.2.2 Generalization

The optimization procedure outlined in the previous section only ensures that the model will perform well on the dataset it was trained on. What about the performance on unseen data, which is the end-goal of most applications? In more precise terminology, in addition to obtaining a good parametrization on the training set, the difference in performance between seen and unseen data, called the *generalization error*, must also be minimized. Provided the training and test sets are sampled from the same distribution, this error can be decomposed in three components [165]:

$$\text{Generalization Error} = \text{Irreducible} + \text{Bias}^2 + \text{Variance}. \tag{4.11}$$

The irreducible component, also known as the Bayes error rate, is caused by ambiguities fundamental to the task at hand and represent the floor of the generalization error. The relative importance of the last two terms, or the Bias–Variance trade-off, is related to the concept of *capacity*, that is, the size of the function space that can be represented by a given

model as can be seen in Figure 4.6. On the bias-dominated end of the trade-off we find low-capacity models, which can approximate a relatively small amount of functions; their average performance is not optimal, but it does not vary much when used on unseen data. On the variance-dominated end of the trade-off, we find high capacity models, which can represent a large amount of functions and thus run the risk of learning statistical fluctuations in the training dataset; therefore, their performance will tend to vary much more when the data is changed.

In the case of neural networks, model capacity scales with the number of trainable parameters as well as the total number of gradient updates performed in training. This points to three different ways of controlling the bias–variance trade-off: restrict the amount of trainable parameters, restrict the possible values that these parameters can take, or restrict the training time. The first method runs the risk of being too restrictive since it caps the "storage space" of the model, and the minimal amount of information needed to solve a given problem is usually not known beforehand. The second technique attempts to restrict solution-space in a dynamical way, for example penalizing large weight values by modifying the loss function:

$$L \longrightarrow L + \lambda_1|\theta| + \lambda_2|\theta|^2. \tag{4.12}$$

The *weight decay* $\lambda_{1,2}$ variables are additional hyperparameters[8] that penalize large weights by an amount proportional to their absolute and squared values, respectively.

The third technique, known as *early-stopping*, keeps an estimate of the generalization error during the training by holding-out a fraction of the training dataset; the training is simply stopped when the generalization error stops improving. This deceptively simple technique is extremely powerful and partially negates overfitting introduced in other parts of the model, most of the time obviating the need for fine-tuning of many hyperparameters [167].

While in practice, these regularization methods generally succeeds at minimizing the amount of overfitting, there is still no free lunch; a well regularized model can still have a large generalization error if the training data is of poor quality [168] or if it is sampled from a significantly different distribution than the test data.

[8]The weight decay variables have to be hyperparameters, else the training procedure would always set them to zero.

4.3 Pixel Clustering Neural Networks

As previously mentionned, the problems caused by merged charge clusters in the ATLAS pixel detector are solved by a neural network algorithm. More precisely, three different sets of neural networks are implemented [169], each operating on single charge clusters:

- The "number" network, a single neural network that classifies clusters as having been produced by one, two, or three or more particles;
- The "position" networks, a set of three neural networks (one for each number class) that estimate the positions of the intersection points in the silicon sensor. The positions are measured in a frame of reference local to the pixel sensor considered, in which the local x and y directions correspond to the transverse and longitudinal directions with respect to the beam line, respectively;
- The "error" networks, a set of six networks (one for each number class for each coordinate) that estimates the uncertainty on the previous measurement.

The simulated training dataset is composed of multijet events generated using Pythia 8.186 [170] with the A14 set of tuned parameters [171] and the NNPDF2.3LO parton distribution function set [31]. In order to obtain a large fraction of high-p_{T} jets and a consequently large amount of merged clusters, a filter keeping only truth-level jets with $1.8 \leq p_{\mathrm{T}} < 2.5$ TeV is applied before handing over to a detailed `GEANT4`-based simulation of the ATLAS detector [172, 173].

12 (5) million clusters are retained to train (validate) each of the ten neural networks. For the number network, the training dataset is sampled such that the fraction of 1–, 2– and 3-or-more particle clusters are of 22%, 26%, and 52%, respectively.

Inspired by the deep learning paradigm, the input set is fairly low-level:

- A 7×7 matrix[9] with entries corresponding to individual pixels (see Figure 4.2), with the charge centroid in the middle;
- A length-7 vector of pixel dimensions in the local y direction, since a fraction of each pixel module's columns are read-out in the same channel, thereby increasing the effective size of the pixels for those merged columns;
- A binary variable encoding the inner detector region (endcap or barrel);

[9]The end goal being estimating positions of particles in the inner detector, this is in principle a three dimensional problem; however, the radial position is completely determined by the coordinates of the silicon sensor's surface, which are precisely known.

- An integer variable representing the cylinder (barrel) or disc (endcap) number;
- Angles of incidence of the track candidate being scored;
- (Error networks only) Hit position estimates for the current cluster.

The two angles of incidence significantly improve the performance of these neural networks, since they help recognize large clusters due to many particles from large single-particle clusters due to high incident angles [169]. To decouple the pixel clustering neural network training from the rest of the track reconstruction chain, the true angle of incidence of the MC-generated particle is currently used as a proxy for the actual track-based measurements. This is believed to be safe since the neural network's performance has been observed to be robust with respect to small perturbations of these angles [174].

The hyperparameters used are listed in Table 4.1. An optimization pass using a random search over hyperparameter combinations [164] has shown that these relatively small, shallow networks are a good trade-off between runtime and performance.

Hyperparameter	Number	Position	Error
Structure	(60)-25-20-(3)	(60)-40-20-(2/4/6)	(62/64/66)-15-10-(30/50/60)
Hidden activation	Sigmoid	Sigmoid	Sigmoid
Output activation	Sigmoid	Identity	Sigmoid
Learning rate (η)	0.08	0.04	0.3
weight decay (λ_2)	10^{-7}	10^{-7}	10^{-6}
Momentum (α)	0.4	0.3	0.7
Minibatch size	60	30	50
Loss function	cross-entropy	mean squared error	cross-entropy

Table 4.1 – Hyperparameters used to train the three sets of neural networks. In the Structure row, the numbers in parenthesis denote the input and output layer sizes (expressed in number of neurons), with numbers separated by slashes corresponding to different sizes in datasets with 1, 2 or 3 particles per cluster, respectively, while the numbers in-between represent the hidden layer sizes.

4.3.1 Number network

The number network is trained to classify charge clusters as comprising one, two, or three or more particles. Since it is a classification task, its performance is better visualized by considering receiver-operating-characteristic (ROC) curves [175]. The network outputs three probabilities for each cluster, and the only constraint is that these must sum to unity. Therefore, there are two output degrees of freedom for each cluster and there is

no one-to-one correspondence between any two given probability bin, in contrast to a two-class scenario [176]. This means that the $2 \times 3 = 6$ ROC curves of Figure 4.7 must be considered to fully determine the performance. In these figures, we see that the number network performs very well in recognizing single-particle clusters from multiple-particle clusters, which is the most important case with respect to tracking performance; it has more difficulty disambiguating the two- vs three-or-more particle cases.

In the ATLAS track reconstruction algorithm, the output probability scores for the two- or three-or-more particle classes, P_2 and P_3, are turned into actual classifications via the following decision rule:

- If $P_2 < 60\%$ and $P_3 < 20\%$: classify as 1-particle cluster
- If $P_2 > 60\%$ and $P_3 < 20\%$: classify as 2-particles cluster
- If $P_3 > 20\%$: classify as ≥ 3-particles cluster

This rule was tuned with a grid scan over all possible combinations, attempting to optimize the trade-off between maximizing the efficiency of reconstructing real tracks inside jets [178] and minimizing the production of fake tracks from shared clusters.

4.3.2 Position networks

After application of the number network, the three position-estimating networks, one for each of the three particle multiplicity classes, implement a solution to a regression problem; they output a 2-dimensional position value for each particle in a given cluster. Thus, their performances are better visualized in Figure 4.8, which show the distribution of the difference between the true and estimated positions, or residuals, in test sets of five million clusters each. In an ideal setting, the errors would be correctly described by single-mode Gaussian distributions. However some cases follow different distributions than the rest, such as 1-particle clusters comprising a single illuminated pixel (Figures 4.8a and 4.8b), leading to non-normally distributed residuals. A first estimation of the position resolution can still be obtained by considering the full width at half-maximum (FWHM) of these distributions, showing that the model yields about three times the intrinsic resolutions (10 μm and 115 μm in local x and y, respectively) in the most difficult case ($\geq$3-particle clusters).

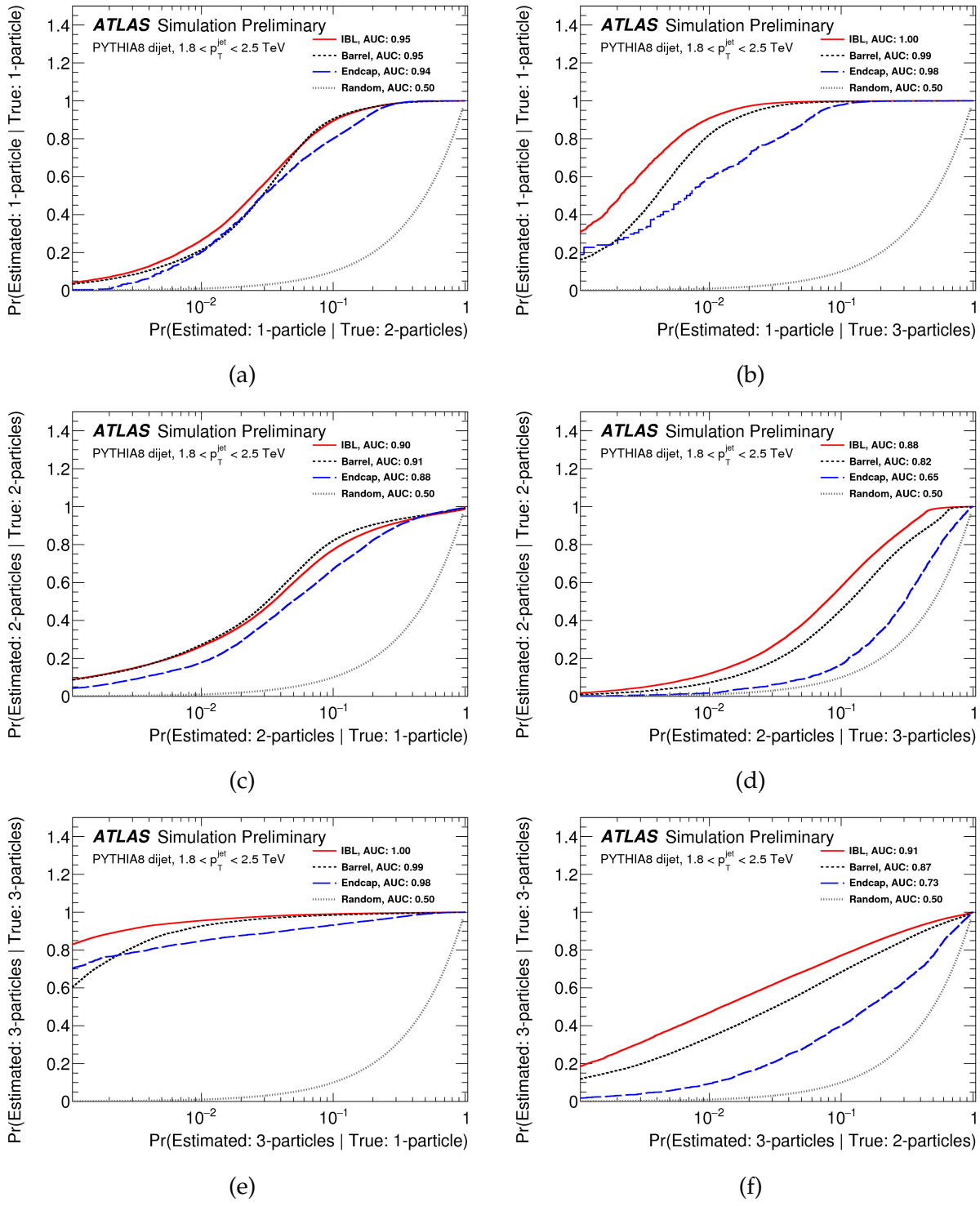


Figure 4.7 – Pairwise receiver operating characteristic (ROC) curves for the network used to estimate the particle multiplicity. (a) 1-particle vs 2-particles clusters. (b) 1-particle vs ≥ 3-particle clusters. (c) 2-particles vs 1-particle clusters. (d) 2-particles vs ≥ 3-particle clusters. (e) ≥ 3-particle vs 1-particle clusters. (f) ≥ 3-particle vs 2-particles clusters. In these figures, curves nearer to the top-left corner represent better performances, and the small-dashed lines correspond to a random classifier with variable bias and constitutes a universal baseline [177].

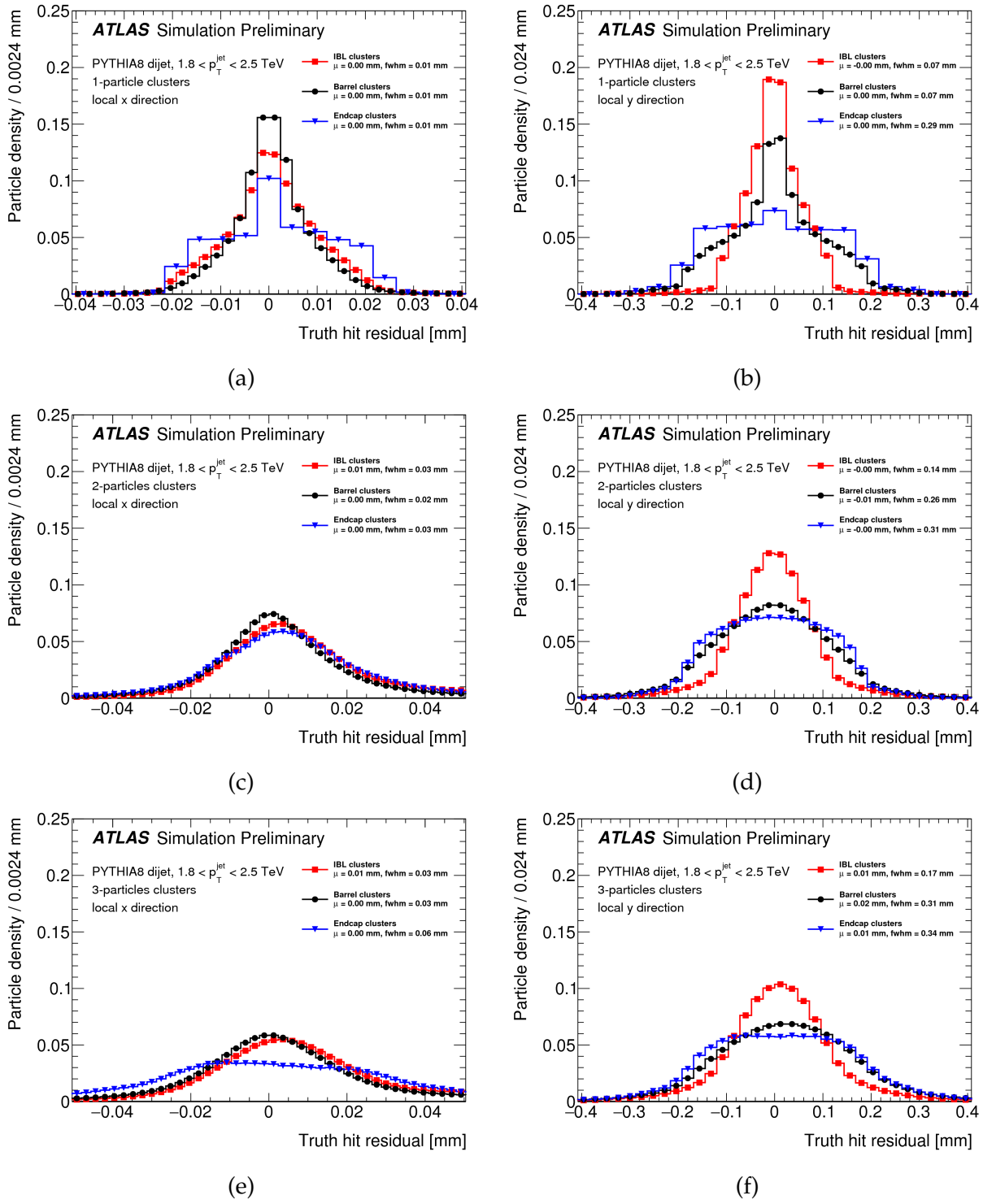


Figure 4.8 – Difference between the neural network position estimation and the true hit position in the (left) local x and (right) local y directions for true (a), (b) 1-particle, (c), (d) 2-particles and (e), (f) 3-particles clusters. All sample means have negligible uncertainties while the full width at half minimum values have relative uncertainties of less than 5%. The different x axis ranges are due to the differing pixel sizes of 50 μm and 400 μm (250 μm in the IBL) in the local x and y directions, respectively [177].

4.3.3 Error networks

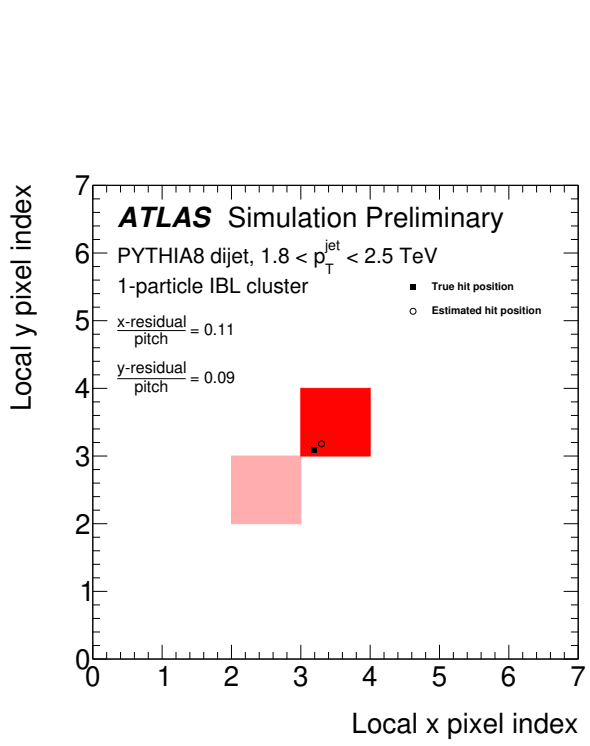
The position estimation networks only furnishes a point estimate of the position of each particle within a cluster while the ATLAS tracking algorithms needs estimates of the position's variance in each direction in order to perform χ^2 fits of a particle's trajectory. Since these quantities are undetermined, the networks try to reconstruct a binned probability distribution over possible residuals for each particles in a given cluster in a given direction; point estimates of the variance are then obtained by taking the root-mean-square (rms) of the resulting distributions. Since the position estimation only yields point estimates of the residual distributions, which are then used to learn the full distributions, this is an instance of semi-supervised learning. An example of the error estimation task is shown in Figure 4.9

The error network's performance is best visualized in the so-called pull distributions of Figure 4.10, which show the distribution of the residual divided by the estimated error in each number/direction pair. If the residuals were perfectly Gaussian and the error perfectly estimated, such pull distributions would always have zero mean and unit variance, which is not always the case in practice. To improve this, the ATLAS collaboration is considering replacing the position and error networks by Mixture Density Networks (MDN) [179, 180], which use a mixture-of-Gaussian probability densities as a learning objective, allowing for modeling of arbitrary distributions [151].

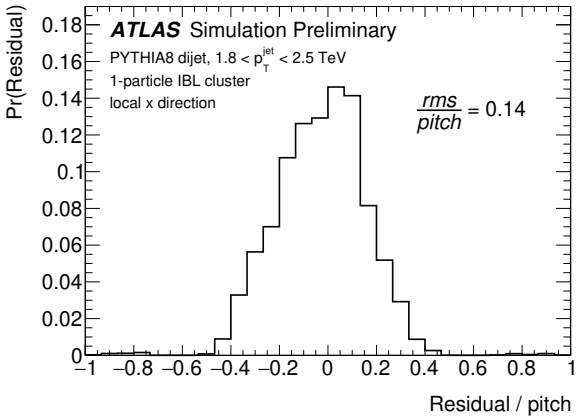
4.4 Conclusion

In this chapter, we've explored in detail a particular aspect of track reconstruction in ATLAS, that of accurately recognizing charge clusters in the pixel detector as being produced by one, two or three or more particles, and of estimating the contributing particle's positions within these clusters. This is done by three sets of neural networks, which exemplifies the power and versatility of such machine learning models.

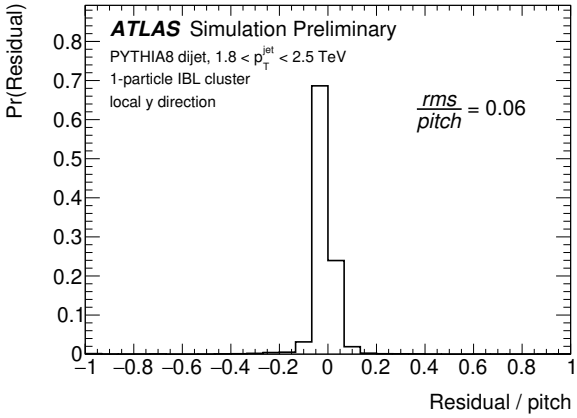
In the next section, we finally move on to discussing a search for supersymmetric particles in ATLAS data in which neural networks play a central part: that of actually labeling events as containing certain kinds of supersymmetric particles or not.



(a)



(b)



(c)

Figure 4.9 – Example use case of the neural network used to estimate the uncertainty for a 1-particle IBL cluster. (a) 1-particle cluster with true hit position marked by the full square and hit position estimated by the neural network marked by the open circle. The cluster is fed to the two neural networks that estimate the probability distribution of this cluster's residual in the (b) local x and (c) y directions, respectively. The neural networks output node are directly mapped to bins of the residual distributions, and the *rms* of these distributions are used as point estimates of the uncertainties. In order to compare the performance in both directions, the residuals and *rms* values are divided by the pitches (50 μm and 250 μm in the local x and y directions, respectively) [177].

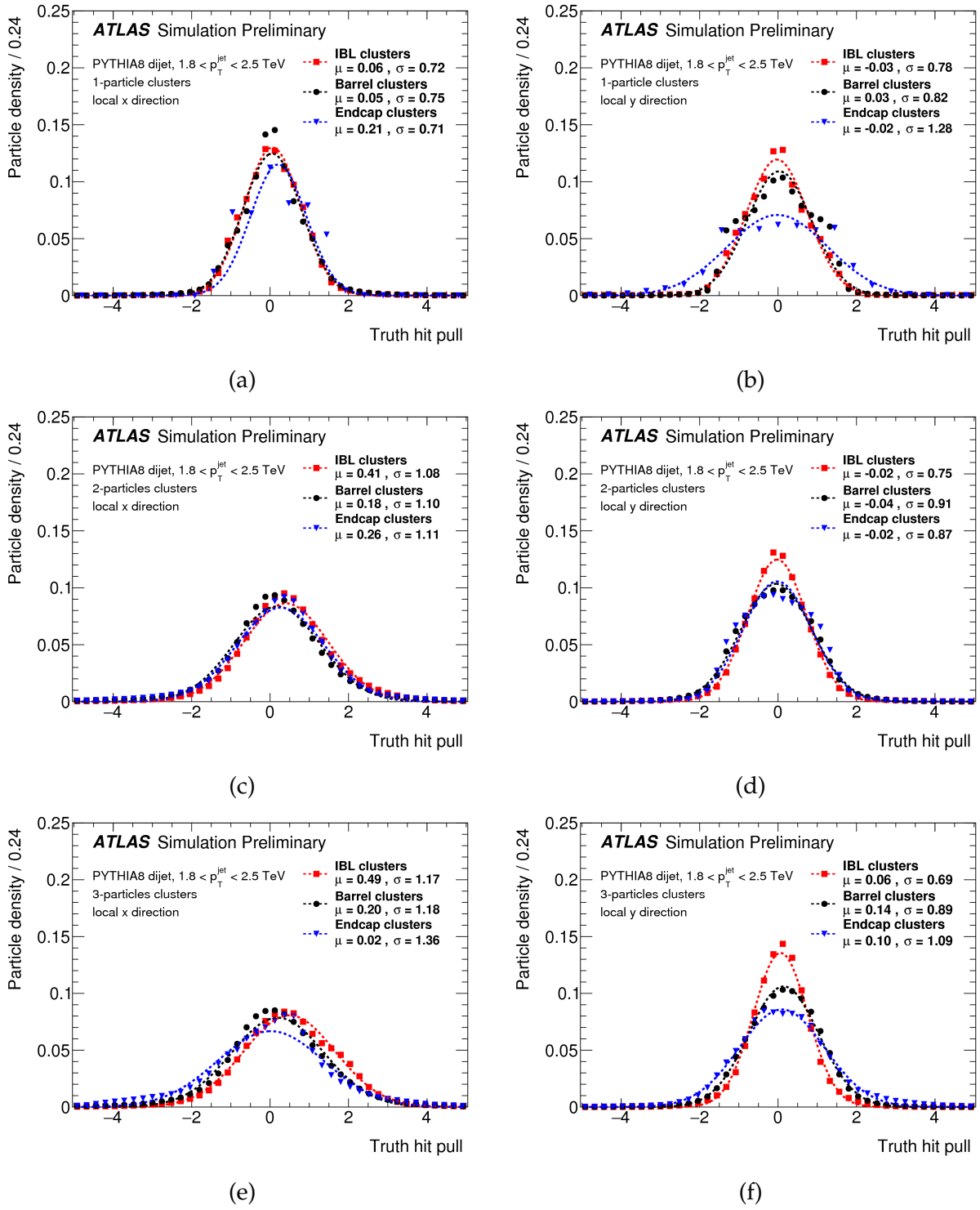


Figure 4.10 – Difference between the neural network position estimation and the true hit position divided by the estimated uncertainty in the (left) local x and (right) local y directions for true (a), (b) 1-particle, (c), (d) 2-particles and (e), (f) 3-particles clusters. The means and standard deviations are estimated with truncated Gaussian fits, which are represented as dashed lines. All means and standard deviations have negligible uncertainties [177].

Chapter 5

Search for supersymmetry in events with many b-jets and significant missing transverse momentum

In this chapter, we use all the theoretical knowledge accrued so far in this thesis to perform an actual search for hypothesized supersymmetric particles. Let's take a moment and briefly review the motivations for this particular search. In Chapter 1, we've touched upon two unsolved problems with the Standard Model; namely, the Higgs mass-related naturalness problem, and dark matter. In Chapter 2 however, we've seen that a solution to both problems arise in R-parity conserving supersymmetric extensions of the SM.

Now, which sparticles exactly do we want to search for? The neutralino, $\tilde{\chi}_1^0$, is an obvious candidate since it provides a perfect WIMP dark matter candidate. Higgs mass considerations lead us to also pay close attention to the gluino and to third-generation squarks. Moreover, to ease the interpretation of results, we want to limit our benchmark models to two free parameters, and cross-section considerations, as shown in Figure 5.1, lead us to consider the gluino as our other primary target, with the sbottom and stop quarks being considered as off-shell; these simplified models, referred to as the Gbb and Gtt models, can be visualized in Figure 5.2.

As mentioned in Section 2.2.2, events from these models lead to final states with striking signatures in the detector. For instance, all-hadronic Gtt events will have 12 jets, four of which originate from b-quarks, while leptonic events can have up to four lepton/neutrino pairs. Boosted W bosons or top quarks can be reconstructed as single large-radius jets, as seen in Figure 5.3. In both grids, the two neutralinos and the potential neutrinos from the W decays in each event are significant sources of missing transverse momentum. The mass splitting has a significant effect on the final state kinematics; areas of the grids where $m_{\tilde{g}} \gg m_{\tilde{\chi}_1^0}$ are characterized by more boosted particles and more $E_{\mathrm{T}}^{\mathrm{miss}}$, since the large

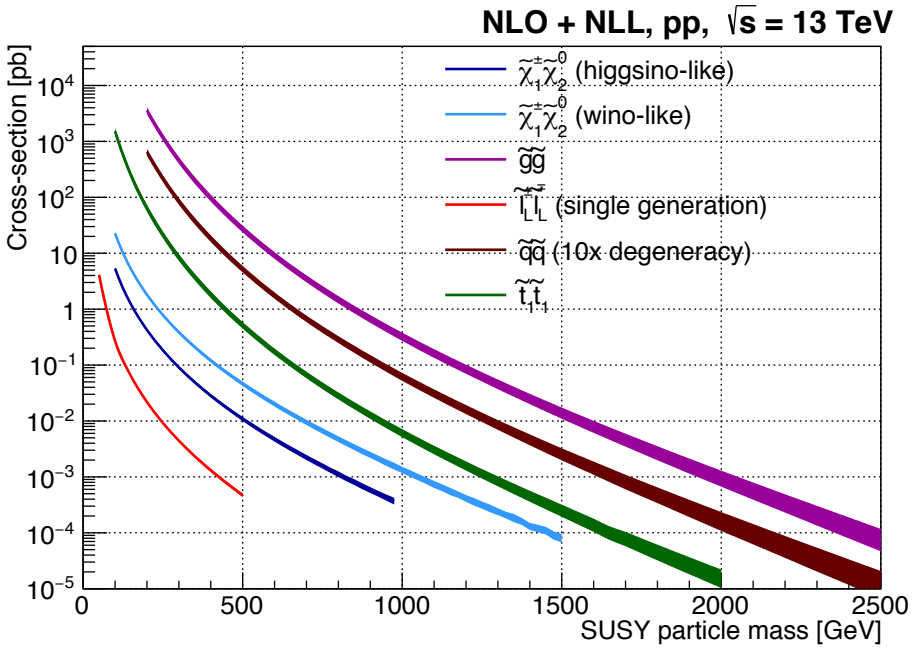


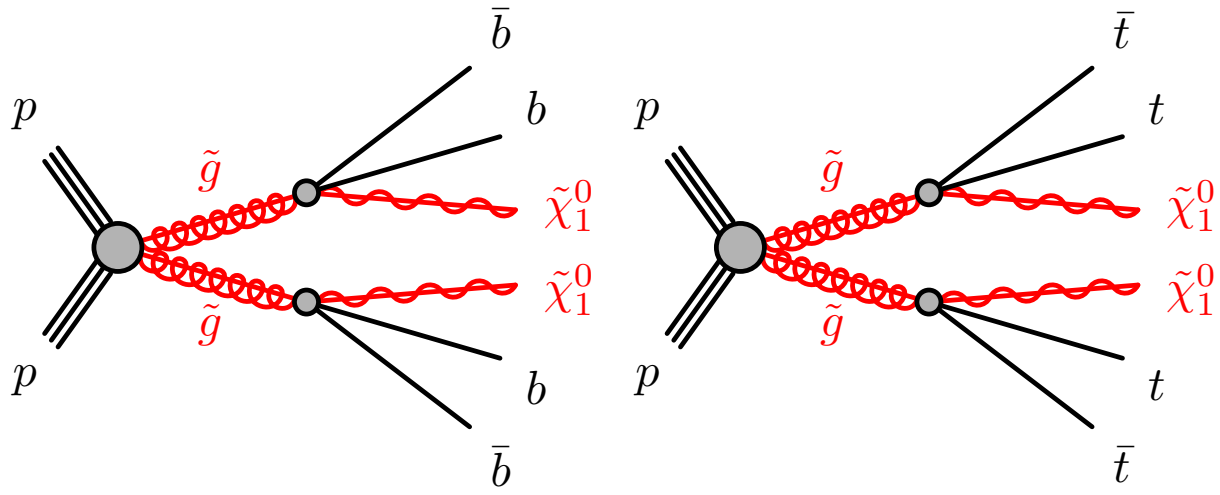
Figure 5.1 – Cross-sections for various scenarios of SUSY particle pair production, showing that gluino pair production is a more luminous search target than, for instance, direct squark production [181].

splitting translates to high available kinematic energy for the decay products, and are generally easier to search for than the compressed regions, where $m_{\tilde{g}} \approx m_{\tilde{\chi}_1^0}$. For instance, in this regime the neutralinos are produced nearly at rest and so the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is more SM-like.

Details on the generated signal samples as well as Standard Model backgrounds and the ATLAS data used in this search are presented in Section 5.1, while the reconstruction of these data into meaningful physics objects is reviewed in Section 5.2. Inspired by the neural networks discussion of Chapter 4, we detail how to train such models to recognize SUSY events from the SM bulk in Section 5.4. Finally, the statistical data analysis and its result are presented in Sections 5.5 and 5.6, respectively.

5.1 Signals, backgrounds, and ATLAS data

Samples produced by Monte Carlo particle collision generators [183] are used to optimize the search strategy and provide the result hypotheses. The generation of signal samples proceeds in two steps. First, matrix elements for $p + p \rightarrow \tilde{g} + \tilde{g}$ events are computed using version 2.3.3 of the `MadGraph5_aMC@NLO` package [184] and the NNPDF 2.3 parton distribution function set [185]; gluino decays to top or bottom quark pairs and a neutralino are subsequently handled by the `Pythia` version 8.212 program [186, 187], considering



(a) $\tilde{g} \to b + b + \tilde{\chi}_1^0$ (b) $\tilde{g} \to t + t + \tilde{\chi}_1^0$

Figure 5.2 – Simplified models of pair-produced gluinos decaying to (a) bottom or (b) top quarks and neutralinos via off-shell stops or sbottoms, colloquially known as the Gbb and Gtt models, respectively.

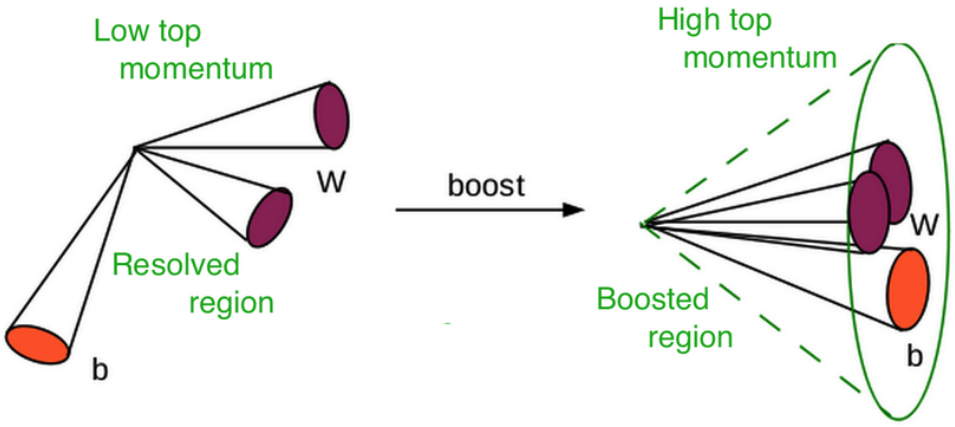


Figure 5.3 – The decay products of the top quark become very collimated in the laboratory frame when its rest frame has a large Lorentz boost [182]. In such cases, the top decay can be reconstructed in a single large-radius jet.

only the kinematic phase-space. The latter stage also includes modeling of soft processes such as hadronization and the underlying event[1]. The two stages are matched at the scale of a quarter of the gluino's mass, using the CKKW-L prescription [188]. In total, nearly 300 distinct samples corresponding to different $\tilde{g}$, $\tilde{\chi}_1^0$ mass pair hypothesis are generated: 152 samples with gluinos decaying to top pairs (the Gtt grid), and 144 samples where gluinos decay to bottom quarks instead (the Gbb grid). In both cases, the gluino masses range from 1.1 TeV to 2.8 TeV, and the lower bound on the neutralino mass is 1 GeV (the $\approx$ massless limit). The upper bound on the neutralino mass depends on the available kinematic phase space: $m_{\tilde{g}} - 2m_{\text{bottom}}$ for the Gbb grid, and $m_{\tilde{g}} - 2m_{\text{top}}$ for Gtt grid.

The most important SM background process with respect to the Gtt and Gbb signals is the strong production of top–anti-top quark pairs with additional high-energy jets arising from initial or final state radiation, henceforth referred to as the $t\bar{t}$ background. Matrix elements for this process are computed by `Powheg-Box v2` [189, 190] with the NNPDF 3.0 PDF set; heavy-flavor hadrons are subsequently decayed by `EVTGEN v1.6.0` [191] while soft processes are handled by `Pythia` 8.230. Additional background sources estimated from MC simulations come from production of single top-quarks, $t\bar{t}$ production associated with vector or scalar bosons, vector boson production with additional high-energy jets, and di-boson processes. The software versions for all MC samples is summarized in table 5.1

In events with no leptons, multijet production is a small but potentially non-negligible background, and the MC samples for such processes usually fail to reproduce the observed data in a reliable way in the kinematic phase-space pertinent for this analysis. Consequently, the contribution from these processes are estimated with the data-driven technique of Ref. [208], in which a template fit to the $\Delta\phi_{\text{min}}^{4j}$ (Sec. 5.2.7) distribution is performed in a statistically-independent region ($\Delta\phi_{\text{min}}^{4j} < 0.1$) and extrapolated to the phase-space targeted by the analysis regions ($\Delta\phi_{\text{min}}^{4j} \geq 0.4$).

The full 2015–2018 ATLAS $\sqrt{s} = 13$ TeV p–p dataset is used, for a total integrated luminosity of 139.0 fb^{-1} after the application of beam, detector and data-quality requirements [209]. In particular, data-taking runs in which the IBL (Section 3.2.2) is off are not included given the impact on b-jet identification performance.

[1]The underlying event describes the effects due to the proton remnants, initial and final state radiation, as well as multiple parton interactions.

Process	Generator + hadronization	Tune	PDF set	Cross-section order
Gbb/Gtt	MadGraph5_aMC@NLO v2.3.3 [184] + Pythia v8.212 [187]	A14 [171]	NNPDF2.3 [185]	$\text{NNLO}_{\text{approx}}$+NNLL [192–200]
$t\bar{t}$	Powheg-Box v2 [189, 190] + Pythia v8.230	A14	NNPDF3.0 (ME) [31] NNPDF2.3 (UE)	NNLO+NNLL [201]
Single top	Powheg-Box v2 + Pythia v8.230	A14	NNPDF3.0 (ME) NNPDF2.3 (UE)	NLO [202] (t/s-channel) NLO+NNLL [203] (Wt)
$t\bar{t}W/t\bar{t}Z$	MadGraph5_aMC@NLO v2.3.3 + Pythia v8.210	A14	NNPDF3.0 (ME) NNPDF2.3 (UE)	NLO
4-tops	MadGraph5_aMC@NLO v2.2.2 + Pythia v8.186 [170]	A14	NNPDF2.3	NLO
$t\bar{t}h$	Powheg-Box v2 + Pythia v8.230	A14	NNPDF3.0 (ME) NNPDF2.3 (UE)	NLO [204]
W/Z+**jets**	Sherpa v2.2.1 [205, 206]	Default	NNPDF3.0	NNLO [207]
WW, WZ, ZZ	Sherpa v2.2.1	Default	NNPDF3.0	NLO

Table 5.1 – Software configurations used to produce various signal and background Monte Carlo samples, detailing the generator, the set of of tuned parameters used for modeling of soft processes, the parton distribution function set, and the order of the cross-section used to normalize the samples. Separate references are given for this latter computation when they differ from that of the generator.

5.2 Physics objects and reconstruction

5.2.1 Trigger

As discussed in Section 3.2.5, the two neutralinos in each signal event, as well as potential neutrinos from the W boson decays, motivate the use of triggers based on missing transverse momentum ($E_{\text{T}}^{\text{miss}}$). This analysis employs the lowest available unprescaled such trigger specific to each data-taking period, which are fully efficient in the offline $E_{\text{T}}^{\text{miss}} \geq 200$ GeV regime.

5.2.2 Small-radius jets

The reconstruction of hadronic jets begins with three-dimensional clustering of energy deposits measured by calorimeter cells into topological clusters ("topoclusters"), a procedure that resolves the energy and the direction of single hadrons [210]. However, partons

created in the hard scattering process will undergo hadronization and so the final state jet comprises many hadrons; therefore, the topoclusters are subsequently used as input to the anti-k_t sequential recombination algorithm [211, 212] with a radius of $R = 0.4$. The energy of the resulting jets is calibrated to account for experimental effects using a combination of MC-based and *in situ* techniques [213].

5.2.3 b-jets

Because of the high b-quark multiplicities stemming from the Gbb and Gtt (via $t \rightarrow W + b$) signals, robust b-jet identification is required for this analysis. A number of low-level algorithms have been implemented for this task in ATLAS [214]:

- **IP3D**, a log-likelihood ratio (LLR) classifier based on the transverse and longitudinal impact parameter of tracks matched to a given jet, taking advantage of the relatively long lifetimes of the lightest b-mesons and b-baryons, which will be present inside a b-quark-initiated jet;
- **SV1**, another LLR classifier also taking advantage of the b-hadron lifetime but by trying to recognized jets originating from a displaced interaction vertex;
- **JetFitter**, a Kalman Filter-based algorithm trying to reconstruct the decay chain of b- and c-hadrons.

These algorithms are then combined with additional kinematic information using a Boosted Decision Tree (BDT) classifier, resulting in the MV2 algorithm[2]. A working point corresponding to a nominal efficiency of 77% is employed; its identification and rejection performance can be seen in Figure 5.4.

5.2.4 Large-radius jets

Given the large hypothesized mass of the gluino and the potentially large difference in mass between the gluino and the neutralino, boosted W bosons and top quarks from the gluino decay can subsequently create decay chains so collimated that the individual constituents are not well resolved by the anti-k_T $R = 0.4$ algorithm, as exemplified in Figure 5.3.

One way to deal with such objects is to go back to the topoclusters and reconstruct another set of jets starting using a larger reconstruction radius. This analysis makes use

[2]Another alternate combination using neural network, the DL1 algorithm, is available but not used in this analysis.

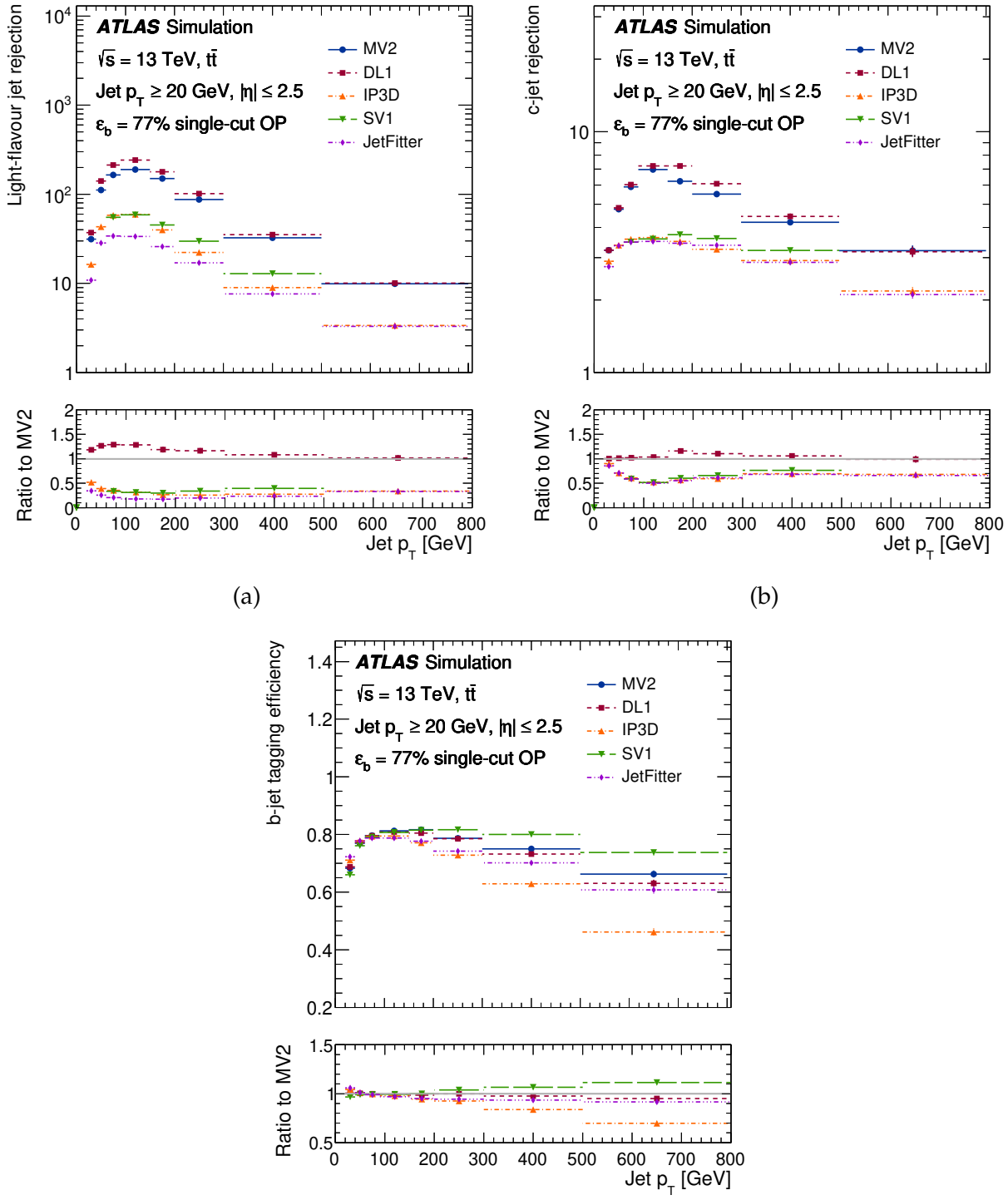


Figure 5.4 – (a) light-flavor jet rejection, (b) c-jet rejection, and (c) b-jet identification efficiency as a function of jet p_{T} for the different ATLAS b-tagging algorithms at their nominally 77% efficient working point [214].

of an alternate strategy, instead using the $R = 0.4$ jets themselves as input to another iteration of the anti-k_t algorithm but with a radius of $R = 0.8$. This technique, known as *jet re-clustering*, allows for easy propagation of calibrations and uncertainties from the input jet collection [215, 216]. Such large-radius jets can properly reconstruct the W boson and top quark masses.

5.2.5 Leptons

Electrons are reconstructed with two different algorithms with differing efficiency and rejection factors. So-called baseline electron candidates are reconstructed using a high-efficiency, low-rejection set of criteria using information from the tracker and the electromagnetic calorimeters in the $p_T > 20$ GeV, $|\eta| < 2.47$ range; they and are used to veto leptonic events in order to define an all-hadronic channel. Signal electrons are used to define the leptonic channels, and are defined by a set of additional criteria on top of the baseline identification, resulting in a lower-efficiency but higher-rejection operating point [217]. The same paradigm is used for identification for muons [218]. No attempt is made to reconstruct τ-leptons; because of their large mass they tend to decay to hadrons, in which case they get reconstructed as jets [11].

Leptons useful for this analysis are qualified as being prompt, that is, they arise from the hard scattering interaction. However, there can be other sources of lepton production, which lead to some ambiguity and potential double-counting of energy contributions; this problem is avoided by the implementation of an *overlap removal* procedure similar to that of Ref. [9].

First, electrons arising from muon bremsstrahlung are avoided by discarding electrons that share an inner detector track with a muon. Contributions from jets arising from hadrons coming from prompt electron decay are avoided by removing any jet whose axis lies within $\Delta R = \sqrt{\Delta\Upsilon^2 + \Delta\phi^2} < 0.2$ of an electron. Conversely, contributions from electrons produced in hadronic cascades are removed by discarding electrons with $E_T < 50$ GeV that are within $\Delta R < 0.4$ of a jet. A slightly more complex strategy is employed for higher energy electrons, discarding them if they are within $\Delta R(p_T) < \min(0.4, 0.04 + 10/p_T)$ of a jet's axis, to increase acceptance in boosted scenarios. Jets arising from muon bremsstrahlung are avoided by removing jets with fewer than three matching ID tracks that contain a muon within $\Delta R < 0.2$ of its axis, and the remaining ambiguity from muon created in hadronic cascades is handled in the same way as for electrons.

5.2.6 Missing transverse energy

The missing transverse momentum is a two-vector in the R–ϕ plane, defined by the negative of the vector sum of the transverse momenta of all physics objects described above. To account for contributions from particles not matched to any selected objects, a so-called "soft term" is computed with unused inner detector tracks that are matched to the primary vertex. The magnitude of the resulting missing transverse momentum vector is referred to as missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$ [219, 220].

As mentioned earlier, this analysis is mainly concerned with the $E_{\mathrm{T}}^{\mathrm{miss}} \geq 200$ GeV regime, for which the selected online $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers are fully efficient.

5.2.7 Kinematic variables

Beyond the direct usage of the four-momenta of the physics objects described above, a few event-level kinematic input variable are built from the four-vectors and are useful in various aspects of the analysis:

- N_{jet}, the number of small-R jets;
- N_b, the number of b-tagged small-R jets at the 77% MV2 operating point;
- N_l^{signal} and N_l^{baseline}, the number of signal and baseline candidate leptons;
- m_{eff}, the effective mass, which correlates to the mass scale of sparticles produced in the hard scattering interaction:

$$m_{\mathrm{eff}} = \sum_{j=1}^{N_{\mathrm{jet}}} p_{\mathrm{T}}^{j} + \sum_{l=1}^{N_l} p_{\mathrm{T}}^{l} + E_{\mathrm{T}}^{\mathrm{miss}}; \tag{5.1}$$

- M_J^{Σ}, the sum of the four leading large-radius jet masses, sensitive to the presence of boosted W bosons or top quarks;
- $\Delta\phi_{\mathrm{min}}^{4j}$, the minimum distance in ϕ between $E_{\mathrm{T}}^{\mathrm{miss}}$ and the leading four small-R jets, useful to recognize events with large $E_{\mathrm{T}}^{\mathrm{miss}}$ due to mismeasured jets:

$$\Delta\phi_{min}^{4j} = \min(|\phi_1 - \phi_{E_{\mathrm{T}}^{\mathrm{miss}}}|, ..., |\phi_4 - \phi_{E_{\mathrm{T}}^{\mathrm{miss}}}|). \tag{5.2}$$

5.3 Data–Monte Carlo agreement

In order to be sure that any potential excess observed above the Standard Model is not caused by imperfect simulations, it is crucial to verify the data and Monte Carlo agreement after application of a set of loose criteria. In these so-called pre-selection regions, the background cross-section is still expected to be much higher than that of the Gtt and Gbb signals, and thus any large mismodelling not due to these models can be easily spotted and corrected, if need be.

A common pre-selection is first performed:

- $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger
- Offline $E_{\mathrm{T}}^{\mathrm{miss}} \geq 200$ GeV
- $N_{\mathrm{jet}} \geq 4$
- $N_b \geq 3$

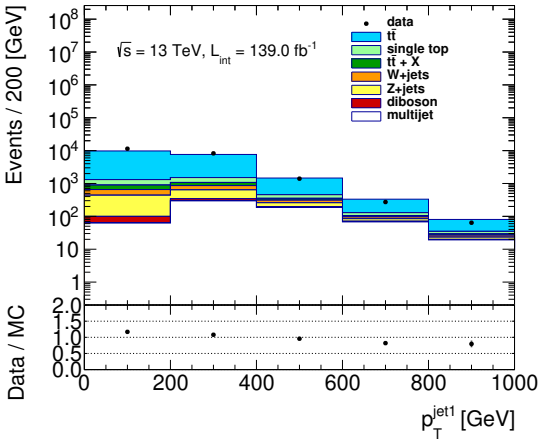
Then, to ease interpretation, two different channels are defined:

- 0-lepton channel:
  - $N_l^{\mathrm{baseline}} = 0$
  - $\Delta\phi_{\mathrm{min}}^{4j} \geq 0.4$
- 1-lepton channel:
  - $N_l^{\mathrm{signal}} \geq 1$

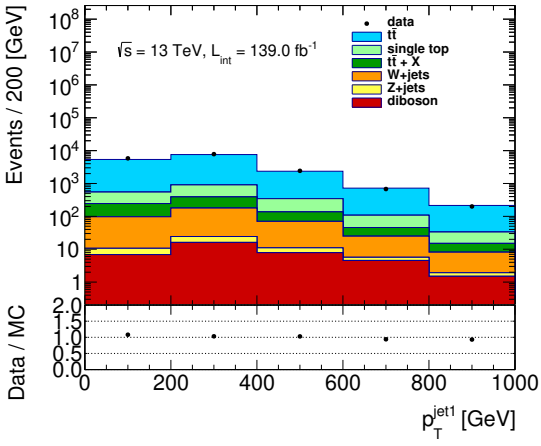
As in previous iterations of this analysis [8, 9], a data–Monte Carlo discrepancy is observed in the 1-lepton channel, for instance in the high-m_{eff} regime.[3] Consequently, a set of normalization factors are derived in dedicated control regions for the $t\bar{t}$ and W+jets backgrounds in the m_{eff}–N_{jet} plane, and for the single-top and Z+jets backgrounds in the m_{eff} distribution. All ≥ 1-lepton plots are shown with the resulting weights applied; the distribution of these weights can be found in Appendix A.

Figures 5.5 to 5.10 show the level of agreement for a few of the variables used in this analysis, taking the statistical uncertainty into account; no large discrepancy is present at pre-selection level.

[3] This is observed by several ATLAS analyses, and the cause is still not fully understood. One hypothesis is that it is caused by quantum interference between different top-related processes which is not yet accounted for in the Monte Carlo simulations.

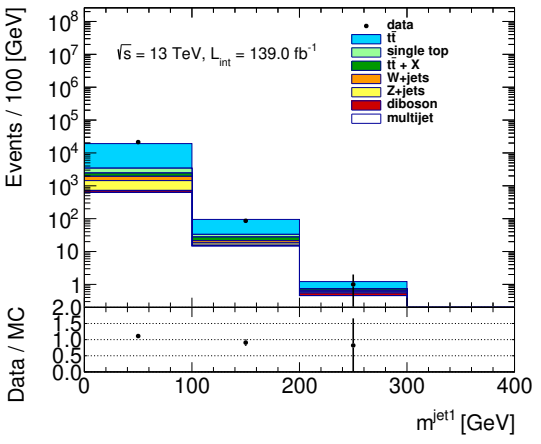


(a) 0ℓ

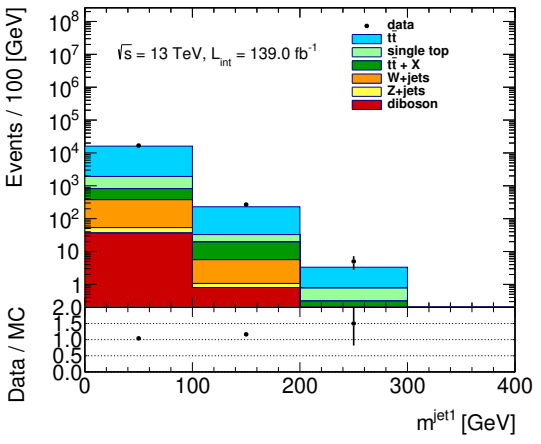


(b) $\geq 1\ell$

Figure 5.5 – Data/MC comparison for the leading small-radius jet p_{T} in the (a) 0-lepton and (b) $\geq$ 1-lepton channels.

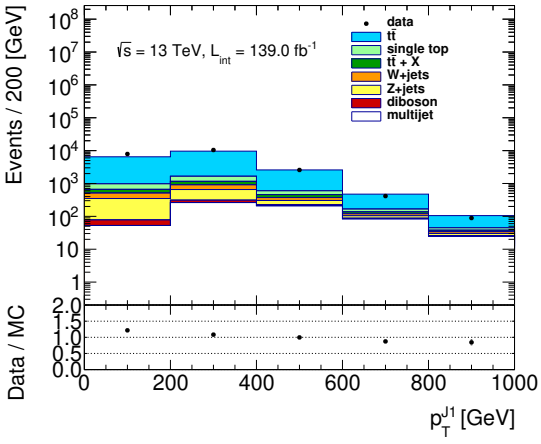


(a) 0ℓ

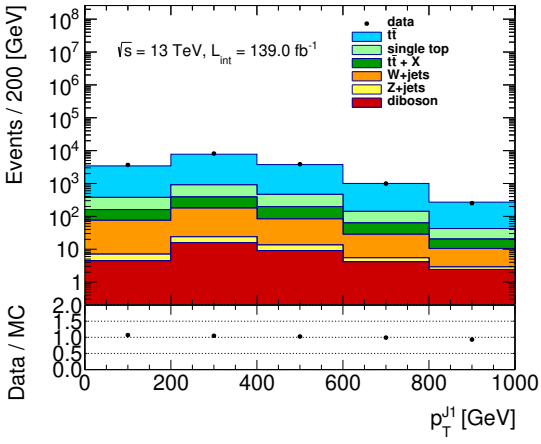


(b) $\geq 1\ell$

Figure 5.6 – Data/MC comparison for the leading small-radius jet mass in the (a) 0-lepton and (b) $\geq$ 1-lepton channels.

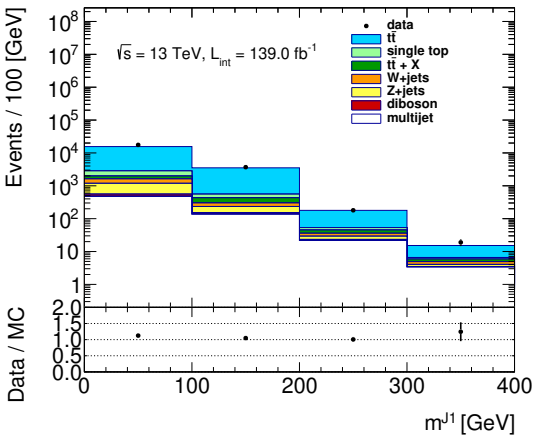


(a) 0ℓ

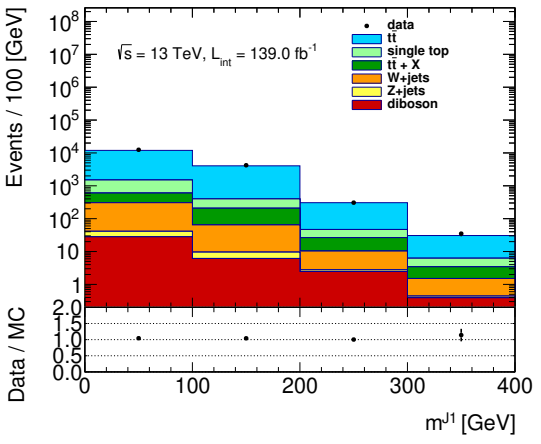


(b) $\geq 1\ell$

Figure 5.7 – Data/MC comparison for the leading large-radius jet p_{T} in the (a) 0-lepton and (b) $\geq$ 1-lepton channels.

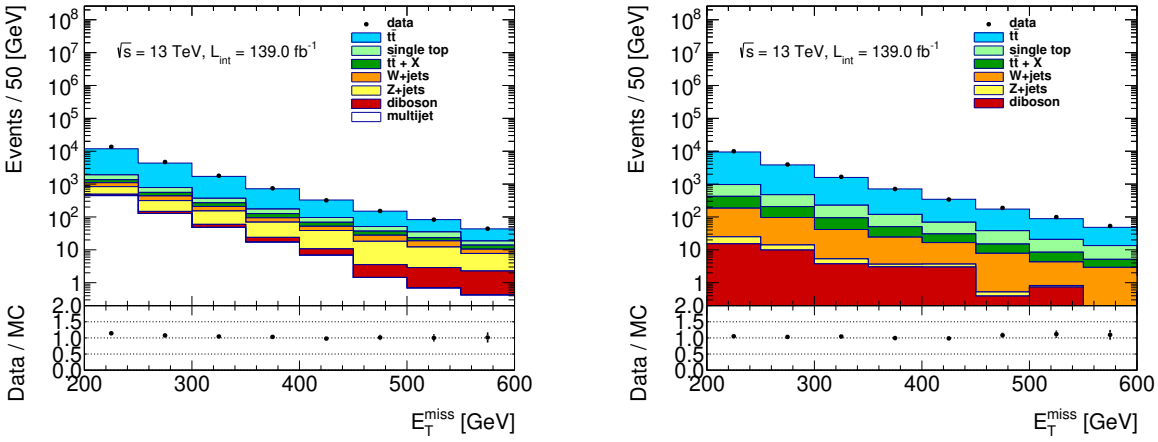


(a) 0ℓ



(b) $\geq 1\ell$

Figure 5.8 – Data/MC comparison for the leading large-radius jet mass in the (a) 0-lepton and (b) $\geq$ 1-lepton channels.



(a) 0ℓ (b) $\geq 1\ell$

Figure 5.9 – Data/MC comparison for $E_{\mathrm{T}}^{\mathrm{miss}}$ in the (a) 0-lepton and (b) $\geq$ 1-lepton channels.

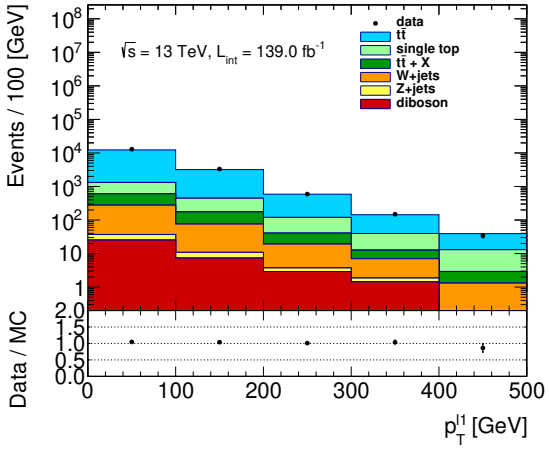


Figure 5.10 – Data/MC comparison for the leading lepton p_{T} in the $\geq$ 1-lepton channel.

5.4 Neural network for event selection

Traditionally, high-energy physics searches use physically-motivated hand-crafted variables such as invariant or transverse masses, which are easily interpreted by humans. However, which high-level variables are better suited to the task at hand is not always obvious, and important but more subtle information may be discarded in the transition away from the lower-level objects. Using neural networks (Section 4.2) for event selection in high energy physics is by no means a new idea [221–224], and analyses using such methods also generally use a restricted set of hand-crafted variables; however, the recent deep learning revolution in machine learning [7] has motivated the use of neural networks performing event selection starting from low-level inputs such as the four-momenta of selected objects [225, 226], potentially allowing more relevant information to be retained.

For the present analysis, a neural network is trained to recognize events from the Gtt and Gbb signals against the Standard Model background (Table 5.1) using the following low-level input set, designed to cover all possible decay topologies of a Gtt event (see Figure 5.2b):

- The four-momenta (p_{T}, η, ϕ, m) of the 10 leading small-R jets, and a set of binary variables indicating which jets are b-tagged (MV2, 77% WP);
- The four-momenta of the four leading large-R jets;
- The four-momenta of the four leading leptons (e or μ);
- The $E_{\mathrm{T}}^{\mathrm{miss}}$ vector.

The simplest way use such a network with parameterized signal grids such as Gbb and Gtt would be to train a specific classifier for each mass point. However, in order to maximize the training set statistics for a single training, a parameterized learning method [227] is used instead, allowing for a more efficient use of computational resources and a potentially more robust classifier, since adding training data is one of the most powerful ways to reduce the generalization error [146]. To further improve the statistics, events with $E_{\mathrm{T}}^{\mathrm{miss}} \in (100, 200)$ GeV as well as events with $N_b = 2$ are also used, albeit for training only.

The goal of the parameterized learning method is to obtain a single classifier that is optimal for every $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0})$ mass hypothesis across both signal grids. To do so, the neural network uses a further three non-discriminating inputs[4], or parameters:

- A binary variable, I_{Gtt}, identifying the event as coming from the Gtt or Gbb grids;

[4]Non-discriminating in the sense that they will be fixed at test time, thus completely removing their discriminating power from the equation.

- The gluino mass, $M_{\tilde{g}}$[5];
- The neutralino mass, $M_{\tilde{\chi}_1^0}$.

The first parameter allows the network to learn different strategies to recognize for Gtt and Gbb events from the SM background, since here the two grids are taken to be mutually exclusive[6]; the last two parameters allow the strategy to be modulated according to the masses of the particles searched for. These parameters are only meaningful for signal events; background events get their parameter values by sampling randomly from the signal distribution, which also removes the correlation of the first parameter with the network's output.

The neural network does not simply output a single signal or background probability score, but also tries to discriminate between different background classes. In total, there are eight softmax output scores:

- Two signal probability scores, one for each signal grid: $P(\mathrm{Gtt})$ and $P(\mathrm{Gbb})$;[7]
- Six background probability scores, one for each Monte Carlo-estimated background: $P(t\bar{t})$, $P(\text{single-top})$, $P(tt + X)$, $P(W + \text{jets})$, $P(Z + \text{jets})$ and $P(\text{diboson})$.

To ensure that no bias is introduced by the training and hyperparameter selection procedures, the whole dataset is split into three statistically independent subsets:

- The training set (30%), used to train all hyperparameter samplings;
- The model selection set (35%), used to pick the best hyperparameters;
- The test set (35%), used to produce the final performance estimate.

The training set comprises about 1.09M and 2.25M signal and background events, respectively. The signal cross-section being much smaller than that of the background processes precludes using the physical cross-sections for training; using unweighted events from signal and background samples with the statistics listed above upweights the signal contribution significantly for training. In the model selection and test sets, the events are always weighted to the physical cross-sections. To select the optimal hyperparameters, a simple random search [228] was performed on the Compute-Canada GPU cluster `Béluga`[8];

[5] An uppercase M is used here to differentiate from the neural network input parameter from m, the physical mass for a given grid point.

[6] A straightforward extension of this method would consider signals with mixed gluino decay, with this binary parameter being replaced by branching ratios to different final states.

[7] Since in this work the Gbb and Gtt grids are taken to be mutually exclusive, this could be reduced to a single signal probability score.

[8] `https://docs.computecanada.ca/wiki/B%C3%A9luga/en`

Hyperparameter	Searched range	Selected Value
Hidden layers	1–4	3
Hidden units	50–1000	737
Learning rate	10^{-5}–10^{-1}	1.30×10^{-5}
Batch	2^5–2^{10}	128
Dropout (input)	Yes or No	No
Dropout (hidden)	Yes or No	No
BatchNorm	Yes or No	Yes
L1 (hidden)	0 or 10^{-7}–10^{-2}	1.57×10^{-5}
L1 (output)	0 or 10^{-7}–10^{-2}	0
L2 (hidden)	0 or 10^{-7}–10^{-2}	2.31×10^{-7}
L2 (output)	0 or 10^{-7}–10^{-2}	0
Early-stop	Cross-entropy loss or f-score	Cross-entropy loss

Table 5.2 – Configuration and result of the hyperparameter optimization.

the searched range as well as the optimal configuration are listed in Table 5.2. In all cases, the ADAM algorithm (Section 4.2.1) is used to tune the weights.

After training, first the I_{Gtt}, $M_{\tilde{g}}$ and $M_{\tilde{\chi}_1^0}$ parameters are set to definite values corresponding to a specific mass hypothesis in a specific grid, the same for all background, signal, and data samples, which removes the discriminating power of these variables; this allows using the neural network to construct a set of output variables amenable to simple counting experiments in signal regions (SR), which are scrupulously kept *blinded* until the very end of the analysis, in order to avoid introducing biases in the SR definitions. To fully explore the discovery and exclusion potential of such regions, the following procedure is followed:

1. Select a signal (Gbb or Gtt), which determines the value of I_{Gtt}, and a mass point $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0})$;
2. Evaluate $P(\mathrm{Signal}|I_{\mathrm{Gtt}}, M_{\tilde{g}} = m_{\tilde{g}}, M_{\tilde{\chi}_1^0} = m_{\tilde{\chi}_1^0})$ for the signal sample and all background samples;
3. Tune a decision threshold on the resulting variable, subject to the following requirements:
    - Maximize the expected statistical significance, which is the magnitude of a statistical fluctuation (measured in standard deviations) above the background-only hypothesis that would explain the yield that is expected under the signal+background hypothesis;[9];

[9]During optimization, the `NumberCountingUtils::BinomialExpZ` formula from the `RooStats` library is

- Obtain ≥ 0.5 expected background events;
- Obtain $\leq 30\%$ Poisson statistical uncertainty on the $t\bar{t}$ background yield;

4. Repeat for all mass pairs in both signal grids.

This procedure yields about 300 signal regions, one per grid point. To select the best hyperparameters, this procedure is carried-out using the model selection dataset for all different trainings and the best model is defined as that which has the largest number of grid points with statistical significance ≥ 1.64, corresponding to a one-sided confidence level (CL) of 95%. For the final performance estimate however, such a high number of signal regions is unreasonable as the discovery potential is weakened by the Look-Elsewhere Effect (LEE) [229], according to which the probability of observing a signal-like statistical fluctuation of the background increases with number of counting-experiments performed. Therefore, after obtaining all these regions on the test set with the final model, we need to find a minimal set that can maximize the exclusion reach and the sensitivity for both grids. This is equivalent to the set cover problem[10], for which a simple solution exists: iteratively select regions from the full set until exclusion and sensitivity stops improving. Following this strategy, optimality is attainable with only four signal regions per grid, for a grand total of eight; their definitions can be found in Tables 5.3 and 5.4.

One signal region is assigned to each grid points of both models; the resulting coverage is shown in Figure 5.11. Although the mass splitting is not taken into account while performing the assignment, the various signal regions end up being assigned to specific $m_{\tilde{g}} - m_{\tilde{\chi}_1^0}$ ranges.

As can be seen in figure 5.12, the neural network is much more powerful than the simple cut-and-count analysis (CCA) performed in Ref. [9], even after accounting for the luminosity difference. In fact, the neural network regions reach a signal efficiency[11] sometimes two or three times higher than that of the CCA regions, while having only one-half to one-tenth of the corresponding background efficiency; this translates to improvements in statistical significance by up to a factor 85 for some mass hypotheses. In such cases, the cut-and-count analysis would need to accumulate roughly 7000× more data to attain an equivalent performance level in single regions[12][13]. This increased performance also

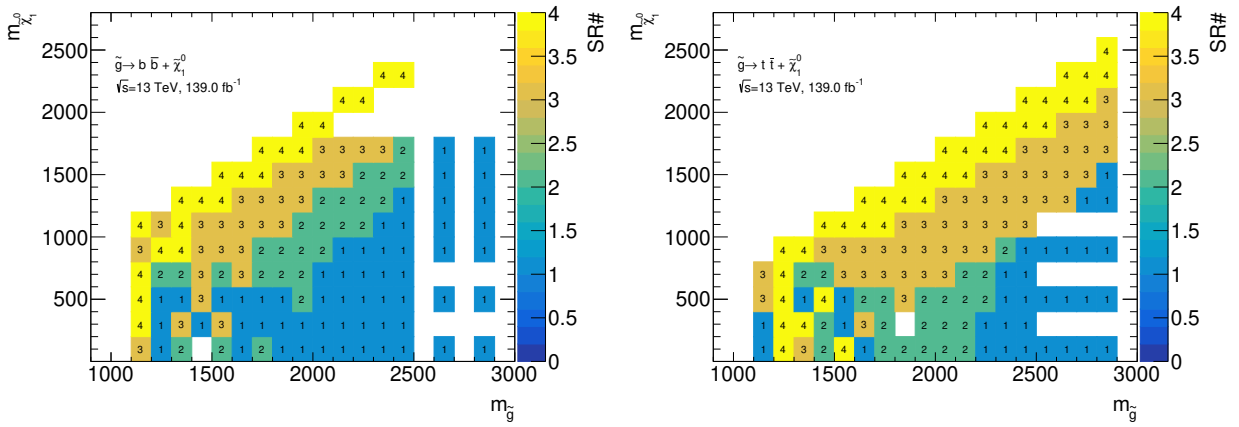
used to compute it, while the profile likelihood fit of Section 5.5 is used for the final results.

[10] `https://en.wikipedia.org/wiki/Set_cover_problem`

[11] The signal (background) efficiency of a signal region is defined as the fraction of events from a signal (background) sample that are accepted into the region.

[12] The statistical significance of a single region can be approximated by $Z = S/\sqrt{B}$. Therefore, scaling the yields by a factor L only improves the significance by $\sqrt{L}$.

[13] To be completely fair, the previous CCA strategy was optimized for 79.9 fb^{-1}, and the 0 ℓ and $\geq 1\ell$ channels were statistically combined. The single-region results have been recomputed with the present



(a) Gbb (b) Gtt

Figure 5.11 – Signal region coverage for the (a) Gbb and (b) Gtt signal grids. In both cases, the regions are numbered in the order in which they are listed in Tables 5.3 and 5.4. The empty squares represent mass pairs for which no signal samples were produced.

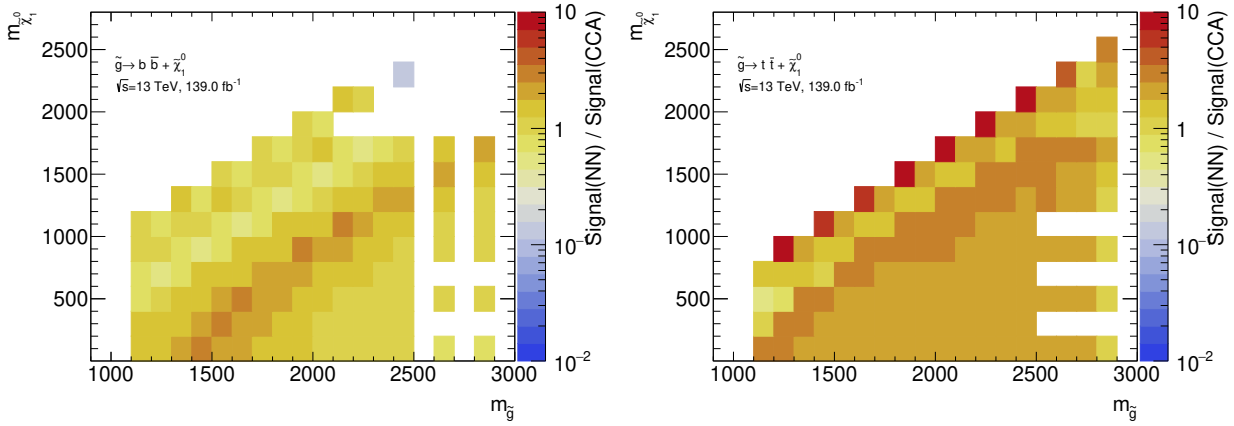
simplifies the analysis by removing the need to perform a statistical combination of many orthogonal regions.

In order to verify and correct potential mis-modeling, each SR has an associated orthogonal control region (CR), in which background-only fits (Section 5.5) are performed to estimate data/MC normalization factors used to correct the normalization of the leading background, $t\bar{t}$. Sets of validation regions (VR), orthogonal to both the SR and CR, are also implemented to verify the remaining mismodelling after application of the normalization factor derived in the CR [230].

While the signal regions are defined only in terms of a decision threshold on the neural network output, the CRs and VRs have additional requirements. For instance, thresholds or ranges in m_{eff} and M_J^{Σ} are applied in some regions to keep them kinetically close to the SRs as the neural network output requirement is loosened, and to reduce the signal contamination to acceptable levels ($\leq$ 20(10)% in VR(CR)). To ensure that the derived $t\bar{t}$ normalization factors are meaningful, a lower threshold on $P(t\bar{t})$ is added to the CRs to maximise the $t\bar{t}$ purity. In the results that follows, two kinds of validation regions are implemented for the Gtt regions:

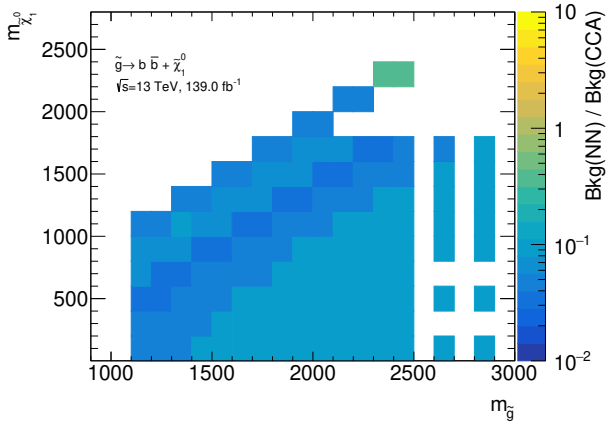
- **VR1**, which are primarily defined by intermediate cuts on the signal probability (*i.e.*, between the CRs and SRs) and allow to verify the level of agreement after the application of the $t\bar{t}$ normalization factors. These regions do not have any cuts on $P(t\bar{t})$ in order to keep the region definition as similar as possible to that of the SRs;

luminosity to calculate the numbers shown here.

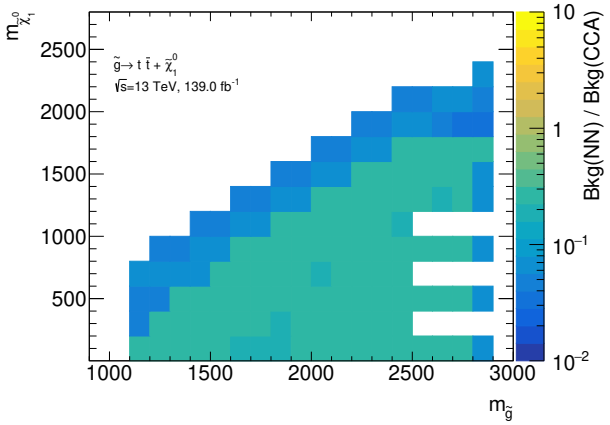


(a) Gbb, signal efficiency

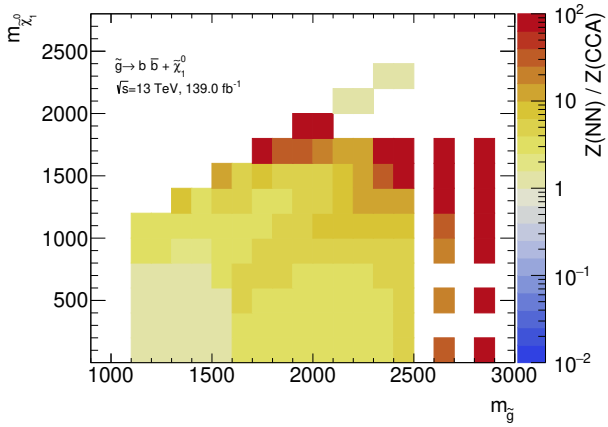
(b) Gtt, signal efficiency



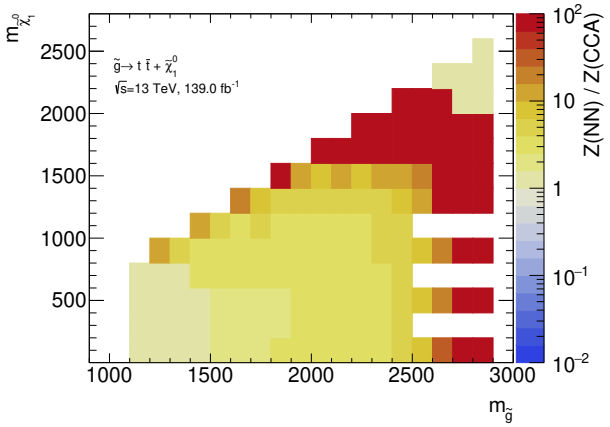
(c) Gbb, background efficiency



(d) Gtt, background efficiency



(e) Gbb, statistical significance



(f) Gtt, statistical significance

Figure 5.12 – (a,b) Signal and (c,d) background efficiencies as well as (e,f) statistical significance of the neural network regions divided by that of the best cut-and-count (CCA) discovery regions of Ref [9] for each mass pair, for the Gbb and Gtt models.

- **VR2**, which are defined by inverting the $P(t\bar{t})$ in the CRs.

The Gbb SRs are usually dominated by the irreducible Z+jets background, primarily because of $Z \to \nu\nu$ events; this background can be more efficiently suppressed in the Gtt SRs because of the higher jet multiplicities due to four top quark decays, for instance. Designing control regions pure enough to derive normalization factors with acceptable uncertainties for the Z+jets background is unfeasible; fortunately, it is possible to define additional validation regions for which the purity requirement is less crucial. This is done by requiring lower thresholds on $P(Z + \text{jets})$ to enhance the fraction of this background, allowing to check for potential mismodeling. Each Gbb SR/CR pair therefore have three associated validation regions:

- **VR1**, again defined by intermediate cuts on the signal probability but with an upper limit on the Z+jets probability;
- **VR2**, again defined by inverting the $P(t\bar{t})$ in the CRs;
- **VR3**, defined like VR1 but with a lower limit on the Z+jets probability instead, to check the modeling of this background.

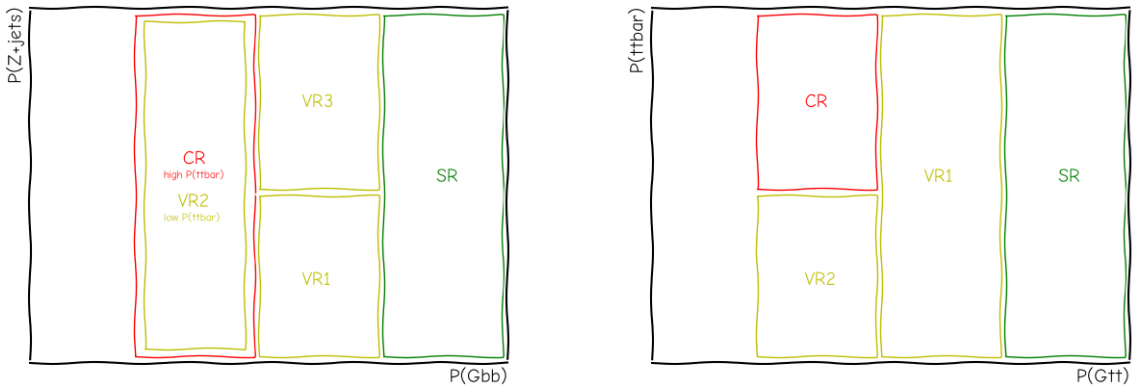
The basic layout of the three types of regions can be seen in Figure 5.13. Region definitions are listed in Tables 5.3 and 5.4 for Gbb and Gtt, respectively; expected background yields and composition can be seen in Figures 5.14 and 5.15. The Gtt control regions achieve 70% to 80% $t\bar{t}$ purity, while the Gbb regions achieve about 50% to 80% purity; the Z+jets purity in the Gbb VR3 set range from 20% to 50%. One can also see in these figures the very stringent nature of the SRs, which typically have $\lessapprox 1$ expected background event.

While the CR and the VRs assigned to a given SR are orthogonal to each other (and to the SR), different SR/VRs/CR triplets are not necessarily orthogonal to other triplets, and the overlap of background events between different regions is shown in Figure 5.16.

The Data–Monte Carlo agreement after the pre-selections described in Section 5.3 for relevant neural network output variables are shown in Figures 5.17–5.23. The signal probabilities (Figures 5.17 and 5.21) are generally well modeled. The $P(t\bar{t})$ distributions (Figures 5.18 and 5.23) show some disagreement at high values, which is the range that is corrected by the $t\bar{t}$–enriched control regions via fitted normalization factors. Such normalization differences are considered normal even though the control regions have negligible expected signal yields, because of the large extrapolation from the bulk of the $t\bar{t}$ sample ($\gtrapprox 100$M events) to a small number of events in the CRs ($\lessapprox 100$ events). The Z+jets probabilities (Figure 5.19) show the same tendency, but there is no dedicated control

Region	P(Gbb)	log10 P($t\bar{t}$)	log10 P(Z+jets)	$\Delta\phi^{4j}_{min}$	m_{eff}	M_J^{Σ}
SR-Gbb-2800-1400	≥ 0.9994	-	-	$\geq$0.6	-	-
VR1-Gbb-2800-1400	$\in (0.79, 0.9994($	-	<-1.7	$\geq$0.4	≥ 2400	< 800
VR2-Gbb-2800-1400	$\in (0.57, 0.79($	<-0.9	-	≥ 0.4	≥ 1400	< 800
VR3-Gbb-2800-1400	$\in (0.79, 0.9994($	-	$\geq$-1.7	$\geq$0.4	$\in (1400, 2000($	-
CR-Gbb-2800-1400	$\in (0.57, 0.79($	$\geq$-0.9	-	$\geq$0.4	≥ 1400	< 800
SR-Gbb-2300-1000	≥ 0.9996	-	-	$\geq$0.6	-	-
VR1-Gbb-2300-1000	$\in (0.89, 0.9996($	-	<-1.7	$\geq$0.4	≥ 2200	< 800
VR2-Gbb-2300-1000	$\in (0.72, 0.89($	<-1.1	-	≥ 0.4	≥ 1200	< 800
VR3-Gbb-2300-1000	$\in (0.89, 0.9996($	-	$\geq$-1.7	$\geq$0.4	$\in (1200, 2000($	-
CR-Gbb-2300-1000	$\in (0.72, 0.89($	$\geq$-1.1	-	$\geq$0.4	≥ 1200	< 800
SR-Gbb-2100-1600	≥ 0.9997	-	-	$\geq$0.4	-	-
VR1-Gbb-2100-1600	$\in (0.87, 0.9997($	-	<-1.3	$\geq$0.4	≥ 800	< 700
VR2-Gbb-2100-1600	$\in (0.85, 0.87($	<-1.4	-	≥ 0.4	≥ 800	< 700
VR3-Gbb-2100-1600	$\in (0.87, 0.9997($	-	$\geq$-1.3	$\geq$0.4	-	-
CR-Gbb-2100-1600	$\in (0.85, 0.87($	$\geq$-1.4	-	$\geq$0.4	≥ 800	< 700
SR-Gbb-2000-1800	≥ 0.9996	-	-	$\geq$0.4	-	-
VR1-Gbb-2000-1800	$\in (0.89, 0.9996($	-	<-1.3	$\geq$0.4	≥ 400	< 400
VR2-Gbb-2000-1800	$\in (0.88, 0.89($	<-1.6	-	≥ 0.4	≥ 400	< 400
VR3-Gbb-2000-1800	$\in (0.89, 0.9996($	-	$\geq$-1.3	$\geq$0.4	-	-
CR-Gbb-2000-1800	$\in (0.88, 0.89($	$\geq$-1.6	-	$\geq$0.4	≥ 400	< 400

Table 5.3 – Gbb region definitions.



(a) (b)

Figure 5.13 – Signal, validation and control region strategy for (a) Gbb and (b) Gtt regions.

Region	P(Gtt)	log10 P($t\bar{t}$)	m_{eff}	M_J^{Σ}
SR-Gtt-2100-1	$\geq$ 0.9998	-	-	-
VR1-Gtt-2100-1	$\in$ (0.85, 0.9998(	-	$\geq$ 1800	-
VR2-Gtt-2100-1	$\in$ (0.7, 0.85(	<-0.9	$\geq$ 1200	-
CR-Gtt-2100-1	$\in$ (0.7, 0.85(	$\geq$-0.9	$\geq$ 1200	-
SR-Gtt-1800-1	$\geq$ 0.9997	-	-	-
VR1-Gtt-1800-1	$\in$ (0.85, 0.9997(	-	$\geq$ 1600	-
VR2-Gtt-1800-1	$\in$ (0.76, 0.85(	<-1.0	$\geq$ 1200	-
CR-Gtt-1800-1	$\in$ (0.76, 0.85(	$\geq$-1.0	$\geq$ 1200	-
SR-Gtt-2300-1200	$\geq$ 0.9997	-	-	-
VR1-Gtt-2300-1200	$\in$ (0.81, 0.9997(	-	$\geq$ 1500	-
VR2-Gtt-2300-1200	$\in$ (0.74, 0.81(	<-1.0	$\geq$ 1200	-
CR-Gtt-2300-1200	$\in$ (0.74, 0.81(	$\geq$-1.0	$\geq$ 1200	-
SR-Gtt-1900-1400	$\geq$ 0.9996	-	-	-
VR1-Gtt-1900-1400	$\in$ (0.89, 0.9996(	-	$\geq$ 600	$<$ 500
VR2-Gtt-1900-1400	$\in$ (0.87, 0.89(	<-1.6	$\geq$ 600	$<$ 500
CR-Gtt-1900-1400	$\in$ (0.87, 0.89(	$\geq$-1.6	$\geq$ 600	$<$ 500

Table 5.4 – Gtt region definitions.

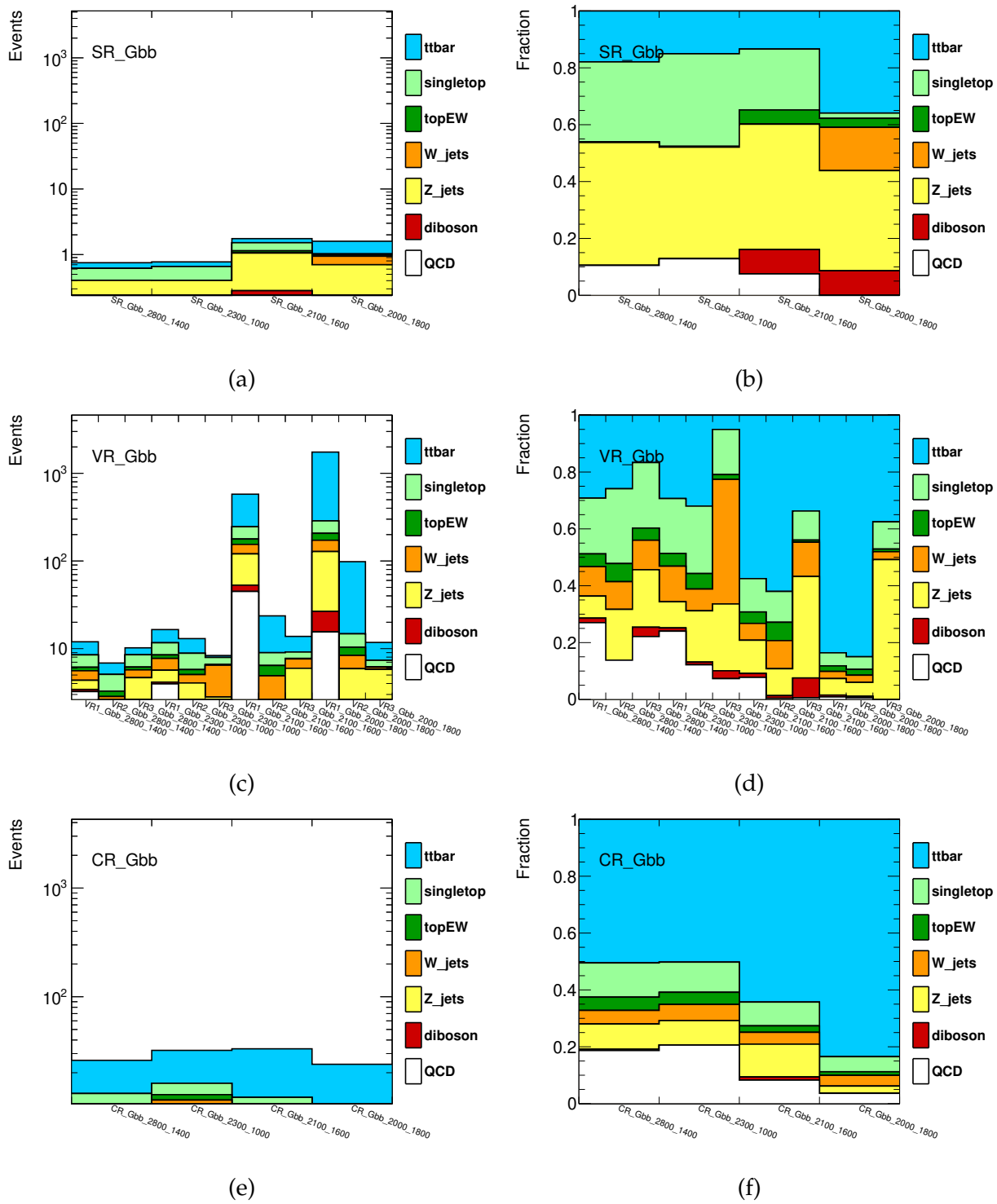


Figure 5.14 – Pre-fit expected background (a,c,e) yields and (b,d,f) compositions in the Gbb regions.

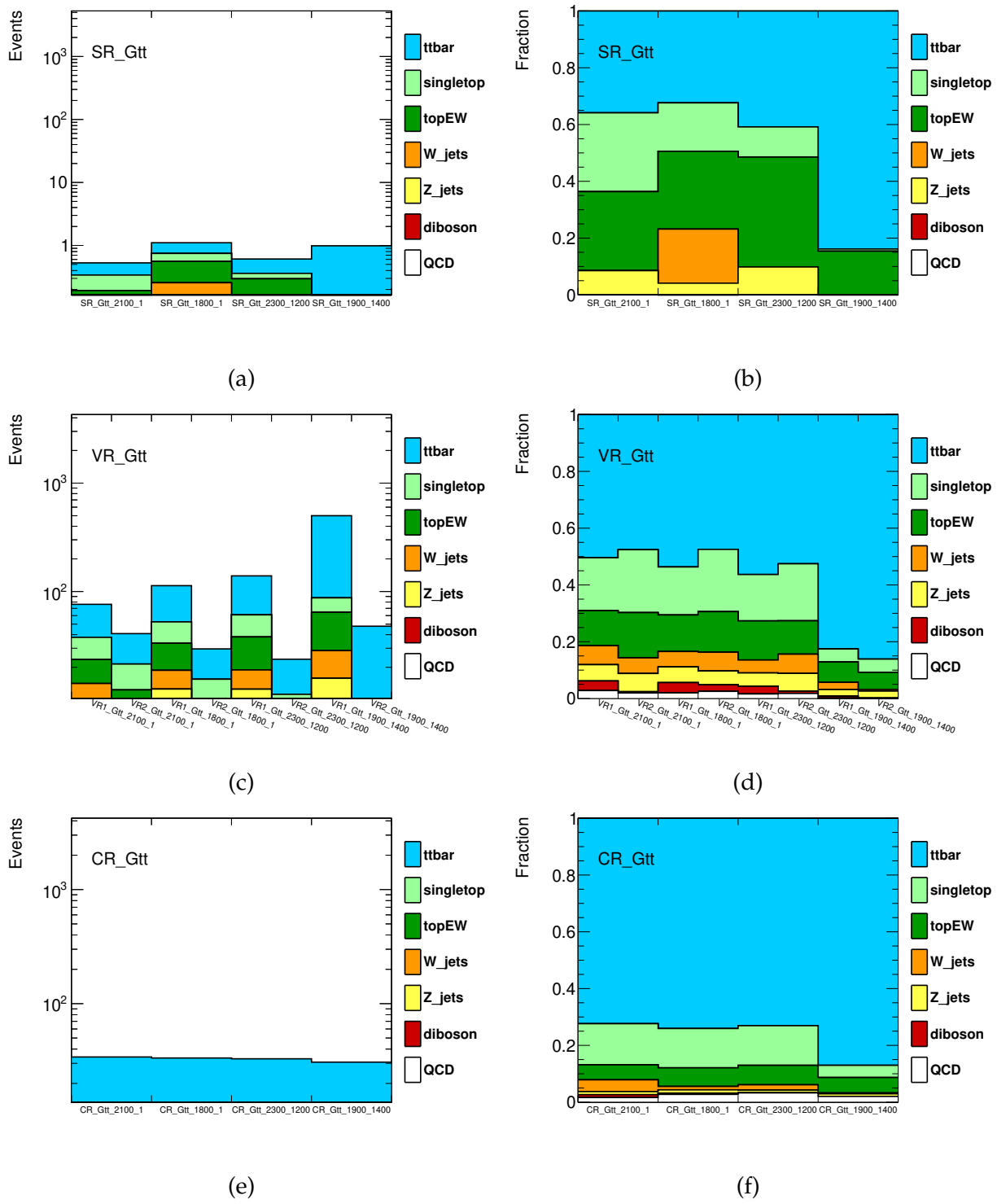
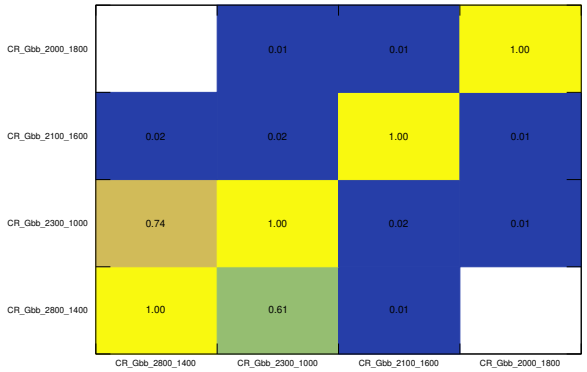
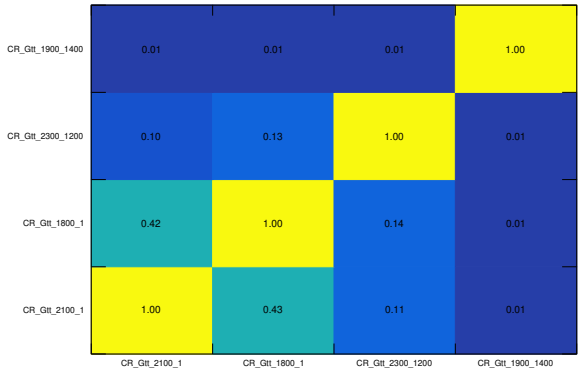


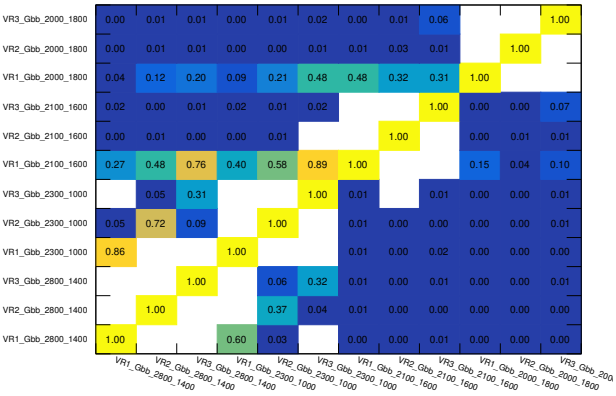
Figure 5.15 – Pre-fit expected background (a,c,e) yields and (b,d,f) compositions in the Gtt regions.



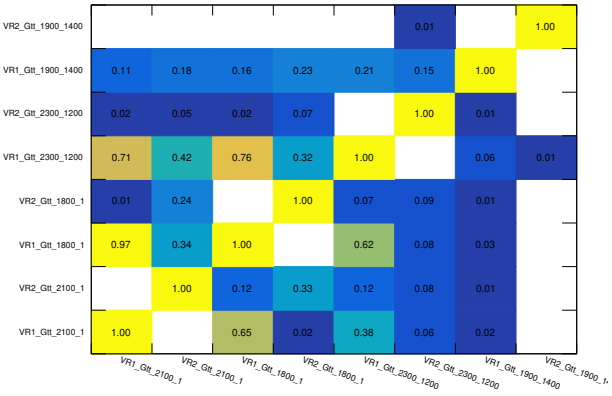
(a) Gbb, CR



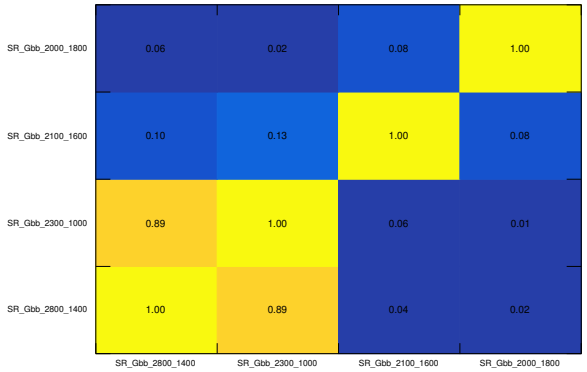
(b) Gtt, CR



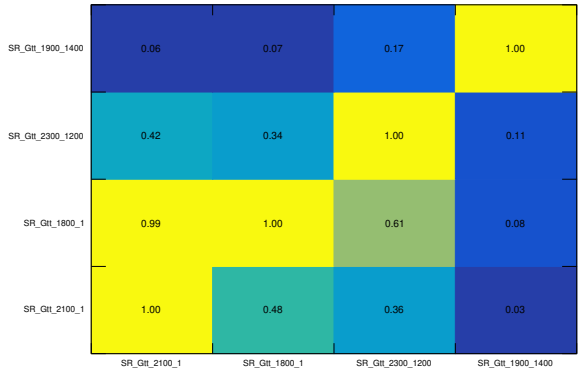
(c) Gbb, VR



(d) Gtt, VR

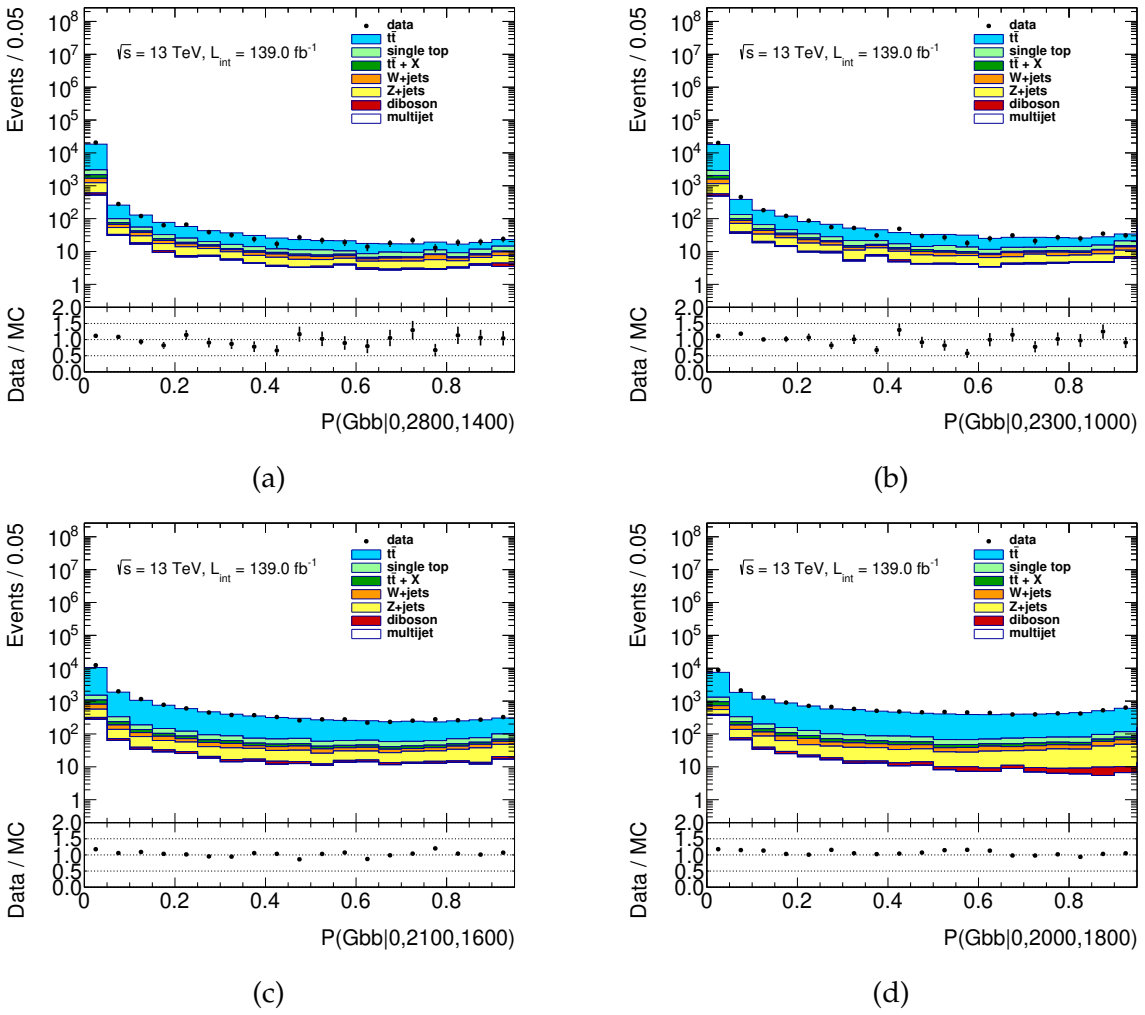


(e) Gbb, SR



(f) Gtt, SR

Figure 5.16 – Overlap of background events in the (a,c,e) Gbb and (b,d,f) Gtt regions. Each cell quantifies the fraction of events in the signal region of the x axis that are also in the signal region of the y axis.



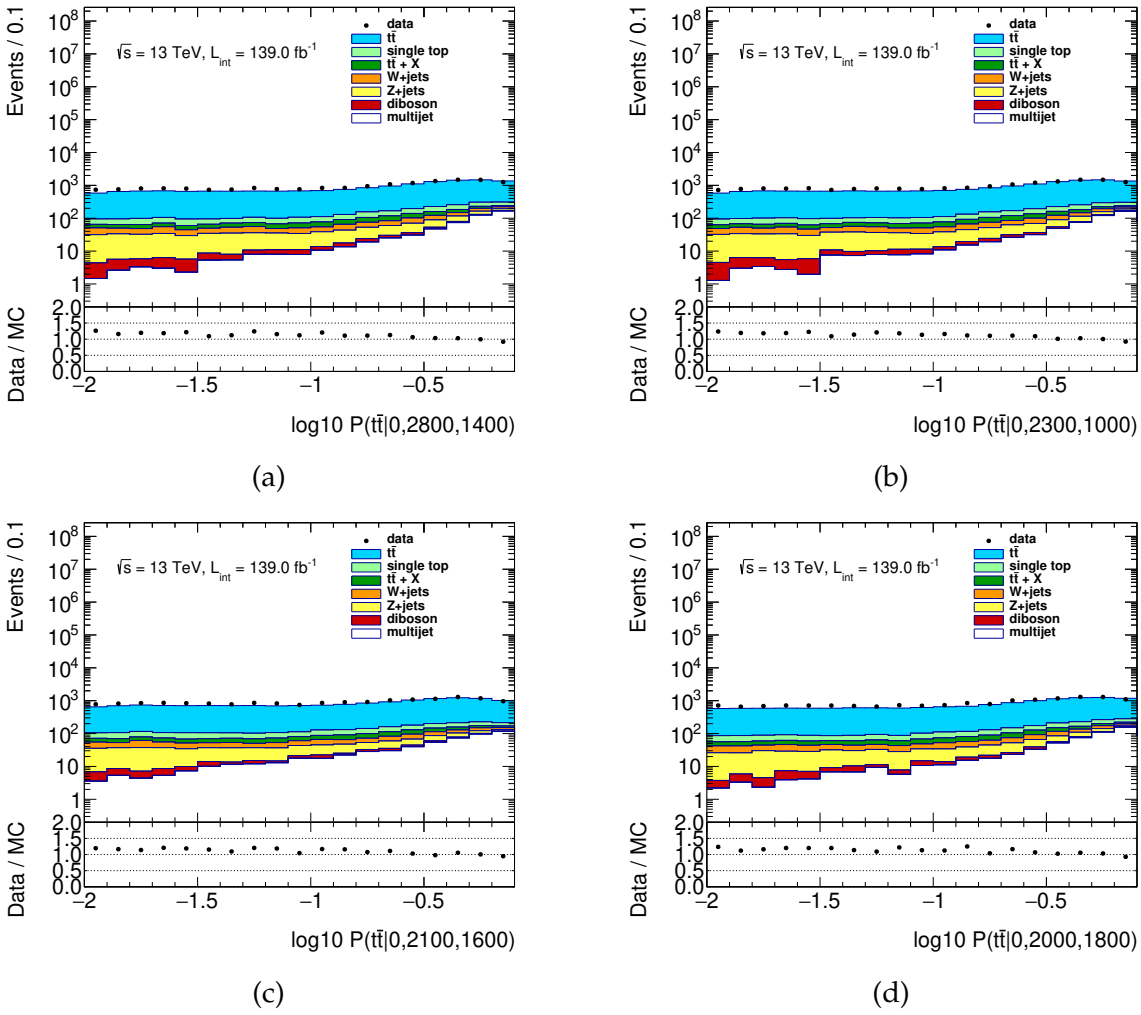
(a) (b) (c) (d)

Figure 5.17 – Data/MC comparison for the $P(\mathrm{Gbb})$ variables used for the Gbb regions, in the 0-lepton channel defined in Section 5.3.

regions to correct these distributions; this range is instead selected by the Z+jets–enriched validation regions and so any mismodelling due to this background can be identified.

5.5 Profile likelihood fits

The statistical analysis of results follows a two-step strategy, both implemented using the `HistFitter` package v0.63 [230]. Firstly, a background-only fit is performed separately in each control region in order to derive the $\mu_{t\bar{t}}$ normalization factors; These per-region normalization factors are extrapolated to their respective SRs before unblinding these regions and quantifying any potential excess. If there is a significant excess, then a discovery can be claimed; if none is observed, we move on to exclusion fits, performed in both the CRs



(a) (b)

(c) (d)

Figure 5.18 – Data/MC comparison for the $P(t\bar{t})$ variables used for the Gbb regions, in the 0-lepton channel defined in Section 5.3.

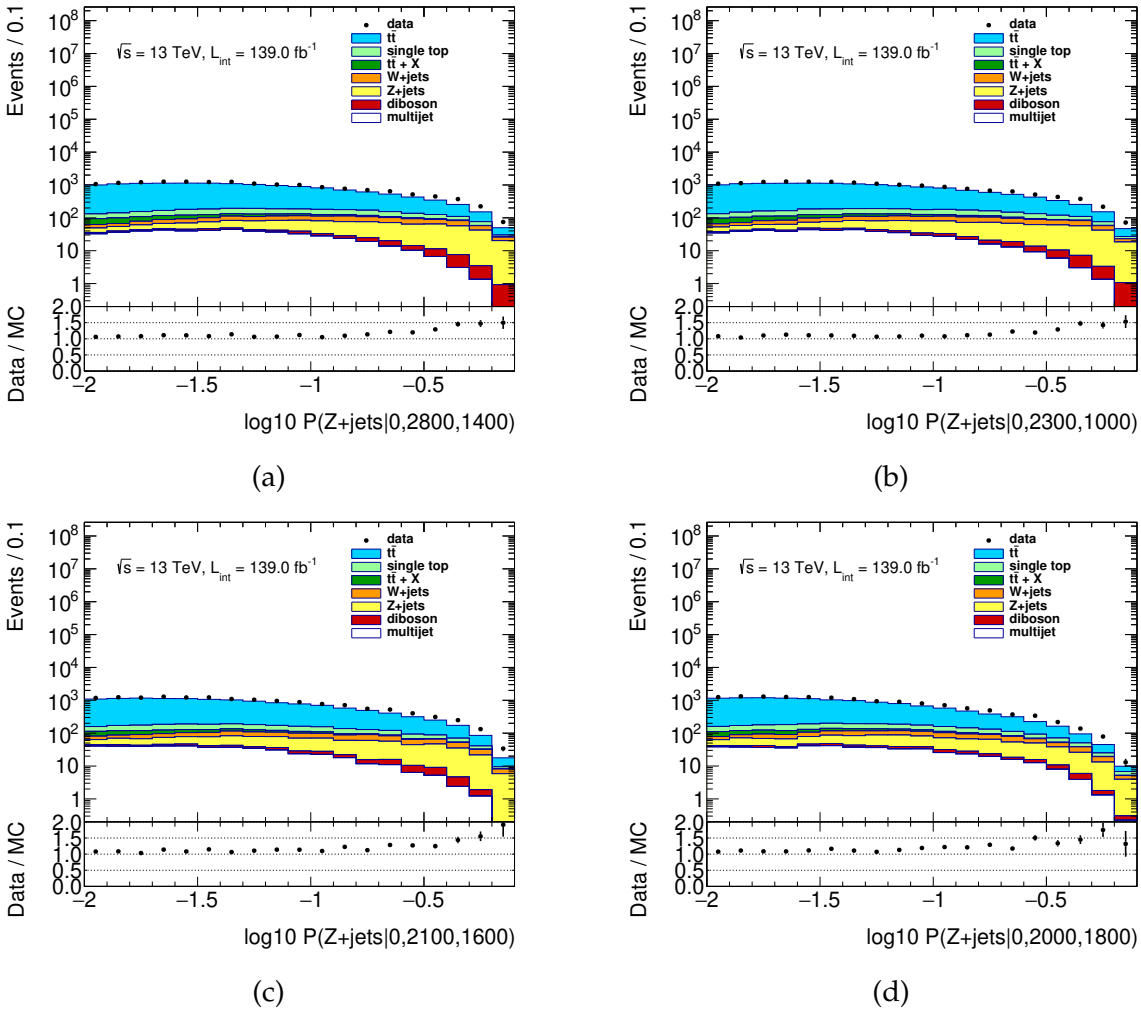


Figure 5.19 – Data/MC comparison for the $P(Z + \text{jets})$ variables used for the Gbb regions, in the 0-lepton channel defined in Section 5.3.

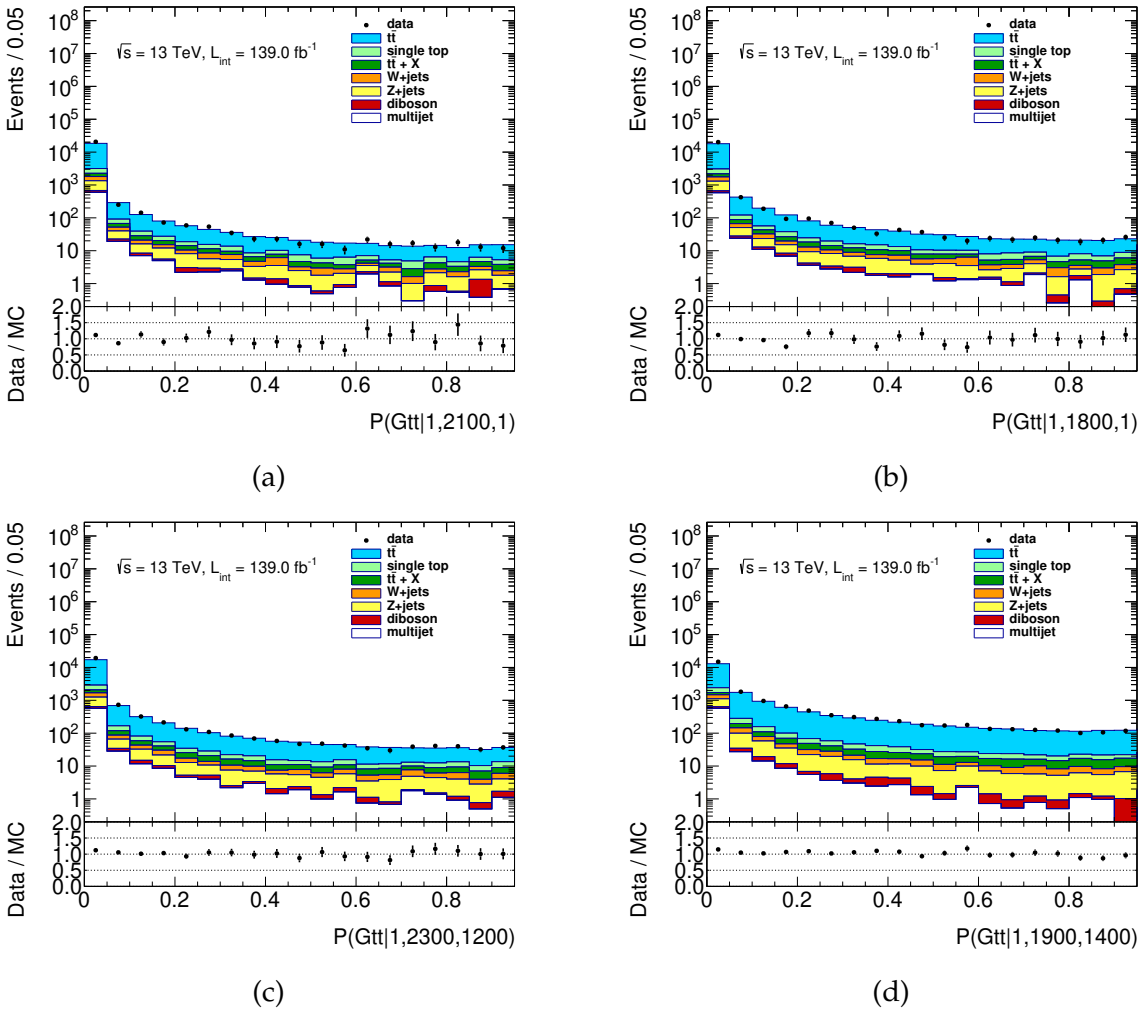


Figure 5.20 – Data/MC comparison for the $P(\mathrm{Gtt})$ variables used for the Gtt regions, in the 0-lepton channel defined in Section 5.3.

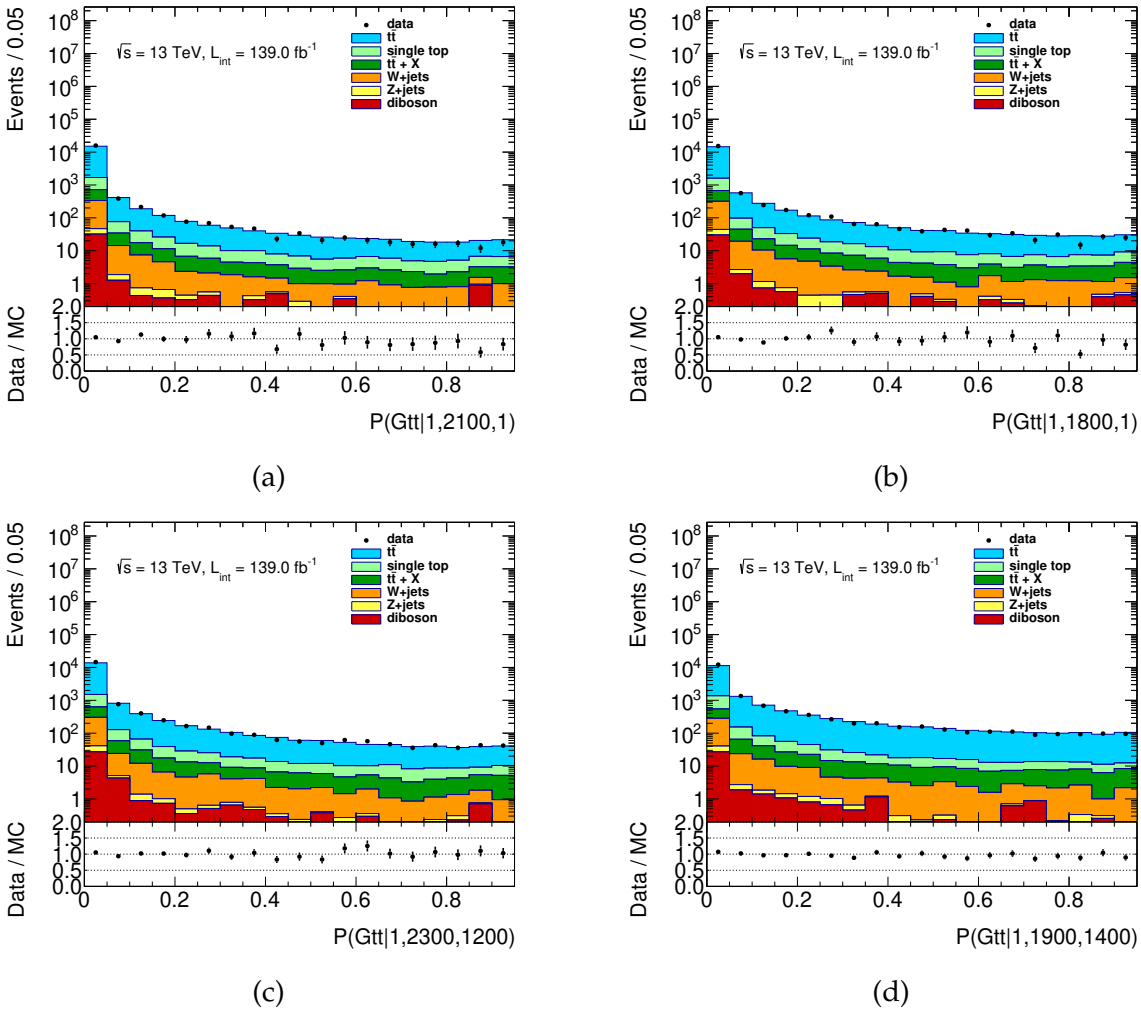


Figure 5.21 – Data/MC comparison for the $P(\mathrm{Gtt})$ variables used for the Gtt regions, in the ≥ 1-lepton channel defined in Section 5.3.

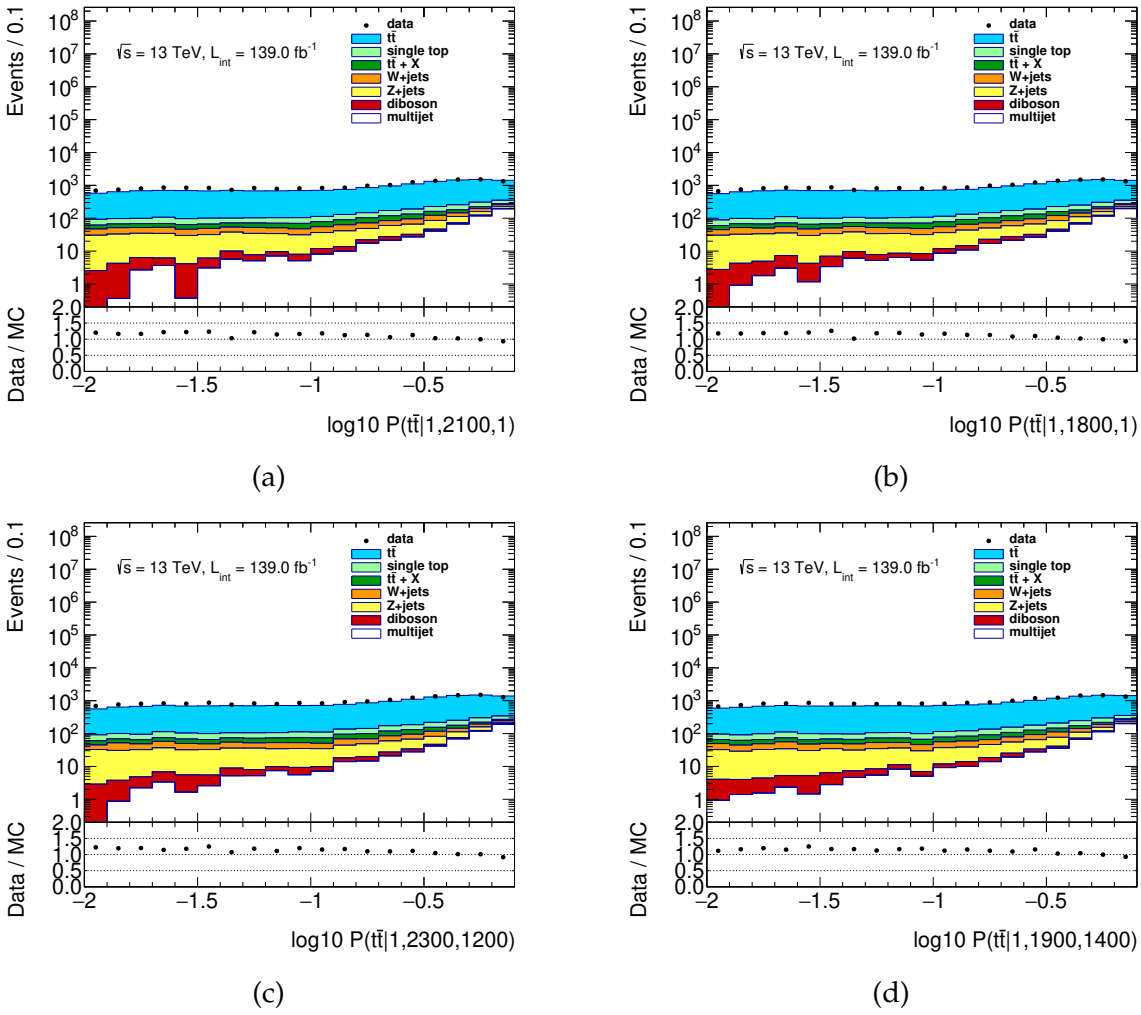
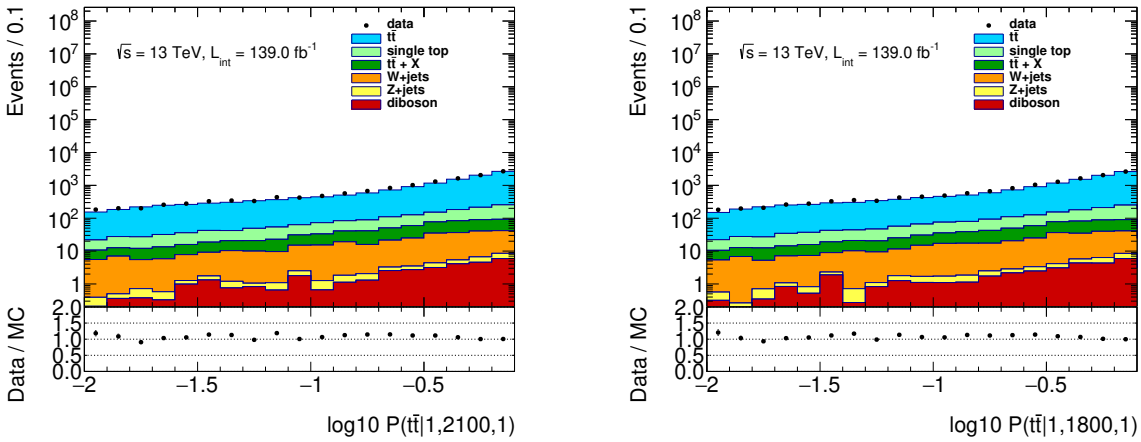
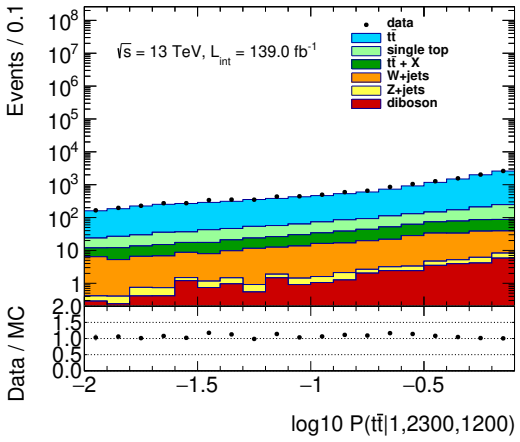


Figure 5.22 – Data/MC comparison for the $P(t\bar{t})$ variables used for the Gtt regions, in the 0-lepton channel defined in Section 5.3.

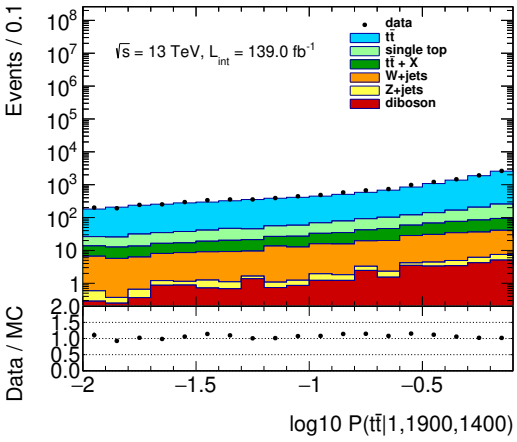


(a)

(b)



(c)



(d)

Figure 5.23 – Data/MC comparison for the $P(t\bar{t})$ variables used for the Gtt regions, in the ≥ 1-lepton channel defined in Section 5.3.

and the SRs simultaneously and used as input to likelihood-ratio tests aiming to probe which parts of the Gbb and Gtt grids are excluded by the analysis.

The likelihood function used for the background-only fit is the following:

$$L = \prod_{\mathrm{CR}} \mathcal{P}_{\mathrm{CR}} \times C_{\mathrm{syst}}. \tag{5.3}$$

The probability densities $\mathcal{P}_{\mathrm{CR}}$ are Poisson distributions:

$$\mathcal{P}_{\mathrm{CR}} = \mathrm{Poisson}(n_{\mathrm{CR}}^{\mathrm{obs}} | n_{\mathrm{CR}}^{\mathrm{exp}}). \tag{5.4}$$

The expression for the expected number of events, n_{CR}^{exp}, is nominally taken from the Monte Carlo background estimate:

$$n_{\mathrm{CR}}^{\mathrm{exp}} = (\mu_{t\bar{t}} n_{\mathrm{CR}}^{t\bar{t}} + n_{\mathrm{CR}}^{\mathrm{other}}). \tag{5.5}$$

The nominal expected yields can get scaled up or down depending on the effect of systematic uncertainties:

$$n^{\mathrm{exp}} \rightarrow n^{\mathrm{exp}}(1 + \kappa_i \alpha_i + \ldots). \tag{5.6}$$

Here, the $\kappa_i \alpha_i$ terms encompass the effects of nuisance parameters: κ_i is equal to the impact on the yield of a one-sigma fluctuation in the α_i nuisance parameter. The values for the α_i parameters are set by the fit procedure and constrained by the C_{syst} term:

$$C_{\mathrm{syst}} = \prod_i \mathcal{N}(0 | \alpha_i, 1). \tag{5.7}$$

This technique is known as *profiling* [231], and allows expressing the likelihood as a function of the parameters of interests only, the nuisance parameters settling to their most probable values given the observed data.

As mentioned earlier, if no significant excesses are observed after the background-only fit, an exclusion fit is performed this time taking into account the signal regions:

$$L = \prod_{\text{SR}} \mathcal{P}_{\text{SR}} \prod_{\text{CR}} \mathcal{P}_{\text{CR}} \times C_{\text{syst}}. \tag{5.8}$$

Again, a profile likelihood fit is performed to remove the dependency on parameters other than μ_{sig}, the signal strength. Following the Neyman-Pearson lemma [232], likelihood ratio tests are then performed for each mass hypothesis with the $q_{\mu_{\text{sig}}}$ variable [233]:

$$q_{\mu_{\text{sig}}} = -2 \log \frac{\hat{L}(\mu_{\text{sig}})}{\hat{L}}. \tag{5.9}$$

Here, $\hat{L}(\mu_{\text{sig}})$ is the result of a profiling fit with a fixed μ_{sig} value, and $\hat{L}$ lets μ_{sig} float as well to fully maximize the likelihood. Results of these hypothesis tests are finally used to produce 95% confidence level (CL) exclusion contours in the $m_{\tilde{g}}$–$m_{\tilde{\chi}_1^0}$ plane using the CLs prescription [234].

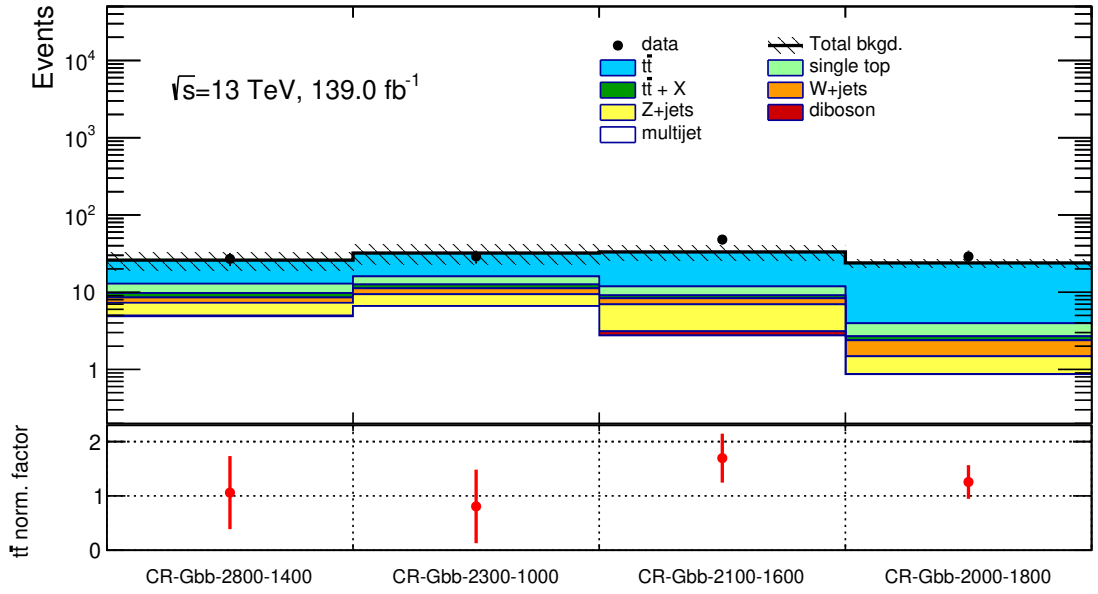
5.6 Results

Background-only fit results are found in Figures 5.24–5.26. Figure 5.24 show the result of the fits in the control regions for the Gbb and Gtt regions, including the $\mu_{t\bar{t}}$ normalization factors, which range from 0.8 to 1.7, depending on the region. The data–Monte Carlo agreement after application of these normalization factors in the validation regions can be found in Figures 5.25. These show no significant disagreement, which means that the $\mu_{t\bar{t}}$ scale factors are reasonable. The fit results in the signal regions are found in Figures 5.26, showing no significant deviation from the Standard Model background. The expected and observed event yields in the signal regions are summarized in Table 5.5; the full yield tables for all regions can be found in Appendix B.

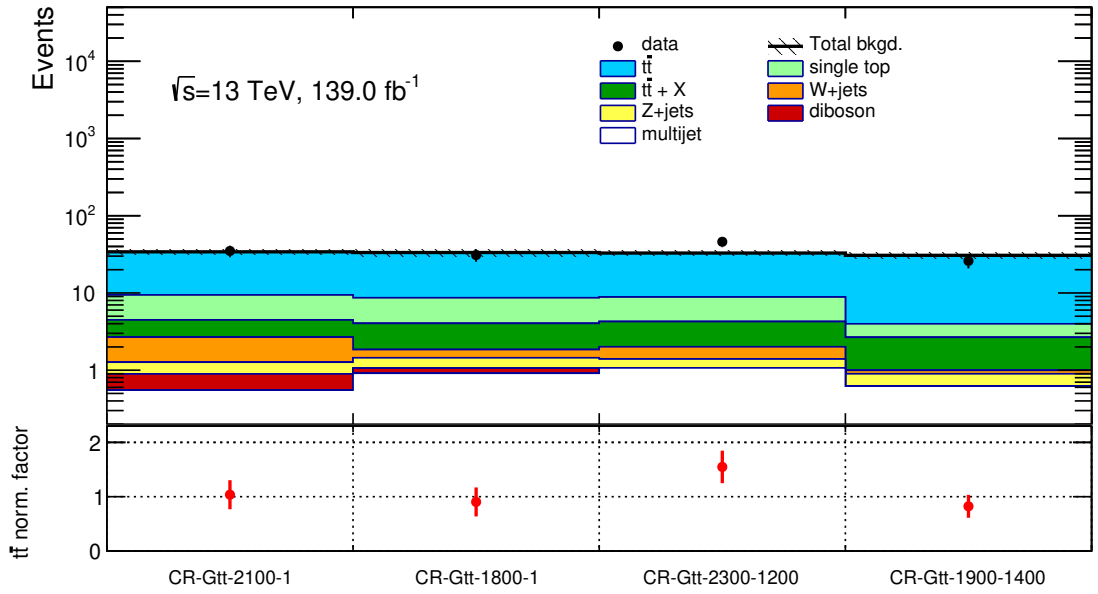
Systematic uncertainties affect the analysis if they impact the expected background yields (the normalization factors in the case of $t\bar{t}$), or the signal yields, which are used to compute the exclusion limits. The systematic uncertainties taken into account can be mostly separated into two categories, experimental and theoretical; the experimental uncertainties affect the reconstruction and identification efficiencies and the energy resolutions of all the physics objects used in the analysis, while the theoretical uncertainties affect the simulation of the various Standard Model backgrounds as well as the SUSY signals, and include variations of the renormalization and factorisation scales, parton density functions, and the amount of QCD radiation. The leading sources of experimental system-

atic uncertainties are due to the weights correcting the data–Monte Carlo discrepancy in the 1-lepton channel, jet energy resolution and scale (JER/JES), and b-tagging. Dedicated terms accounting for the generator, radiation and parton shower theoretical uncertainties are computed in all regions for the $t\bar{t}$, single-top, and V+jets backgrounds. Dummy, uncorrelated 50% uncertainty terms are included on the yields of all other backgrounds except QCD to account for the lack of dedicated estimation; the QCD term has a larger, 100% uncertainty, since this background is not estimated from a dedicated sample but with a procedure involving the rest of the Monte Carlo backgrounds. A term accounting for luminosity uncertainty is also included in the fit. The uncertainties in the regions SR-Gbb-2800-1400 and SR-Gbb-2300-1000 are dominated by the Z+jets theory uncertainty, which can reach up to 70%; this is expected since the selected kinematic phase-space (highly energetic events with many particles) in the SRs represent a very small fraction of the whole sample and therefore are highly sensitive to variations in theory parameters. The two other Gbb signal regions and the Gtt signal-regions are statistically limited. A summary of of the statistical and systematic uncertainties in the various signal regions can be found in Table 5.6, while the full breakdown of the uncertainties in all regions can be found in Appendix C.

As discussed in the previous section, since no excesses are seen after the background-only fits, the data is used to set limits in the $m_{\tilde{g}}$–$m_{\tilde{\chi}_1^0}$ plane for both the Gbb and Gtt models. The results of the model-dependent exclusion fits are shown in Figure 5.27. In the Gbb model, gluino masses below 2.3 TeV are excluded at 95% CL in the massless neutralino case; the highest excluded neutralino mass is of approximately 1.7 TeV, for a 2.3 TeV gluino. In the Gtt model, gluinos with less than 2.45 TeV of mass are excluded in the massless neutralino case, while the highest limit on the neutralino mass is of 1.65 TeV, for a 2.1 TeV gluino.

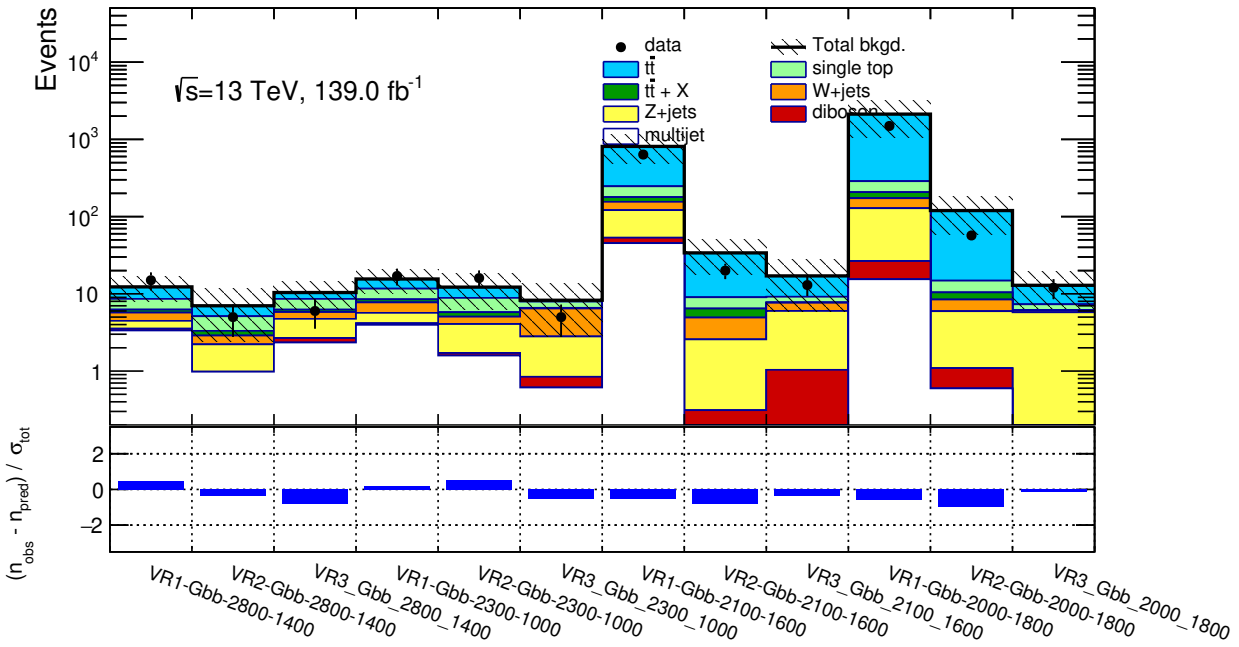


(a)

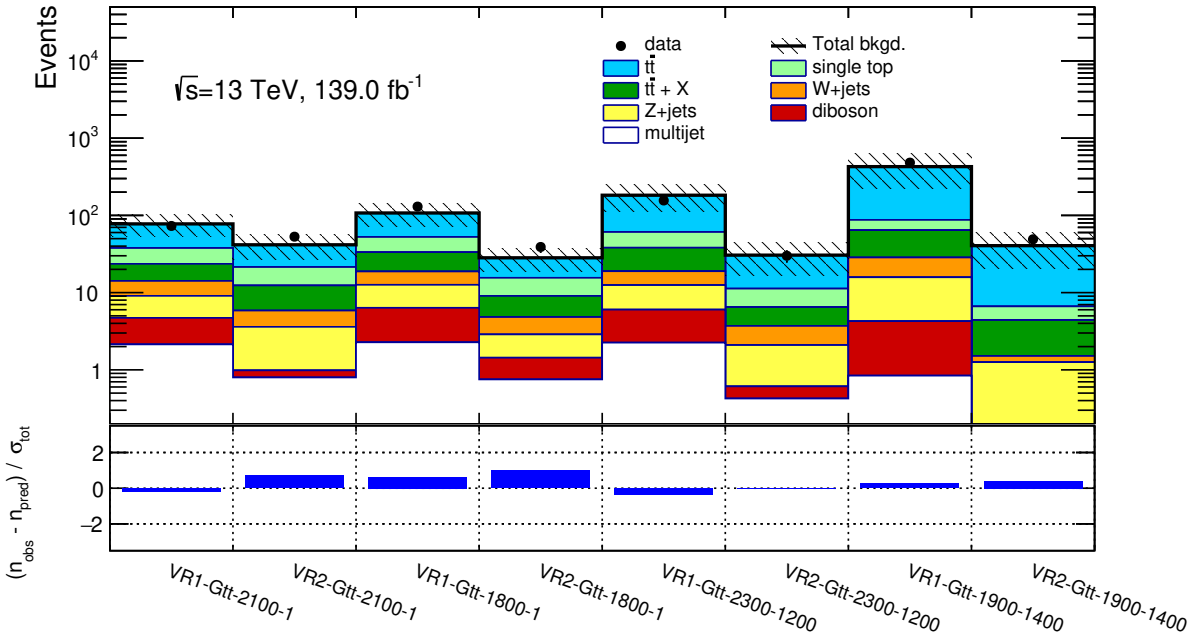


(b)

Figure 5.24 – Top pannel: Number of observed events in 139 fb^{-1} of 13 TeV ATLAS data, shown as points, and the corresponding number of expected background events, shown as histograms, in each of the (a) Gbb and (b) Gtt control regions. Bottom pannel: fitted $t\bar{t}$ normalization factors in each signal regions.

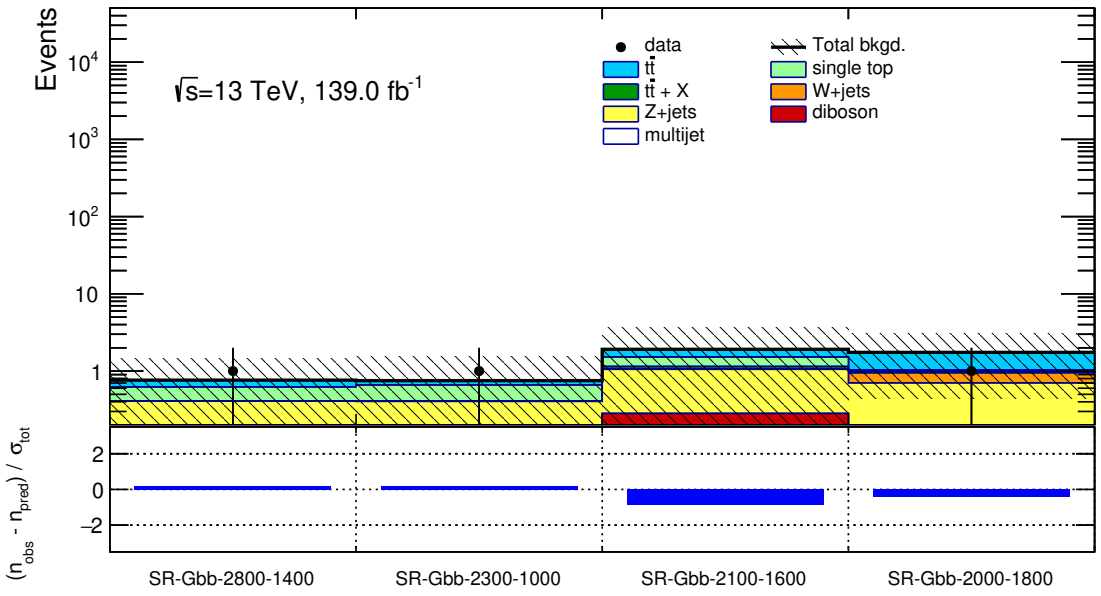


(a)

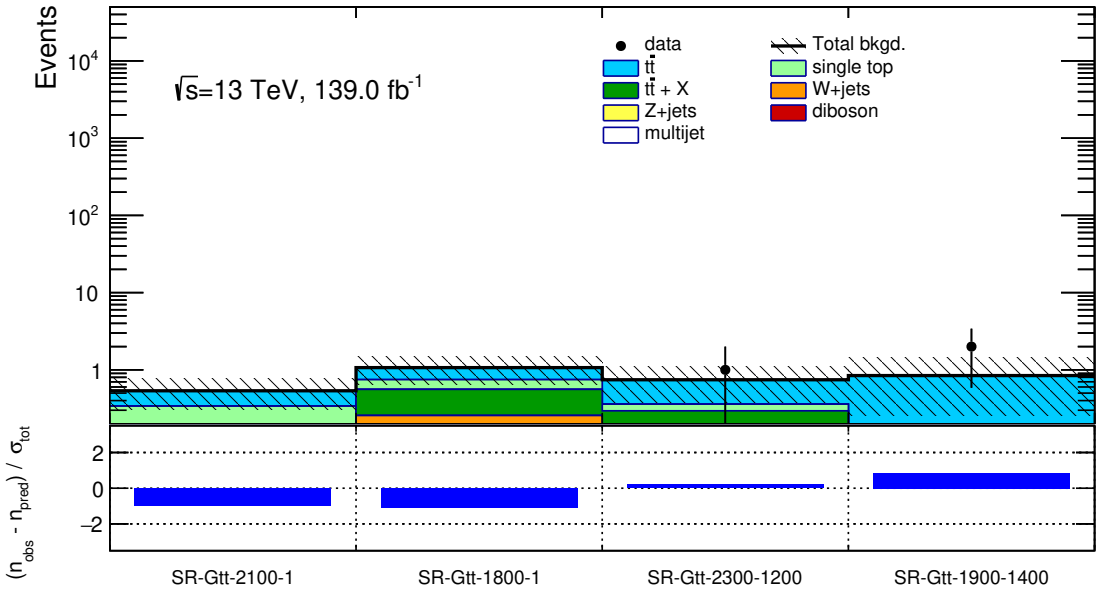


(b)

Figure 5.25 – Top pannel: Number of observed events in 139 fb^{-1} of 13 TeV ATLAS data, shown as points, and the corresponding number of expected background events, shown as histograms, in each of the (a) Gbb and (b) Gtt validation regions. Bottom pannel: pulls in each validation regions.



(a)



(b)

Figure 5.26 – Top pannel: Number of observed events in 139 fb^{-1} of 13 TeV ATLAS data, shown as points, and the corresponding number of expected background events, shown as histograms, in each of the (a) Gbb and (b) Gtt signal regions. Bottom pannel: pulls in each signal regions.

SR	Gbb-2800-1400	Gbb-2300-1000	Gbb-2100-1600	Gbb-2000-1800
Observed events	1	1	0	1
Fitted background	0.8 ± 0.7	0.8 ± 0.8	1.9 ± 1.8	1.7 ± 2.6
$t\bar{t}$	0.14 ± 0.14	0.09 ± 0.10	0.40 ± 0.31	0.7 ± 1.6
Single-top	0.21 ± 0.16	0.25 ± 0.21	0.37 ± 0.26	0.03 ± 0.04
$tt + X$	0.00 ± 0.05	0.00 ± 0.04	0.09 ± 0.07	0.05 ± 0.08
Z+jets	0.3 ± 0.6	0.3 ± 0.8	0.8 ± 1.7	0.6 ± 0.6
W+jets	-	-	$0.00^{+0.17}_{-0.00}$	0.2 ± 1.1
Diboson	-	-	0.15 ± 0.09	0.14 ± 0.17
Multijet	0.08 ± 0.09	0.10 ± 0.11	0.13 ± 0.14	0.00 ± 0.08
MC-only background	0.8 ± 0.7	0.8 ± 0.8	1.7 ± 1.8	1.6 ± 2.3

(a)

SR	Gtt-2100-1	Gtt-1800-1	Gtt-2300-1200	Gtt-1900-1400
Observed events	0	0	1	2
Fitted background	0.5 ± 0.5	1.1 ± 0.8	0.7 ± 0.7	0.8 ± 1.2
$t\bar{t}$	0.20 ± 0.25	0.3 ± 0.4	0.4 ± 0.5	0.7 ± 1.1
Single-top	0.15 ± 0.17	0.19 ± 0.23	0.06 ± 0.08	0.01 ± 0.05
$tt + X$	0.15 ± 0.21	0.3 ± 0.4	0.2 ± 0.4	0.15 ± 0.30
Z+jets	0.04 ± 0.13	0.05 ± 0.13	0.06 ± 0.23	-
W+jets	-	0.21 ± 0.35	-	-
Diboson	-	-	-	-
Multijet	$0.00^{+0.09}_{-0.00}$	$0.00^{+0.09}_{-0.00}$	$0.00^{+0.09}_{-0.00}$	$0.00^{+0.10}_{-0.00}$
MC-only background	0.5 ± 0.4	1.1 ± 0.9	0.6 ± 0.6	1.0 ± 1.3

(b)

Table 5.5 – Summary of the expected and observed event counts in the (a) Gbb and (b) Gtt signal regions. The MC-only background represents the nominally expected background level before the profile likelihood fit, while the background-specific counts are post-fit.

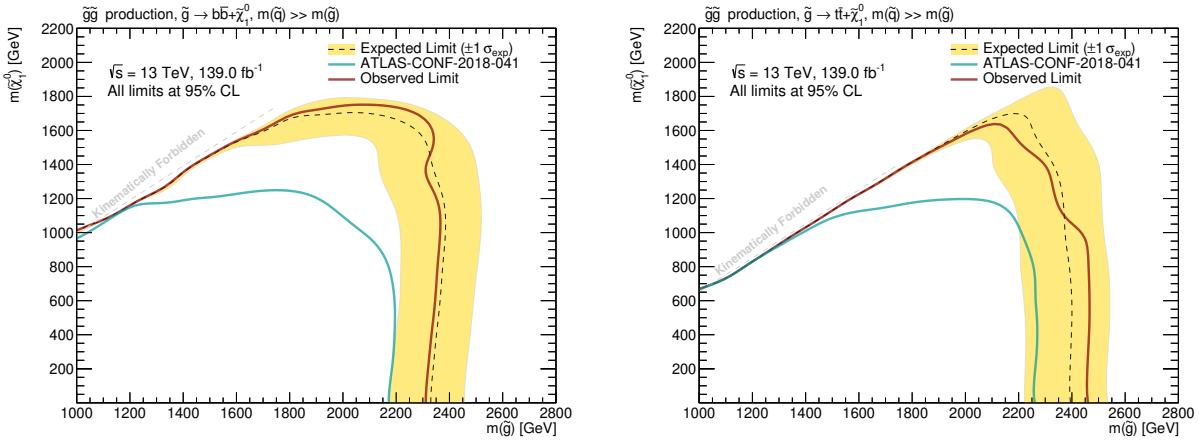
SR	Gbb-2800-1400	Gbb-2300-1000	Gbb-2100-1600	Gbb-2000-1800
Background expectation	0.76	0.75	1.91	1.75
Statistical uncertainty	±0.87	±0.87	±1.38	±1.32
Systematic uncertainty	±0.70	±0.79	±1.76	±1.30

(a)

SR	Gtt-2100-1	Gtt-1800-1	Gtt-2300-1200	Gtt-1900-1400
Background expectation	0.54	1.07	0.74	0.84
Statistical uncertainty	±0.73	±1.03	±0.86	±0.92
Systematic uncertainty	±0.23	±0.41	±0.36	±0.58

(b)

Table 5.6 – Summary of the absolute statistical and systematic uncertainties on the expected event counts in the (a) Gbb and (b) Gtt signal regions.



(a) (b)

Figure 5.27 – 95% CL exclusion contours in the $m_{\tilde{g}}$–$m_{\tilde{\chi}_1^0}$ plane for the (a) Gbb and (b) Gtt signals.

These results represent a substantial improvement over the previous ATLAS limits for these models, which were based on statistical combinations of many simple cut-and-count regions [9]. For massless neutralinos, the upper limit on the gluino mass is moved by 100 GeV and 200 GeV for the Gbb and Gtt models, respectively; this is a significant improvement even when taking into account the luminosity increase, considering that the gluino pair production cross-section falls rapidly with increasing gluino mass as seen in Figure 5.1. Moreover, the neural network results allow to set world-leading limits in the compressed mass-splitting regime, where the improvement relative to cut-based analyses is particularly significant; in this difficult regime, the neural network moves the upper limit on the neutralino mass by about 500 GeV and 400 GeV in the Gbb and Gtt models, respectively.

Conclusion

This thesis has presented the implementation of a novel deep learning-based approach to the search for supersymmetry, using only the four-momenta of selected physics objects. We've begun this journey by considering the current status of the Standard Model, an extremely successful theory of fundamental interactions, including a few of its problems. We've then turned our attention to supersymmetry, an hypothesized SM extension that could solve many of the aforementioned problems in an elegant and consistent way. Doing so, we've hopefully convinced ourselves that natural SUSY still hasn't said its final word, motivating the need to carry on the search with new and improved methods. We've then discussed the LHC and the ATLAS detector, and we've seen the huge amount of information that is recorded about each collisions by all the subsystems. Consequently, a lot of work is needed to reconstruct meaningful physics objects from this raw data and we've discussed one area in which I've been involved, that of track reconstruction in the ATLAS inner detector; my contribution consisted of characterizing and optimizing a set of neural network used to recognize and split overlapping charge clusters in the pixel detector. In later chapters, we've seen how a parameterized neural network operating on low-level inputs is able to achieve a very high sensitivity to the presence of gluinos and neutralinos in ATLAS $\sqrt{s} = 13$ TeV data for two different simplified models with off-shell stops (Gtt) or sbottoms (Gbb).

No significant excess above the SM background is recorded. In the Gbb model, gluino masses below 2.3 TeV and neutralino masses below 1.7 TeV are excluded at the 95% confidence level, while in the Gtt model, gluino and neutralino masses of less than 2.45 TeV and 1.65 TeV are ruled out, greatly increasing the previous limit on these two simplified models. For some mass hypotheses, the sensitivity increase is substantial; for instance, the ($m_{\tilde{g}} = 1.6$ TeV, $m_{\tilde{\chi}_1^0} = 1.245$ TeV) mass point of the Gtt grid has seen its statistical significance increase by 85× that of the previous results, after accounting for the difference in luminosity.

The neural network algorithm implemented for this search has proved to be very efficient and powerful, and can readily be applied to other scenarios in the future, particularly

in cases where many related signal models can share a common background model, which amplifies the benefits of the parameterized learning paradigm. Possible ways of improving future searches based on the present one include performing a shape fit on the neural network output rather than doing a simple counting experiment, and using more advanced classifiers such as recurrent neural networks (RNN), which can make it easier to use a larger selection of input physics objects that varies between events. The exclusion results presented in this thesis still do not decisively rule-out the natural spectrum, but will contribute to put pressure on this particular style of supersymmetric theories; searches such as this one should ideally be repeated in the future as the ATLAS detector continues to gather more data to make a discovery or continue pushing the boundary well outside of the low TeV range.

Bibliography

[1] J. Thomson. *Cathode rays*. Phil. Mag. Ser. 5 44 (1897), DOI: 10.1080/14786449708621070.

[2] ATLAS Collaboration. *Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC*. Physics Letters B 716.1 (2012), DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214.

[3] CMS Collaboration. *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*. Physics Letters B 716.1 (2012), DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235.

[4] ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider*. J. Inst. 3.08 (2008), DOI: 10.1088/1748-0221/3/08/S08003.

[5] CMS Collaboration. *The CMS experiment at the CERN LHC*. J. Inst. 3.08 (2008), DOI: 10.1088/1748-0221/3/08/S08004.

[6] L. Evans and P. Bryant. *LHC Machine*. J. Inst. 3.08 (2008), DOI: 10.1088/1748-0221/3/08/S08001.

[7] Y. LeCun, Y. Bengio, and G. Hinton. *Deep Learning*. Nature 521.7553 (2015), DOI: 10.1038/nature14539.

[8] ATLAS Collaboration. *Search for Supersymmetry in Final States with Missing Transverse Momentum and Multiple b-Jets in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*. JHEP 2018.6 (2018), DOI: 10.1007/JHEP06(2018)107.

[9] ATLAS Collaboration. *Search for supersymmetry in final states with missing transverse momentum and multiple b-jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. ATLAS-CONF-2018-041. URL: https://cds.cern.ch/record/2632347.

[10] "Particle Data Group". *Review of Particle Physics*. Physical Review D 98.3 (2018). DOI: 10.1103/PhysRevD.98.030001.

[11] M. Thomson. *Modern Particle Physics*. Cambridge, United Kingdom: Cambridge University Press, 2013.

[12] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. Reading, Mass: Addison-Wesley Pub. Co, 1995.

[13] P. H. Frampton. *Gauge Field Theories*. 3rd, enlarged and improved ed. Weinheim: Wiley-VCH, 2008.

[14] A. Hobson. *There Are No Particles, There Are Only Fields*. American Journal of Physics 81.3 (2013), DOI: 10.1119/1.4789885.

[15] G. F. Giudice. *The Dawn of the Post-Naturalness Era* (2017). arXiv: 1710.07663.

[16] S. Tomonaga. *On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields*. Prog Theor Phys 1.2 (1946), DOI: 10.1143/PTP.1.27.

[17] J. Schwinger. *On Quantum-Electrodynamics and the Magnetic Moment of the Electron*. Phys. Rev. 73.4 (1948), DOI: 10.1103/PhysRev.73.416.

[18] J. Schwinger. *Quantum Electrodynamics. I. A Covariant Formulation*. Phys. Rev. 74.10 (1948), DOI: 10.1103/PhysRev.74.1439.

[19] R. P. Feynman. *Space-Time Approach to Quantum Electrodynamics*. Phys. Rev. 76.6 (1949), DOI: 10.1103/PhysRev.76.769.

[20] R. P. Feynman. *The Theory of Positrons*. Phys. Rev. 76.6 (1949), DOI: 10.1103/PhysRev.76.749.

[21] R. P. Feynman. *Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction*. Phys. Rev. 80.3 (1950), DOI: 10.1103/PhysRev.80.440.

[22] T. Aoyama, T. Kinoshita, and M. Nio. *Revised and Improved Value of the QED Tenth-Order Electron Anomalous Magnetic Moment*. Phys. Rev. D 97.3 (2018), DOI: 10.1103/PhysRevD.97.036001.

[23] M. Gell-Mann. *A Schematic Model of Baryons and Mesons*. Physics Letters 8.3 (1964), DOI: 10.1016/S0031-9163(64)92001-3.

[24] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. *Advantages of the Color Octet Gluon Picture*. Physics Letters B 47.4 (1973), DOI: 10.1016/0370-2693(73)90625-4.

[25] D. J. Gross and F. Wilczek. *Ultraviolet Behavior of Non-Abelian Gauge Theories*. Phys. Rev. Lett. 30.26 (1973), DOI: 10.1103/PhysRevLett.30.1343.

[26] H. D. Politzer. *Reliable Perturbative Results for Strong Interactions?* Phys. Rev. Lett. 30.26 (1973), DOI: 10.1103/PhysRevLett.30.1346.

[27] F. J. Dyson. *The Radiation Theories of Tomonaga, Schwinger, and Feynman*. Phys. Rev. 75.3 (1949), DOI: 10.1103/PhysRev.75.486.

[28] Particle Data Group. *Quantum Chromodynamics*. *Review of Particle Physics*. Vol. 98. 3. 2018. DOI: 10.1103/PhysRevD.98.030001.

[29] B. Andersson et al. *Parton Fragmentation and String Dynamics*. Physics Reports 97.2 (1983), DOI: 10.1016/0370-1573(83)90080-7.

[30] ATLAS Collaboration. *Properties of jet fragmentation using charged particles measured with the ATLAS detector in pp collisions at* $\sqrt{s} = 13$ *TeV*. Phys. Rev. D 100.5 (2019), DOI: 10.1103/PhysRevD.100.052011. arXiv: 1906.09254.

[31] T. N. Collaboration et al. *Parton Distributions for the LHC Run II*. JHEP 2015.4 (2015), DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849.

[32] D. Lindley. *Focus: Landmarks—Breaking the Mirror*. Physics 22 (2008).

[33] C. S. Wu et al. *Experimental Test of Parity Conservation in Beta Decay*. Phys. Rev. 105.4 (1957), DOI: 10.1103/PhysRev.105.1413.

[34] F. Englert and R. Brout. *Broken Symmetry and the Mass of Gauge Vector Mesons*. Phys. Rev. Lett. 13.9 (1964), DOI: 10.1103/PhysRevLett.13.321.

[35] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. *Global Conservation Laws and Massless Particles*. Phys. Rev. Lett. 13.20 (1964), DOI: 10.1103/PhysRevLett.13.585.

[36] P. W. Higgs. *Broken Symmetries and the Masses of Gauge Bosons*. Phys. Rev. Lett. 13.16 (1964), DOI: 10.1103/PhysRevLett.13.508.

[37] Particle Data Group. *Status of Higgs Boson Physics*. *Review of Particle Physics*. Vol. 98. 3. 2018. DOI: 10.1103/PhysRevD.98.030001.

[38] S. L. Glashow. *The Renormalizability of Vector Meson Interactions*. Nuclear Physics 10 (1959), DOI: 10.1016/0029-5582(59)90196-8.

[39] A. Salam and J. C. Ward. *Weak and Electromagnetic Interactions*. Nuovo Cim 11.4 (1959), DOI: 10.1007/BF02726525.

[40] S. Weinberg. *A Model of Leptons*. Phys. Rev. Lett. 19.21 (1967), DOI: 10.1103/PhysRevLett.19.1264.

[41] N. Cabibbo. *Unitary Symmetry and Leptonic Decays*. Phys. Rev. Lett. 10.12 (1963), DOI: 10.1103/PhysRevLett.10.531.

[42] M. Kobayashi and T. Maskawa. *CP-Violation in the Renormalizable Theory of Weak Interaction*. Prog Theor Phys 49.2 (1973), DOI: 10.1143/PTP.49.652.

[43] ATLAS Collaboration. *Standard Model Summary Plots Spring 2020*. ATL-PHYS-PUB-2020-010. 2020. URL: https://cdsweb.cern.ch/record/2718937.

[44] P. Nelson. *Naturalness in Theoretical Physics: Internal Constraints on Theories, Especially the Requirement of Naturalness, Play a Pivotal Role in Physics*. American Scientist 73.1 (1985),

[45] L. Susskind. *Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory*. Phys. Rev. D 20.10 (1979), DOI: 10.1103/PhysRevD.20.2619.

[46] M. Veltman. *The Infrared-Ultraviolet Connection*. *Origin of Symmetries*. WORLD SCIENTIFIC, 1991, pp. 526–546. DOI: 10.1142/9789814329057_0034.

[47] G. Hooft. *Naturalness, Chiral Symmetry, and Spontaneous Chiral Symmetry Breaking*. *Recent Developments in Gauge Theories*. Ed. by G. Hooft et al. NATO Advanced Study Institutes Series. Boston, MA: Springer US, 1980, pp. 135–157. DOI: 10.1007/978-1-4684-7571-5_9.

[48] S. P. Martin. *A Supersymmetry Primer* (1997). DOI: 10.1142/9789812839657_0001. arXiv: hep-ph/9709356.

[49] G. F. Giudice. *Naturally Speaking: The Naturalness Criterion and Physics at the LHC* (2008), DOI: 10.1142/9789812779762_0010. arXiv: 0801.2562.

[50] C. M. Will. *The Confrontation between General Relativity and Experiment*. Living Rev. Relativ. 17.1 (2014), DOI: 10.12942/lrr-2014-4.

[51] C. P. Burgess. *Quantum Gravity in Everyday Life: General Relativity as an Effective Field Theory*. Living Rev. Relativ. 7.1 (2004), DOI: 10.12942/lrr-2004-5.

[52] H. Baer and X. Tata. *Weak Scale Supersymmetry: From Superfields to Scattering Events*. Cambridge, UK ; New York: Cambridge University Press, 2006.

[53] M. van Beekveld, S. Caron, and R. Ruiz de Austri. *The current status of fine-tuning in supersymmetry*. JHEP 01 (2020), DOI: 10.1007/JHEP01(2020)147. arXiv: 1906.10706 [hep-ph].

[54] C. Csaki. *TASI Lectures on Extra Dimensions and Branes* (2004). arXiv: hep-ph/0404096.

[55] M. J. Dugan, H. Georgi, and D. B. Kaplan. *Anatomy of a Composite Higgs Model*. Nuclear Physics B 254 (1985), DOI: 10.1016/0550-3213(85)90221-4.

[56] K. Garrett and G. Duda. *Dark Matter: A Primer*. Advances in Astronomy 2011 (2011), DOI: 10.1155/2011/968283. arXiv: 1006.2483.

[57] J. de Swart, G. Bertone, and J. van Dongen. *How Dark Matter Came to Matter*. Nat Astron 1.3 (2017), DOI: 10.1038/s41550-017-0059. arXiv: 1703.00013.

[58] Y. Sofue and V. Rubin. *Rotation Curves of Spiral Galaxies*. Annu. Rev. Astron. Astrophys. 39.1 (2001), DOI: 10.1146/annurev.astro.39.1.137.

[59] V. C. Rubin. *Dark Matter in Spiral Galaxies*. Scientific American 248.6 (1983),

[60] M. Markevitch et al. *Direct Constraints on the Dark Matter Self-Interaction Cross-Section from the Merging Galaxy Cluster 1E0657-56*. ApJ 606.2 (2004), DOI: 10.1086/383178. arXiv: astro-ph/0309303.

[61] N. Jarosik et al. *Seven-Year WMAP Observations: Sky Maps, Systematic Errors, and Basic Results*. ApJS 192.2 (2011), DOI: 10.1088/0067-0049/192/2/14.

[62] Planck Collaboration. *Planck 2018 Results. VI. Cosmological Parameters* (2019). arXiv: 1807.06209 [astro-ph].

[63] Particle Data Group. *Dark Matter*. *Review of Particle Physics*. Vol. 98. 3. 2018. DOI: 10.1103/PhysRevD.98.030001.

[64] M. Milgrom. *A Modification of the Newtonian Dynamics as a Possible Alternative to the Hidden Mass Hypothesis*. The Astrophysical Journal 270 (1983), DOI: 10.1086/161130.

[65] S. S. McGaugh. *A Tale of Two Paradigms: The Mutual Incommensurability of LCDM and MOND*. Can. J. Phys. 93.2 (2015), DOI: 10.1139/cjp-2014-0203. arXiv: 1404.7525.

[66] G. Bertone, D. Hooper, and J. Silk. *Particle Dark Matter: Evidence, Candidates and Constraints*. Physics Reports 405.5 (2005), DOI: 10.1016/j.physrep.2004.08.031.

[67] Yu. A. Golfand and E. P. Likhtman. *Extension of the Algebra of Poincare Group Generators and Violation of p Invariance*. JETP Lett. 13 (1971). [Pisma Zh. Eksp. Teor. Fiz. **13** (1971) 452],

[68] D. V. Volkov and V. P. Akulov. *Is the Neutrino a Goldstone Particle?* Phys. Lett. B 46 (1973), DOI: 10.1016/0370-2693(73)90490-5.

[69] J. Wess and B. Zumino. *Supergauge Transformations in Four-Dimensions*. Nucl. Phys. B 70 (1974), DOI: 10.1016/0550-3213(74)90355-1.

[70] J. Wess and B. Zumino. *Supergauge Invariant Extension of Quantum Electrodynamics*. Nucl. Phys. B 78 (1974), DOI: 10.1016/0550-3213(74)90112-6.

[71] S. Ferrara and B. Zumino. *Supergauge Invariant Yang-Mills Theories*. Nucl. Phys. B 79 (1974), DOI: 10.1016/0550-3213(74)90559-8.

[72] A. Salam and J. A. Strathdee. *Supersymmetry and Nonabelian Gauges*. Phys. Lett. B 51 (1974), DOI: 10.1016/0370-2693(74)90226-3.

[73] G. L. Kane and M. A. Shifman. *The Supersymmetric World: The Beginnings of the Theory*. River Edge, NJ: World Scientific, 2001.

[74] E. Witten. *Dynamical Breaking of Supersymmetry*. Nuclear Physics B 188.3 (1981), DOI: 10.1016/0550-3213(81)90006-7.

[75] R. K. Kaul. *Gauge Hierarchy in a Supersymmetric Model*. Physics Letters B 109.1 (1982), DOI: 10.1016/0370-2693(82)90453-1.

[76] Particle Data Group. *Supersymmetry, Part I (Theory). Review of Particle Physics*. Vol. 98. 3. 2018. DOI: 10.1103/PhysRevD.98.030001.

[77] Particle Data Group. *Supersymmetry, Part II (Experiment). Review of Particle Physics*. Vol. 98. 3. 2018. DOI: 10.1103/PhysRevD.98.030001.

[78] M. Drees, P. Roy, and R. M. Godbole. *Theory and Phenomenology of Sparticles: An Account of Four-Dimensional N=1 Supersymmetry in High Energy Physics*. Hackensack, NJ: World Scientific, 2004.

[79] P. Fayet. *Supersymmetry and Weak, Electromagnetic and Strong Interactions*. Phys. Lett. B 64 (1976), DOI: 10.1016/0370-2693(76)90319-1.

[80] P. Fayet. *Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions*. Phys. Lett. B 69 (1977), DOI: 10.1016/0370-2693(77)90852-8.

[81] H. Georgi and S. L. Glashow. *Gauge Theories Without Anomalies*. Phys. Rev. D 6.2 (1972), DOI: 10.1103/PhysRevD.6.429.

[82] *Going Nowhere Fast*. https://www.americanscientist.org/article/going-nowhere-fast. 2018.

[83] M. Dine and W. Fischler. *A Phenomenological Model of Particle Physics Based on Supersymmetry*. Phys. Lett. B 110 (1982), DOI: 10.1016/0370-2693(82)91241-2.

[84] L. Alvarez-Gaume, M. Claudson, and M. B. Wise. *Low-Energy Supersymmetry*. Nucl. Phys. B 207 (1982), DOI: 10.1016/0550-3213(82)90138-9.

[85] C. R. Nappi and B. A. Ovrut. *Supersymmetric Extension of the SU(3) x SU(2) x U(1) Model*. Phys. Lett. B 113 (1982), DOI: 10.1016/0370-2693(82)90418-X.

[86] A. H. Chamseddine, R. L. Arnowitt, and P. Nath. *Locally Supersymmetric Grand Unification*. Phys. Rev. Lett. 49 (1982), DOI: 10.1103/PhysRevLett.49.970.

[87] R. Barbieri, S. Ferrara, and C. A. Savoy. *Gauge Models with Spontaneously Broken Local Supersymmetry*. Phys. Lett. B 119 (1982), DOI: 10.1016/0370-2693(82)90685-2.

[88] G. L. Kane et al. *Study of constrained minimal supersymmetry*. Phys. Rev. D 49 (1994), DOI: 10.1103/PhysRevD.49.6173. arXiv: hep-ph/9312272.

[89] J. K. K. Liu. *Search for electroweak production of supersymmetric states in scenarios with compressed mass spectra at $\sqrt{s} = 13$ TeV with ATLAS*. ATL-PHYS-SLIDE-2018-891. URL: https://cds.cern.ch/record/2642408.

[90] R. Barbieri and G. F. Giudice. *Upper Bounds on Supersymmetric Particle Masses*. Nucl. Phys. B 306 (1988), DOI: 10.1016/0550-3213(88)90171-X.

[91] M. R. Buckley, A. Monteux, and D. Shih. *Precision Corrections to Fine Tuning in SUSY*. JHEP 2017.6 (2017), DOI: 10.1007/JHEP06(2017)103. arXiv: 1611.05873.

[92] H. Baer et al. *Radiative Natural SUSY with a 125 GeV Higgs Boson*. Phys. Rev. Lett. 109.16 (2012), DOI: 10.1103/PhysRevLett.109.161802. arXiv: 1207.3343.

[93] H. Baer, V. Barger, and D. Mickelson. *How Conventional Measures Overestimate Electroweak Fine-Tuning in Supersymmetric Theory*. Phys. Rev. D 88.9 (2013), DOI: 10.1103/PhysRevD.88.095013. arXiv: 1309.2984.

[94] M. van Beekveld, S. Caron, and R. R. de Austri. *The Current Status of Fine-Tuning in Supersymmetry*. JHEP 2020.1 (2020), DOI: 10.1007/JHEP01(2020)147. arXiv: 1906.10706.

[95] M. Peiro and S. Robles. *Low-Mass Neutralino Dark Matter in Supergravity Scenarios: Phenomenology and Naturalness*. J. Cosmol. Astropart. Phys. 2017.05 (2017), DOI: 10.1088/1475-7516/2017/05/010. arXiv: 1612.00460.

[96] G. R. Farrar and P. Fayet. *Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry*. Phys. Lett. B 76 (1978), DOI: 10.1016/0370-2693(78)90858-4.

[97] H. Goldberg. *Constraint on the Photino Mass from Cosmology*. Phys. Rev. Lett. 50 (1983), DOI: 10.1103/PhysRevLett.50.1419. Erratum: *Erratum: Constraint on the Photino Mass from Cosmology*. Phys. Rev. Lett. 103 (2009), DOI: 10.1103/PhysRevLett.103.099905.

[98] J. R. Ellis et al. *Supersymmetric Relics from the Big Bang*. Nucl. Phys. B 238 (1984), DOI: 10.1016/0550-3213(84)90461-9.

[99] A. Djouadi et al. *The Minimal supersymmetric standard model: Group summary report*. 1998. arXiv: hep-ph/9901246.

[100] C. F. Berger et al. *Supersymmetry Without Prejudice*. JHEP 02 (2009), DOI: 10.1088/1126-6708/2009/02/023. arXiv: 0812.0980 [hep-ph].

[101] ATLAS Collaboration. *Summary of the ATLAS Experiment's Sensitivity to Supersymmetry after LHC Run 1 - Interpreted in the Phenomenological MSSM*. JHEP 2015.10 (2015), DOI: 10.1007/JHEP10(2015)134. arXiv: 1508.06608.

[102] J. Alwall et al. *Searching for Directly Decaying Gluinos at the Tevatron*. Phys. Lett. B 666 (2008), DOI: 10.1016/j.physletb.2008.06.065. arXiv: 0803.0019 [hep-ph].

[103] J. Alwall, P. Schuster, and N. Toro. *Simplified Models for a First Characterization of New Physics at the LHC*. Phys. Rev. D 79 (2009), DOI: 10.1103/PhysRevD.79.075020. arXiv: 0810.3921 [hep-ph].

[104] D. Alves et al. *Simplified Models for LHC New Physics Searches*. J. Phys. G 39 (2012), DOI: 10.1088/0954-3899/39/10/105005. arXiv: 1105.2838 [hep-ph].

[105] R. Aβmann, M. Lamont, and S. Myers. *A Brief History of the LEP Collider*. Nuclear Physics B - Proceedings Supplements. Proceedings of the 7th Topical Seminar 109.2 (2002), DOI: 10.1016/S0920-5632(02)90005-8.

[106] E. Boltezar et al. *The New CERN 50-MeV Linac*. Proceedings of the 1979 Linear Accelerator Conference (1979).

[107] J.-P. Burnet et al. *Fifty years of the CERN Proton Synchrotron: Volume 1*. CERN Yellow Reports: Monographs. Geneva: CERN, 2011. DOI: 10.5170/CERN-2011-004. URL: https://cds.cern.ch/record/1359959.

[108] S. Gilardoni et al. *Fifty years of the CERN Proton Synchrotron: Volume 2*. Ed. by S. Gilardoni. CERN Yellow Reports: Monographs. Comments: 58 pages, published as CERN Yellow Report https://cds.cern.ch/record/1597087?ln=en. Geneva: CERN, 2013. DOI: 10.5170/CERN-2013-005. URL: https://cds.cern.ch/record/1597087.

[109] *Super Proton Synchrotron marks its 25th birthday*. CERN Courier 41.6 (2001), URL: https://cds.cern.ch/record/1733215.

[110] M. Banner et al. *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN Pp Collider*. Physics Letters B 122.5 (1983), DOI: 10.1016/0370-2693(83)91605-2.

[111] G. Arnison et al. *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at S=540 GeV*. Physics Letters B 122.1 (1983), DOI: 10.1016/0370-2693(83)91177-2.

[112] G. Arnison et al. *Experimental Observation of Lepton Pairs of Invariant Mass around 95 GeV/C2 at the CERN SPS Collider*. Physics Letters B 126.5 (1983), DOI: 10.1016/0370-2693(83)90188-0.

[113] P. Bagnaia et al. *Evidence for $Z^0 \to e^+e^-$ at the CERN Pp Collider*. Physics Letters B 129.1 (1983), DOI: 10.1016/0370-2693(83)90744-X.

[114] J. Haffner. *The CERN accelerator complex* (2013). URL: https://cds.cern.ch/record/1621894.

[115] J.-L. Caron. *LHC Layout* (1997). URL: https://cds.cern.ch/record/841573.

[116] W. Herr, ed. *CAS - CERN Accelerator School: Advanced Accelerator Physics Course*. CAS. CERN. Geneva: CERN, 2014. DOI: 10.5170/CERN-2014-009. arXiv: 1601.07311 [physics.acc-ph].

[117] Particle Data Group. *Particle Detectors at Accelerators*. *Review of Particle Physics*. Vol. 98. 3. 2018. DOI: 10.1103/PhysRevD.98.030001.

[118] J.-P. Burnet. *Putting It into Practice* (2016). DOI: 10.5170/CERN-2015-003.445. arXiv: 1607.01596 [physics].

[119] D. Denisov et al. *Proposal for Running the Tevatron Colliderin 2012-2014*. Materials for P5 on Tevatron Extension (2010).

[120] Particle Data Group. *High-Energy Collider Parameters*. *Review of Particle Physics*. Vol. 98. 3. 2018. DOI: 10.1103/PhysRevD.98.030001.

[121] H. Braun. *Emittance Diagnostics*. CERN accelerator school: Beam diagnostics (2008).

[122] Hofle, Wolfgang and Li, Kevin. *Transverse Feedback in the HL-LHC Era*. 5th Joint HiLumi LHC-LARP Annual Meeting (2015).

[123] R. A. Fernandez. *LHC Report: focus on luminosity*. BUL-NA-2016-084 (2016), URL: https://cds.cern.ch/record/2155680.

[124] LHCb Collaboration. *The LHCb Detector at the LHC*. J. Inst. 3.08 (2008), DOI: 10.1088/1748-0221/3/08/S08005.

[125] ALICE Collaboration. *The ALICE experiment at the CERN LHC*. J. Inst. 3.08 (2008), DOI: 10.1088/1748-0221/3/08/S08002.

[126] LHCf Collaboration. *The LHCf detector at the CERN Large Hadron Collider*. J. Inst. 3.08 (2008), DOI: 10.1088/1748-0221/3/08/S08006.

[127] TOTEM Collaboration. *The TOTEM Experiment at the CERN Large Hadron Collider*. J. Inst. 3.08 (2008), DOI: 10.1088/1748-0221/3/08/S08007.

[128] MoEDAL Collaboration. *Technical Design Report of the MoEDAL Experiment*. CERN-LHCC-2009-006. URL: http://cds.cern.ch/record/1181486.

[129] J. L. Feng et al. *FASER: ForwArd Search ExpeRiment at the LHC*. Phys. Rev. D 97.3 (2018), DOI: 10.1103/PhysRevD.97.035001. arXiv: 1708.09389.

[130] J. Pequenao. *Computer generated image of the whole ATLAS detector*. 2008. URL: https://cds.cern.ch/record/1095924.

[131] J. Pequenao. *Computer generated image of the ATLAS inner detector*. 2008. URL: https://cds.cern.ch/record/1095926.

[132] M. Capeans et al. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19. 2010. URL: https://cds.cern.ch/record/1291633.

[133] J. Pequenao. *Computer generated images of the Pixel, part of the ATLAS inner detector.* 2008. URL: https://cds.cern.ch/record/1095925.

[134] C. Marcelloni De Oliveira. *IBL installation into the inner detector of the ATLAS Experiment side C*. General Photo. 2014. URL: https://cds.cern.ch/record/1702006.

[135] J. Pequenao. *Computer Generated image of the ATLAS calorimeter*. 2008. URL: https://cds.cern.ch/record/1095927.

[136] L. M. de Andrade Filho et al. *Three-Dimensional Event Visualization for the ATLAS Calorimeter*. Computer Physics Communications 183.2 (2012), DOI: 10.1016/j.cpc.2011.09.004.

[137] L. Mandelli. *The Atlas Electromagnetic Calorimeters: Features and Performance*. Int. J. Mod. Phys. A 25.09 (2010), DOI: 10.1142/S0217751X10049323.

[138] M. L. Andrieux et al. *Construction and Test of the First Two Sectors of the ATLAS Barrel Liquid Argon Presampler*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 479.2 (2002), DOI: 10.1016/S0168-9002(01)00943-3.

[139] Particle Data Group. *Passage of Particles Through Matter*. *Review of Particle Physics*. Vol. 98. 3. 2018. DOI: 10.1103/PhysRevD.98.030001.

[140] *ATLAS Approved DAQ Plots*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ.

[141] ATLAS Collaboration. *Performance of the ATLAS global transverse-momentum triggers at $\sqrt{s}$ = 8 TeV*. ATL-DAQ-PUB-2018-001. 2018. URL: https://cds.cern.ch/record/2311730.

[142] ATLAS Collaboration. *Measurement of inclusive jet and dijet cross-sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. JHEP 2018.5 (2018), DOI: 10.1007/JHEP05(2018)195. arXiv: 1711.02692.

[143] ATLAS Collaboration. *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2*. Eur. Phys. J. C 77.10 (2017), DOI: 10.1140/epjc/s10052-017-5225-7. arXiv: 1704.07983.

[144] R. Frühwirth. *Application of Kalman Filtering to Track and Vertex Fitting*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 262.2-3 (1987), DOI: 10.1016/0168-9002(87)90887-4.

[145] J. Schmidhuber. *Deep Learning in Neural Networks: An Overview*. Neural Networks 61 (2015), DOI: 10.1016/j.neunet.2014.09.003.

[146] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Adaptive Computation and Machine Learning. Cambridge, Massachusetts: The MIT Press, 2016.

[147] G. Cybenko. *Approximation by Superpositions of a Sigmoidal Function*. Math. Control Signal Systems 2.4 (1989), DOI: 10.1007/BF02551274.

[148] K. Hornik. *Approximation Capabilities of Multilayer Feedforward Networks*. Neural Networks 4.2 (1991), DOI: 10.1016/0893-6080(91)90009-T.

[149] M. Leshno et al. *Multilayer Feedforward Networks with a Nonpolynomial Activation Function Can Approximate Any Function*. Neural Networks 6.6 (1993), DOI: 10.1016/S0893-6080(05)80131-5.

[150] User:Dhp1080. *Wikipedia: English: Neuron Description.* 2019.

[151] C. M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. New York: Springer, 2006.

[152] W. Maass, G. Schnitger, and E. D. Sontag. *A Comparison of the Computational Power of Sigmoid and Boolean Threshold Circuits*. *Theoretical Advances in Neural Computation and Learning*. Ed. by V. Roychowdhury, K.-Y. Siu, and A. Orlitsky. Boston, MA: Springer US, 1994, pp. 127–151. DOI: 10.1007/978-1-4615-2696-4_4.

[153] W. Maass. *Bounds for the Computational Power and Learning Complexity of Analog Neural Nets*. SIAM J. Comput. 26.3 (1997), DOI: 10.1137/S0097539793256041.

[154] R. Pascanu, G. Montufar, and Y. Bengio. *On the Number of Response Regions of Deep Feed Forward Networks with Piece-Wise Linear Activations* (2014). arXiv: 1312.6098.

[155] G. F. Montufar et al. *On the Number of Linear Regions of Deep Neural Networks*. Advances in Neural Information Processing Systems 27 (2014). Ed. by Z. Ghahramani et al., arXiv: 1402.1869.

[156] Y. A. LeCun et al. *Efficient BackProp*. *Neural Networks: Tricks of the Trade: Second Edition*. Ed. by G. Montavon, G. B. Orr, and K.-R. Müller. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2012, pp. 9–48. DOI: 10.1007/978-3-642-35289-8_3.

[157] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Learning Representations by Back-Propagating Errors*. Nature 323.6088 (1986), DOI: 10.1038/323533a0.

[158] R. Pascanu et al. *On the Saddle Point Problem for Non-Convex Optimization* (2014). arXiv: 1405.4604 [cs].

[159] Y. Dauphin et al. *Identifying and Attacking the Saddle Point Problem in High-Dimensional Non-Convex Optimization*. *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al. 2014, pp. 2933–2941. arXiv: 1406.2572.

[160] G. Goh. *Why Momentum Really Works*. Distill 2.4 (2017), DOI: 10.23915/distill.00006.

[161] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization* (2017). arXiv: 1412.6980.

[162] S. Ioffe and C. Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift* (2015), arXiv: 1502.03167.

[163] D. Wolpert and W. Macready. *No Free Lunch Theorems for Optimization*. IEEE Transactions on Evolutionary Computation 1.1 (1997), DOI: 10.1109/4235.585893.

[164] J. Bergstra and Y. Bengio. *Random Search for Hyper-Parameter Optimization*. Journal of Machine Learning Research 13.10 (2012),

[165] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics. New York, NY: Springer, 2009.

[166] G. C. Cawley and N. L. C. Talbot. *On Over-Fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation*. Journal of Machine Learning Research 11.70 (2010),

[167] Y. Bengio. *Practical Recommendations for Gradient-Based Training of Deep Architectures*. *Neural Networks: Tricks of the Trade: Second Edition*. Ed. by G. Montavon, G. B. Orr, and K.-R. Müller. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2012, pp. 437–478. DOI: 10.1007/978-3-642-35289-8_26.

[168] C. Zhang et al. *Understanding Deep Learning Requires Rethinking Generalization* (2017). arXiv: 1611.03530.

[169] ATLAS Collaboration. *A Neural Network Clustering Algorithm for the ATLAS Silicon Pixel Detector*. J. Inst. 9.09 (2014), DOI: 10.1088/1748-0221/9/09/P09009.

[170] T. Sjöstrand, S. Mrenna, and P. Skands. *A Brief Introduction to PYTHIA 8.1* (2007). DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[171] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV datas*. ATL-PHYS-PUB-2014-021. 2014. URL: http://cdsweb.cern.ch/record/1966419/.

[172] J. Allison et al. *Recent Developments in Geant4*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 835 (2016), DOI: 10.1016/j.nima.2016.06.125.

[173] ATLAS Collaboration. *The ATLAS Simulation Infrastructure* (2010). DOI: 10.1140/epjc/s10052-010-1429-9. arXiv: 1005.4568.

[174] ATLAS Collaboration. *Robustness of the Artificial Neural Networks Used for Clustering in the ATLAS Pixel Detector*. ATL-PHYS-PUB-2015-052. 2015. URL: http://cdsweb.cern.ch/record/2116350/.

[175] T. Fawcett. *An Introduction to ROC Analysis*. Pattern Recognition Letters. ROC Analysis in Pattern Recognition 27.8 (2006), DOI: 10.1016/j.patrec.2005.10.010.

[176] D. J. Hand and R. J. Till. *A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems*. Machine Learning 45.2 (2001), DOI: 10.1023/A:1010920819831.

[177] ATLAS Collaboration. *Training and validation of the ATLAS pixel clustering neural networks*. ATL-PHYS-PUB-2018-002. 2018. URL: https://cds.cern.ch/record/2309474.

[178] ATLAS Collaboration. *Modelling of Track Reconstruction Inside Jets with the 2016 ATLAS $\sqrt{s}$ = 13 TeV pp dataset*. ATL-PHYS-PUB-2017-016. 2017. URL: http://cdsweb.cern.ch/record/2275639.

[179] C. M. Bishop. *Mixture density networks*. 1994. URL: http://publications.aston.ac.uk/id/eprint/373/.

[180] E. Khoda. *ATLAS pixel cluster splitting using Mixture Density Networks*. ATL-PHYS-PROC-2019-082. 2019. URL: https://cds.cern.ch/record/2687968.

[181] LHC SUSY Cross Section Working Group. *SUSY Cross Sections for 13, 14, 33 and 100 TeV pp Collisions*. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections.

[182] G. Martínez Solaeche. *Top tagging at the LHC experiments with proton-proton collisions at 13 TeV*. PhD thesis. 2015. DOI: 10.13140/RG.2.1.4521.4560.

[183] Particle Data Group. *Monte Carlo Event Generators*. *Review of Particle Physics*. Vol. 98. 3. 2018. DOI: 10.1103/PhysRevD.98.030001.

[184] J. Alwall et al. *The Automated Computation of Tree-Level and next-to-Leading Order Differential Cross Sections, and Their Matching to Parton Shower Simulations*. JHEP 2014.7 (2014), DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301.

[185] R. D. Ball et al. *Parton Distributions with LHC Data*. Nuclear Physics B 867.2 (2013), DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303.

[186] T. Sjostrand, S. Mrenna, and P. Skands. *PYTHIA 6.4 Physics and Manual*. J. High Energy Phys. 2006.05 (2006), DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175.

[187] T. Sjöstrand et al. *An Introduction to PYTHIA 8.2*. Computer Physics Communications 191 (2015), DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012.

[188] L. Lonnblad. *Correcting the Colour-Dipole Cascade Model with Fixed Order Matrix Elements*. J. High Energy Phys. 2002.05 (2002), DOI: 10.1088/1126-6708/2002/05/046. arXiv: hep-ph/0112284.

[189] S. Frixione, P. Nason, and C. Oleari. *Matching NLO QCD Computations with Parton Shower Simulations: The POWHEG Method*. J. High Energy Phys. 2007.11 (2007), DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092.

[190] S. Alioli et al. *A General Framework for Implementing NLO Calculations in Shower Monte Carlo Programs: The POWHEG BOX*. JHEP 2010.6 (2010), DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581.

[191] D. J. Lange. *The EvtGen Particle Decay Simulation Package*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 462.1-2 (2001), DOI: 10.1016/S0168-9002(01)00089-4.

[192] W. Beenakker et al. *NNLL-fast: predictions for coloured supersymmetric particle production at the LHC with threshold and Coulomb resummation*. JHEP 12 (2016), DOI: 10.1007/JHEP12(2016)133. arXiv: 1607.07741 [hep-ph].

[193] W. Beenakker et al. *NNLL resummation for squark and gluino production at the LHC*. JHEP 12 (2014), DOI: 10.1007/JHEP12(2014)023. arXiv: 1404.3134 [hep-ph].

[194] W. Beenakker et al. *Towards NNLL resummation: hard matching coefficients for squark and gluino hadroproduction*. JHEP 10 (2013), DOI: 10.1007/JHEP10(2013)120. arXiv: 1304.6354 [hep-ph].

[195] W. Beenakker et al. *NNLL resummation for squark-antisquark pair production at the LHC*. JHEP 01 (2012), DOI: 10.1007/JHEP01(2012)076. arXiv: 1110.2446 [hep-ph].

[196] W. Beenakker et al. *Soft-gluon resummation for squark and gluino hadroproduction*. JHEP 12 (2009), DOI: 10.1088/1126-6708/2009/12/041. arXiv: 0909.4418 [hep-ph].

[197] A. Kulesza and L. Motyka. *Soft gluon resummation for the production of gluino-gluino and squark-antisquark pairs at the LHC*. Phys. Rev. D 80 (2009), DOI: 10.1103/PhysRevD.80.095004. arXiv: 0905.4749 [hep-ph].

[198] A. Kulesza and L. Motyka. *Threshold resummation for squark-antisquark and gluino-pair production at the LHC*. Phys. Rev. Lett. 102 (2009), DOI: 10.1103/PhysRevLett.102.111802. arXiv: 0807.2405 [hep-ph].

[199] W. Beenakker et al. *Squark and gluino production at hadron colliders*. Nucl. Phys. B 492 (1997), DOI: 10.1016/S0550-3213(97)00084-9. arXiv: hep-ph/9610490.

[200] J. Butterworth et al. *PDF4LHC Recommendations for LHC Run II*. J. Phys. G: Nucl. Part. Phys. 43.2 (2016), DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865.

[201] M. Czakon and A. Mitov. *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*. Computer Physics Communications 185.11 (2014), DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675.

[202] P. Kant et al. *HATHOR for Single Top-Quark Production: Updated Predictions and Uncertainty Estimates for Single Top-Quark Production in Hadronic Collisions*. Computer Physics Communications 191 (2015), DOI: 10.1016/j.cpc.2015.02.001. arXiv: 1406.4403.

[203] N. Kidonakis. *NNLL Resummation for S-Channel Single Top Quark Production*. Phys. Rev. D 81.5 (2010), DOI: 10.1103/PhysRevD.81.054028. arXiv: 1001.5034.

[204] The LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties* (2013). DOI: 10.5170/CERN-2013-004. arXiv: 1307.1347 [hep-ex, physics:hep-ph].

[205] T. Gleisberg et al. *Event Generation with SHERPA 1.1*. J. High Energy Phys. 2009.02 (2009), DOI: 10.1088/1126-6708/2009/02/007. arXiv: 0811.4622.

[206] S. Hoeche et al. *QCD Matrix Elements + Parton Showers: The NLO Case*. JHEP 2013.4 (2013), DOI: 10.1007/JHEP04(2013)027. arXiv: 1207.5030.

[207] S. Catani et al. *Vector Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at NNLO*. Phys. Rev. Lett. 103.8 (2009), DOI: 10.1103/PhysRevLett.103.082001. arXiv: 0903.2120.

[208] ATLAS Collaboration. *Search for Pair Production of Up-Type Vector-like Quarks and for Four-Top-Quark Events in Final States with Multiple b-Jets with the ATLAS Detector*. JHEP 2018.7 (2018), DOI: 10.1007/JHEP07(2018)089. arXiv: 1803.09678.

[209] ATLAS Collaboration. *ATLAS data quality operations and performance for 2015–2018 data-taking* (2019). arXiv: 1911.04632 [hep-ex].

[210] ATLAS Collaboration. *Topological Cell Clustering in the ATLAS Calorimeters and Its Performance in LHC Run 1*. Eur. Phys. J. C 77.7 (2017), DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934.

[211] M. Cacciari, G. P. Salam, and G. Soyez. *The Anti-K_t Jet Clustering Algorithm*. J. High Energy Phys. 2008.04 (2008), DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189.

[212] ATLAS Collaboration. *Properties of jets and inputs to jet reconstruction and calibration with the ATLAS detector using proton–proton collisions at* $\sqrt{s} = 13$ *TeV*. ATL-PHYS-PUB-2015-036. 2015. URL: https://cds.cern.ch/record/2044564.

[213] ATLAS Collaboration. *Jet Energy Scale Measurements and Their Systematic Uncertainties in Proton-Proton Collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS Detector*. Phys. Rev. D 96.7 (2017), DOI: 10.1103/PhysRevD.96.072002. arXiv: 1703.09665.

[214] 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 79.11 (2019), DOI: 10.1140/epjc/s10052-019-7450-8. arXiv: 1907.05120.

[215] B. Nachman et al. *Jets from Jets: Re-Clustering as a Tool for Large Radius Jet Reconstruction and Grooming at the LHC*. JHEP 2015.2 (2015), DOI: 10.1007/JHEP02(2015)075. arXiv: 1407.2922.

[216] ATLAS Collaboration. *Jet reclustering and close-by effects in ATLAS Run 2*. ATLAS-CONF-2017-062. 2017. URL: https://cds.cern.ch/record/2275649.

[217] ATLAS Collaboration. *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*. JINST 14.12 (2019), DOI: 10.1088/1748-0221/14/12/P12006. arXiv: 1908.00005 [hep-ex].

[218] ATLAS Collaboration. *Muon Reconstruction Performance of the ATLAS Detector in Proton–Proton Collision Data at $\sqrt{s}$=13 TeV*. Eur. Phys. J. C 76.5 (2016), DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598.

[219] 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 78.11 (2018), DOI: 10.1140/epjc/s10052-018-6288-9. arXiv: 1802.08168.

[220] ATLAS Collaboration. *E_T^{miss} performance in the ATLAS detector using 2015–2016 LHC pp collisions*. ATLAS-CONF-2018-023. 2018. URL: https://cds.cern.ch/record/2625233.

[221] B. Denby et al. *Neural Networks for Triggering*. IEEE Transactions on Nuclear Science 37.2 (1990), DOI: 10.1109/23.106627.

[222] A. Chilingarian. *Neural Classification Technique for Background Rejection in High Energy Physics Experiments*. Neurocomputing 6.5 (1994), DOI: 10.1016/0925-2312(94)90002-7.

[223] B. Denby. *Neural Networks in High Energy Physics: A Ten Year Perspective*. Computer Physics Communications 119.2 (1999), DOI: 10.1016/S0010-4655(98)00199-4.

[224] S. Whiteson and D. Whiteson. *Machine Learning for Event Selection in High Energy Physics*. Engineering Applications of Artificial Intelligence 22.8 (2009), DOI: 10.1016/j.engappai.2009.05.004.

[225] P. Baldi, P. Sadowski, and D. Whiteson. *Enhanced Higgs to $\tau^+\tau^-$ Searches with Deep Learning* (2014). DOI: 10.1103/PhysRevLett.114.111801. arXiv: 1410.3469 [hep-ph].

[226] P. Baldi, P. Sadowski, and D. Whiteson. *Searching for Exotic Particles in High-Energy Physics with Deep Learning*. Nat Commun 5.1 (2014), DOI: 10.1038/ncomms5308. arXiv: 1402.4735.

[227] P. Baldi et al. *Parameterized Machine Learning for High-Energy Physics*. Eur. Phys. J. C 76.5 (2016), DOI: 10.1140/epjc/s10052-016-4099-4. arXiv: 1601.07913.

[228] J. Bergstra and Y. Bengio. *Random Search for Hyper-Parameter Optimization*. Journal of Machine Learning Research 13.10 (2012), URL: http://jmlr.org/papers/v13/bergstra12a.html.

[229] E. Gross and O. Vitells. *Trial Factors for the Look Elsewhere Effect in High Energy Physics*. Eur. Phys. J. C 70.1 (2010), DOI: 10.1140/epjc/s10052-010-1470-8.

[230] M. Baak et al. *HistFitter Software Framework for Statistical Data Analysis*. Eur. Phys. J. C 75.4 (2015), DOI: 10.1140/epjc/s10052-015-3327-7.

[231] Particle Data Group. *Monte Carlo Event Generators*. *Review of Particle Physics*. Vol. 98. 3. 2018. DOI: 10.1103/PhysRevD.98.030001.

[232] J. Neyman and E. S. Pearson. *IX. On the Problem of the Most Efficient Tests of Statistical Hypotheses*. Phil. Trans. R. Soc. Lond. A 231.694-706 (1933), DOI: 10.1098/rsta.1933.0009.

[233] G. Cowan et al. *Asymptotic Formulae for Likelihood-Based Tests of New Physics*. Eur. Phys. J. C 71.2 (2011), DOI: 10.1140/epjc/s10052-011-1554-0.

[234] A. L. Read. *Presentation of Search Results: The CLs Technique*. J. Phys. G: Nucl. Part. Phys. 28.10 (2002), DOI: 10.1088/0954-3899/28/10/313.

Appendices

A Kinematic reweighting in the 1-lepton channel

Note: these plots, and all the work that went into deriving these weights, are due to Egor Antipov. Many thanks!

Table 7 – Definitions of the control regions used to derive the kinematic reweighting scale-factors. The N_b requirements ensure these control regions are orthogonal to all signal regions of the analysis, which include a $N_b \geq 3$ requirement. The Z-enriched region uses a definition of $E_{\mathrm{T}}^{\mathrm{miss}}$ that includes the lepton pair momentum, to simulate $Z \to \nu\nu$ events.

Criteria common to all regions: $N_{jet} \geq 4$, $E_{\mathrm{T}}^{\mathrm{miss}} \geq 200$ GeV				
Control-region	N_{lepton}	N_b	mtb (GeV)	M_Z (GeV)
$t\bar{t}$-enriched	$= 1$	$= 2$	≤ 350	
single-top-enriched	$= 1$	$= 2$	> 350	
W-enriched	$= 1$	$= 0$		
Z-enriched	$= 2$, opposite charge	$= 0$		$\in$ (60,120(

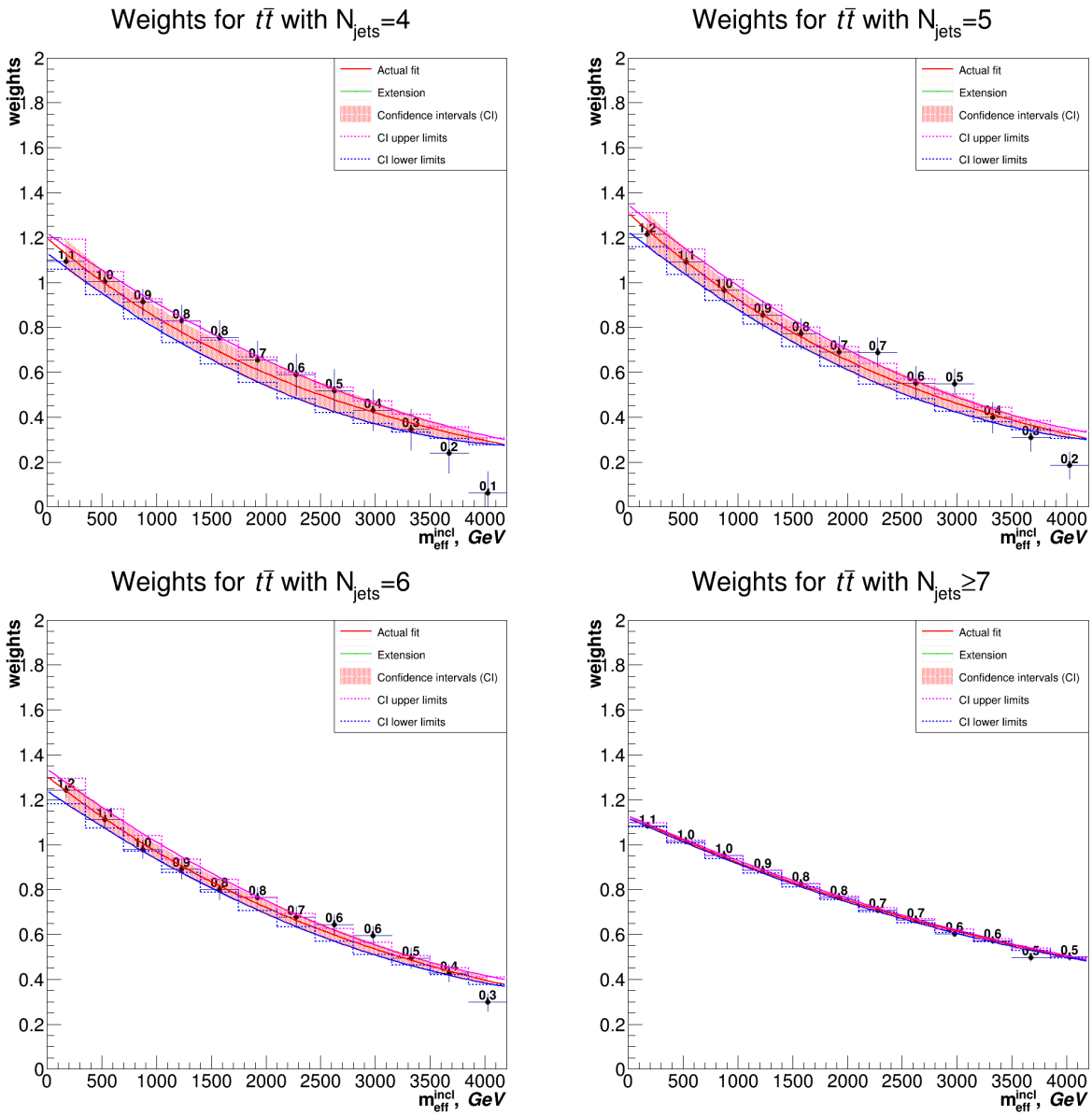


Figure 28 – Result of the fit to the data/MC distribution for the $t\bar{t}$–enriched control region in four different bins of N_{jet}.

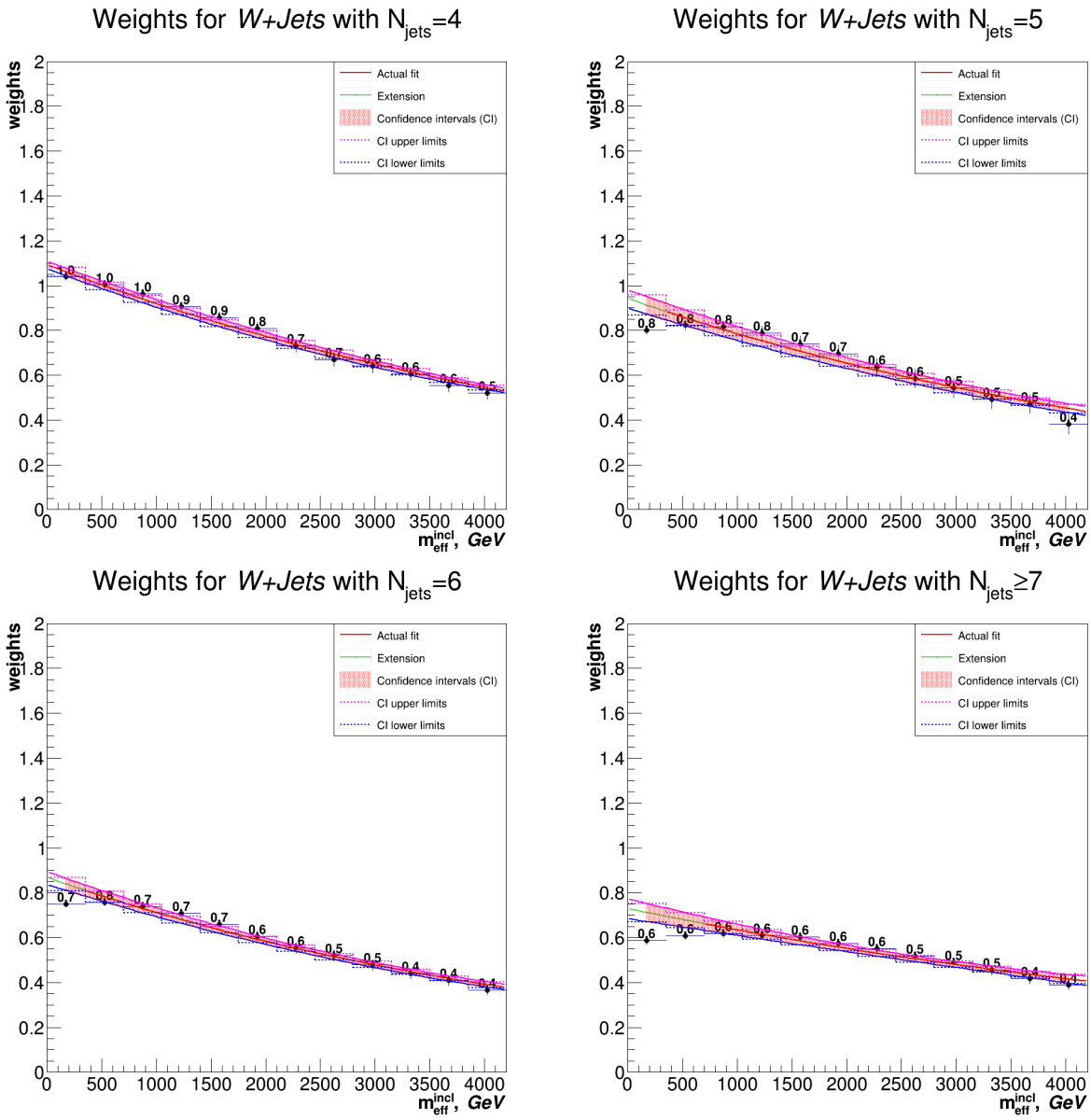


Figure 29 – Result of the fit to the data/MC distribution for the W–enriched control region in four different bins of N_{jet}.

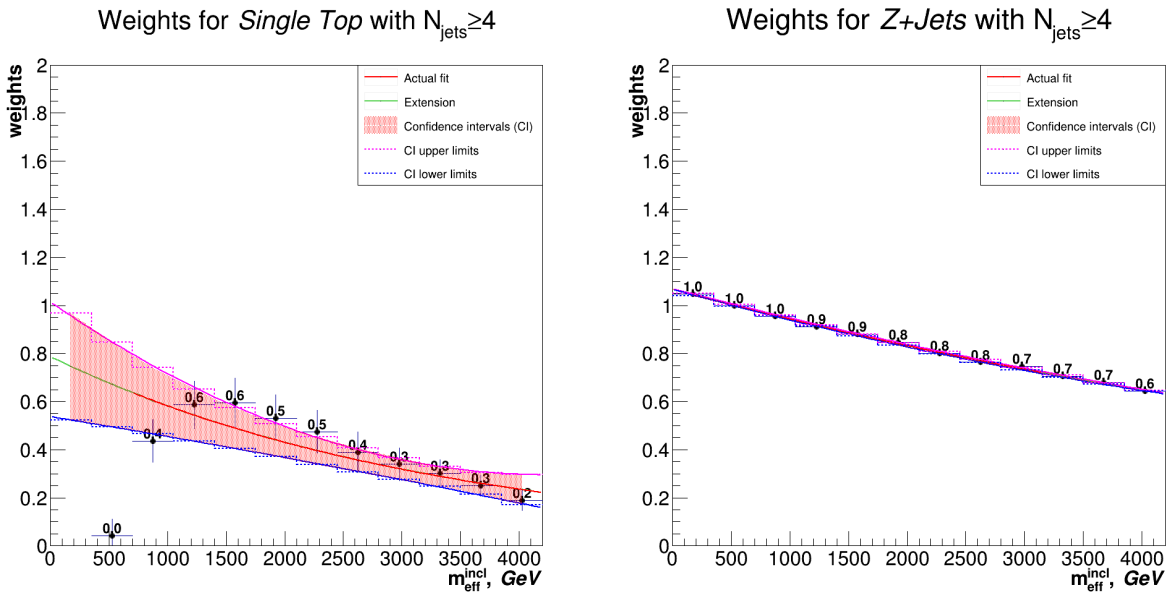


Figure 30 – Result of the fit to the data/MC distribution for the (a) single-top– and (b) Z–enriched control regions.

B Yield tables

	CR_Gbb_2800_1400	VR1	VR2	VR3	SR
Observed events	27	15	5	6	1
Fitted bkg events	26.98 ± 5.35	12.31 ± 4.26	7.01 ± 4.56	10.40 ± 4.21	0.76 ± 0.70
Fitted ttbar events	13.94 ± 8.59	3.71 ± 3.24	$1.88^{+2.03}_{-1.88}$	1.81 ± 1.54	0.14 ± 0.14
Fitted singletop events	3.13 ± 1.58	2.35 ± 1.41	1.80 ± 1.12	2.35 ± 1.35	0.21 ± 0.16
Fitted topEW events	1.21 ± 0.69	0.54 ± 0.47	$0.44^{+0.85}_{-0.44}$	0.44 ± 0.33	$0.00^{+0.05}_{-0.00}$
Fitted W_jets events	$1.24^{+1.24}_{-1.24}$	1.24 ± 1.04	0.67 ± 0.57	1.06 ± 0.97	0.00 ± 0.00
Fitted Z_jets events	$2.32^{+3.31}_{-2.32}$	$0.93^{+1.62}_{-0.93}$	$1.23^{+2.76}_{-1.23}$	$2.07^{+3.31}_{-2.07}$	$0.33^{+0.63}_{-0.33}$
Fitted diboson events	0.11 ± 0.06	$0.19^{+0.50}_{-0.19}$	0.00 ± 0.00	0.33 ± 0.32	0.00 ± 0.00
Fitted QCD events	5.04 ± 4.92	3.35 ± 3.30	0.98 ± 0.97	2.34 ± 2.31	$0.08^{+0.09}_{-0.08}$
MC exp. SM events	26.01	11.97	6.85	10.21	0.75
MC exp. ttbar events	13.14	3.49	1.77	1.70	0.13
MC exp. singletop events	3.13	2.35	1.80	2.35	0.21
MC exp. topEW events	1.21	0.54	0.44	0.44	0.00
MC exp. W_jets events	1.24	1.24	0.67	1.06	0.00
MC exp. Z_jets events	2.32	0.93	1.22	2.07	0.32
MC exp. diboson events	0.11	0.19	0.00	0.33	0.00
data-driven exp. QCD events	4.86	3.23	0.95	2.26	0.08

Table 8 – Background-only fit results for region Gbb_2800_1400

	CR_Gbb_2300_1000	VR1	VR2	VR3	SR
Observed events	29	17	16	5	1
Fitted bkg events	29.00 ± 5.86	15.57 ± 5.32	12.22 ± 5.91	8.26 ± 5.43	$0.75^{+0.79}_{-0.75}$
Fitted ttbar events	13.02 ± 10.75	$3.90^{+4.07}_{-3.90}$	$3.36^{+3.47}_{-3.36}$	$0.34^{+0.79}_{-0.34}$	$0.09^{+0.10}_{-0.09}$
Fitted singletop events	3.41 ± 1.12	3.19 ± 2.09	3.09 ± 1.73	1.33 ± 0.90	0.25 ± 0.21
Fitted topEW events	1.37 ± 0.93	0.74 ± 0.48	$0.71^{+1.19}_{-0.71}$	0.14 ± 0.11	$0.00^{+0.04}_{-0.00}$
Fitted W_jets events	$1.81^{+2.18}_{-1.81}$	2.07 ± 1.70	$1.00^{+1.07}_{-1.00}$	$3.65^{+4.81}_{-3.65}$	0.00 ± 0.00
Fitted Z_jets events	$2.76^{+4.72}_{-2.76}$	$1.52^{+2.30}_{-1.52}$	$2.35^{+3.56}_{-2.35}$	$1.96^{+2.71}_{-1.96}$	$0.31^{+0.76}_{-0.31}$
Fitted diboson events	0.00 ± 0.00	$0.19^{+0.50}_{-0.19}$	0.12 ± 0.12	0.23 ± 0.14	0.00 ± 0.00
Fitted QCD events	6.62 ± 6.58	3.96 ± 3.96	1.59 ± 1.58	$0.61^{+0.64}_{-0.61}$	$0.10^{+0.11}_{-0.10}$
MC exp. SM events	32.12	16.50	13.01	8.35	0.77
MC exp. ttbar events	16.14	4.84	4.16	0.42	0.12
MC exp. singletop events	3.41	3.19	3.09	1.33	0.25
MC exp. topEW events	1.37	0.74	0.71	0.14	0.00
MC exp. W_jets events	1.82	2.06	1.00	3.66	0.00
MC exp. Z_jets events	2.77	1.52	2.35	1.96	0.30
MC exp. diboson events	0.00	0.19	0.12	0.23	0.00
data-driven exp. QCD events	6.62	3.96	1.59	0.61	0.10

Table 9 – Background-only fit results for region Gbb_2300_1000

	CR_Gbb_2100_1600	VR1	VR2	VR3	SR
Observed events	48	636	20	13	0
Fitted bkg events	48.07 ± 7.29	810.65 ± 329.95	33.91 ± 16.10	17.07 ± 11.18	1.91 ± 1.76
Fitted ttbar events	36.19 ± 8.95	564.07 ± 331.10	24.89 ± 16.13	7.91 ± 6.88	0.40 ± 0.31
Fitted singletop events	2.77 ± 1.58	67.89 ± 34.87	2.57 ± 1.49	1.41 ± 1.02	0.37 ± 0.26
Fitted topEW events	$0.76^{+0.91}_{-0.76}$	23.25 ± 11.86	1.52 ± 0.95	$0.10^{+0.17}_{-0.10}$	0.09 ± 0.07
Fitted W_jets events	$1.38^{+2.17}_{-1.38}$	34.31 ± 27.01	2.36 ± 2.20	1.66 ± 1.57	$0.00^{+0.17}_{-0.00}$
Fitted Z_jets events	$3.84^{+7.43}_{-3.84}$	$67.83^{+89.72}_{-67.83}$	$2.26^{+4.05}_{-2.26}$	$4.96^{+6.40}_{-4.96}$	$0.77^{+1.68}_{-0.77}$
Fitted diboson events	$0.37^{+0.75}_{-0.37}$	8.04 ± 5.63	$0.22^{+0.54}_{-0.22}$	0.95 ± 0.92	0.15 ± 0.09
Fitted QCD events	2.77 ± 2.75	45.27 ± 44.75	0.09 ± 0.09	0.08 ± 0.08	$0.13^{+0.14}_{-0.13}$
MC exp. SM events	33.20	578.01	23.65	13.79	1.75
MC exp. ttbar events	21.34	332.32	14.67	4.65	0.23
MC exp. singletop events	2.78	67.84	2.57	1.41	0.37
MC exp. topEW events	0.76	23.24	1.52	0.10	0.09
MC exp. W_jets events	1.38	34.24	2.34	1.66	0.00
MC exp. Z_jets events	3.83	67.42	2.24	4.92	0.77
MC exp. diboson events	0.36	8.05	0.22	0.96	0.15
data-driven exp. QCD events	2.75	44.90	0.09	0.08	0.13

Table 10 – Background-only fit results for region Gbb_2100_1600

	CR_Gbb_2000_1800	VR1	VR2	VR3	SR
Observed events	29	1492	57	12	1
Fitted bkg events	29.04 ± 5.44	2123.48 ± 1078.83	119.74 ± 61.78	12.90 ± 6.50	1.75 ± 1.30
Fitted ttbar events	25.09 ± 5.78	1836.42 ± 1079.22	104.93 ± 61.48	5.58 ± 3.80	$0.72^{+0.82}_{-0.72}$
Fitted singletop events	1.26 ± 0.91	80.03 ± 41.21	4.38 ± 3.48	$1.12^{+1.62}_{-1.12}$	0.03 ± 0.02
Fitted topEW events	$0.30^{+0.49}_{-0.30}$	34.28 ± 17.70	2.00 ± 1.42	0.11 ± 0.09	0.05 ± 0.04
Fitted W_jets events	$0.91^{+1.60}_{-0.91}$	44.46 ± 19.88	$2.48^{+3.38}_{-2.48}$	$0.33^{+0.51}_{-0.33}$	$0.25^{+0.53}_{-0.25}$
Fitted Z_jets events	$0.62^{+1.82}_{-0.62}$	101.72 ± 62.78	4.86 ± 4.66	5.76 ± 5.17	0.56 ± 0.28
Fitted diboson events	0.00 ± 0.00	11.14 ± 5.82	$0.49^{+2.86}_{-0.49}$	0.00 ± 0.00	0.14 ± 0.08
Fitted QCD events	0.86 ± 0.86	15.43 ± 15.33	0.60 ± 0.59	0.00 ± 0.04	0.00 ± 0.04
MC exp. SM events	23.93	1746.96	98.19	11.77	1.60
MC exp. ttbar events	19.98	1460.34	83.44	4.42	0.57
MC exp. singletop events	1.27	80.04	4.38	1.13	0.03
MC exp. topEW events	0.30	34.29	2.00	0.11	0.05
MC exp. W_jets events	0.90	44.20	2.47	0.33	0.24
MC exp. Z_jets events	0.61	101.46	4.82	5.79	0.56
MC exp. diboson events	0.00	11.12	0.48	0.00	0.14
data-driven exp. QCD events	0.87	15.50	0.60	0.00	0.00

Table 11 – Background-only fit results for region Gbb_2000_1800

	CR_Gtt_2100_1	VR1	VR2	SR
Observed events	35	73	53	0
Fitted bkg events	35.00 ± 5.92	77.61 ± 24.41	41.62 ± 14.45	0.54 ± 0.23
Fitted ttbar events	25.58 ± 6.60	39.82 ± 23.15	20.18 ± 11.80	0.20 ± 0.12
Fitted singletop events	4.94 ± 1.34	14.19 ± 7.56	9.06 ± 4.73	0.15 ± 0.09
Fitted topEW events	1.80 ± 1.00	9.42 ± 4.86	6.52 ± 3.56	0.15 ± 0.11
Fitted W_jets events	$1.41^{+1.94}_{-1.41}$	5.09 ± 4.20	2.25 ± 2.18	0.00 ± 0.00
Fitted Z_jets events	$0.37^{+1.27}_{-0.37}$	$4.37^{+5.32}_{-4.37}$	$2.62^{+4.08}_{-2.62}$	$0.04^{+0.07}_{-0.04}$
Fitted diboson events	$0.34^{+0.44}_{-0.34}$	2.58 ± 1.52	$0.19^{+0.70}_{-0.19}$	0.00 ± 0.00
Fitted QCD events	0.55 ± 0.55	2.13 ± 2.12	0.80 ± 0.79	$0.00^{+0.04}_{-0.00}$
MC exp. SM events	34.10	76.21	40.91	0.53
MC exp. ttbar events	24.67	38.42	19.47	0.19
MC exp. singletop events	4.94	14.19	9.06	0.15
MC exp. topEW events	1.80	9.42	6.52	0.15
MC exp. W_jets events	1.41	5.10	2.25	0.00
MC exp. Z_jets events	0.37	4.37	2.62	0.05
MC exp. diboson events	0.34	2.58	0.19	0.00
data-driven exp. QCD events	0.55	2.14	0.80	0.00

Table 12 – Background-only fit results for region Gtt_2100_1

	CR_Gtt_1800_1	VR1	VR2	SR
Observed events	31	130	39	0
Fitted bkg events	31.00 ± 5.57	107.44 ± 34.76	28.24 ± 9.35	1.07 ± 0.41
Fitted ttbar events	22.36 ± 6.60	54.90 ± 31.51	12.68 ± 7.43	0.32 ± 0.20
Fitted singletop events	4.62 ± 0.80	19.15 ± 10.21	6.50 ± 3.54	0.19 ± 0.12
Fitted topEW events	2.19 ± 1.52	14.60 ± 7.59	4.23 ± 2.39	0.30 ± 0.21
Fitted W_jets events	$0.40^{+0.77}_{-0.40}$	6.16 ± 5.13	1.95 ± 1.73	0.21 ± 0.18
Fitted Z_jets events	$0.37^{+1.37}_{-0.37}$	$6.27^{+8.39}_{-6.27}$	$1.44^{+2.73}_{-1.44}$	$0.05^{+0.07}_{-0.05}$
Fitted diboson events	$0.15^{+0.53}_{-0.15}$	4.09 ± 2.15	$0.68^{+0.80}_{-0.68}$	0.00 ± 0.00
Fitted QCD events	0.92 ± 0.91	2.27 ± 2.24	0.75 ± 0.74	$0.00^{+0.04}_{-0.00}$
MC exp. SM events	33.40	113.28	29.57	1.10
MC exp. ttbar events	24.74	60.79	14.04	0.36
MC exp. singletop events	4.62	19.15	6.49	0.19
MC exp. topEW events	2.19	14.59	4.23	0.30
MC exp. W_jets events	0.40	6.15	1.95	0.21
MC exp. Z_jets events	0.38	6.26	1.44	0.05
MC exp. diboson events	0.15	4.09	0.68	0.00
data-driven exp. QCD events	0.91	2.26	0.75	0.00

Table 13 – Background-only fit results for region Gtt_1800_1

	CR_Gtt_2300_1200	VR1	VR2	SR
Observed events	46	156	30	1
Fitted bkg events	46.04 ± 6.79	182.49 ± 68.56	30.57 ± 13.93	0.74 ± 0.36
Fitted ttbar events	37.18 ± 7.19	121.57 ± 66.22	19.31 ± 11.79	0.38 ± 0.24
Fitted singletop events	4.59 ± 1.21	22.79 ± 11.95	4.77 ± 2.76	0.06 ± 0.04
Fitted topEW events	2.26 ± 1.76	19.21 ± 9.91	2.79 ± 2.11	0.24 ± 0.19
Fitted W_jets events	0.62 ± 0.57	6.41 ± 5.20	$1.61^{+2.17}_{-1.61}$	0.00 ± 0.00
Fitted Z_jets events	$0.32^{+0.86}_{-0.32}$	$6.48^{+10.12}_{-6.48}$	$1.48^{+2.01}_{-1.48}$	$0.06^{+0.11}_{-0.06}$
Fitted diboson events	0.00 ± 0.00	3.77 ± 2.17	$0.19^{+0.43}_{-0.19}$	0.00 ± 0.00
Fitted QCD events	1.07 ± 1.06	2.25 ± 2.23	0.43 ± 0.42	$0.00^{+0.04}_{-0.00}$
MC exp. SM events	32.88	139.47	23.75	0.61
MC exp. ttbar events	24.02	78.55	12.48	0.25
MC exp. singletop events	4.59	22.79	4.77	0.06
MC exp. topEW events	2.26	19.21	2.79	0.24
MC exp. W_jets events	0.61	6.36	1.61	0.00
MC exp. Z_jets events	0.32	6.55	1.49	0.06
MC exp. diboson events	0.00	3.77	0.19	0.00
data-driven exp. QCD events	1.07	2.25	0.43	0.00

Table 14 – Background-only fit results for region Gtt_2300_1200

	CR_Gtt_1900_1400	VR1	VR2	SR
Observed events	26	482	49	2
Fitted bkg events	25.99 ± 5.11	427.93 ± 204.03	40.55 ± 20.27	0.84 ± 0.58
Fitted ttbar events	22.00 ± 5.45	340.51 ± 201.31	33.92 ± 19.75	0.68 ± 0.55
Fitted singletop events	1.31 ± 0.95	23.06 ± 12.20	2.22 ± 1.82	$0.01^{+0.02}_{-0.01}$
Fitted topEW events	1.67 ± 1.36	35.87 ± 18.64	2.90 ± 1.64	0.15 ± 0.15
Fitted W_jets events	$0.09^{+0.83}_{-0.09}$	12.70 ± 8.38	$0.25^{+1.48}_{-0.25}$	0.00 ± 0.00
Fitted Z_jets events	$0.28^{+0.75}_{-0.28}$	$11.53^{+12.68}_{-11.53}$	$1.16^{+1.99}_{-1.16}$	0.00 ± 0.00
Fitted diboson events	0.00 ± 0.00	3.43 ± 1.80	$0.10^{+0.28}_{-0.10}$	0.00 ± 0.00
Fitted QCD events	0.63 ± 0.62	0.84 ± 0.83	$0.00^{+0.05}_{-0.00}$	$0.00^{+0.05}_{-0.00}$
MC exp. SM events	30.72	500.77	47.88	0.99
MC exp. ttbar events	26.74	413.34	41.23	0.83
MC exp. singletop events	1.32	23.05	2.23	0.01
MC exp. topEW events	1.67	35.88	2.90	0.15
MC exp. W_jets events	0.10	12.66	0.27	0.00
MC exp. Z_jets events	0.28	11.59	1.16	0.00
MC exp. diboson events	0.00	3.42	0.09	0.00
data-driven exp. QCD events	0.62	0.82	0.00	0.00

Table 15 – Background-only fit results for region Gtt_1900_1400

C Systematics tables

	CR_Gbb_2800_1400	VR1	VR2	VR3	SR
Total background expectation	26.98	12.31	7.01	10.40	0.76
Total statistical ($\sqrt{N_{\text{exp}}}$)	±5.19	±3.51	±2.65	±3.22	±0.87
Total background systematic	±5.35 [19.85%]	±4.26 [34.63%]	±4.56 [65.10%]	±4.21 [40.48%]	±0.70 [91.20%]
mu_ttbar_Gbb_2800_1400	±8.86 [32.8%]	±2.36 [19.2%]	±1.20 [17.1%]	±1.15 [11.0%]	±0.09 [11.9%]
alpha_QCDHundred	±4.90 [18.1%]	±3.26 [26.5%]	±0.95 [13.6%]	±2.28 [21.9%]	±0.08 [10.4%]
alpha_WZ_ckkw_syst	±2.65 [9.8%]	±1.30 [10.6%]	±1.81 [25.9%]	±2.79 [26.8%]	±0.54 [71.2%]
gamma_stat_CR_Gbb_2800_1400_cuts_bin_0	±2.52 [9.3%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JER1	±1.52 [5.6%]	±0.66 [5.3%]	±2.23 [31.8%]	±0.66 [6.3%]	±0.00 [0.26%]
alpha_JER0	±1.21 [4.5%]	±0.23 [1.9%]	±1.34 [19.1%]	±0.16 [1.5%]	±0.01 [1.4%]
alpha_JER2	±1.11 [4.1%]	±0.24 [2.0%]	±1.20 [17.2%]	±0.58 [5.6%]	±0.05 [6.3%]
alpha_JES0	±1.06 [3.9%]	±0.05 [0.37%]	±0.05 [0.73%]	±0.35 [3.4%]	±0.04 [5.1%]
alpha_WZ_renorm_syst	±0.98 [3.6%]	±0.90 [7.3%]	±0.55 [7.9%]	±0.87 [8.4%]	±0.04 [5.2%]
alpha_WZ_fac_syst	±0.96 [3.6%]	±0.27 [2.2%]	±0.71 [10.2%]	±1.08 [10.4%]	±0.28 [36.2%]
alpha_JES1	±0.91 [3.4%]	±0.01 [0.07%]	±0.09 [1.3%]	±0.37 [3.5%]	±0.05 [6.8%]
alpha_JES6	±0.63 [2.3%]	±0.03 [0.24%]	±0.31 [4.4%]	±0.76 [7.3%]	±0.04 [4.7%]
alpha_topEW_syst_CR_Gbb_2800_1400	±0.60 [2.2%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
Lumi	±0.38 [1.4%]	±0.25 [2.0%]	±0.15 [2.1%]	±0.25 [2.4%]	±0.02 [2.3%]
alpha_JER5	±0.31 [1.1%]	±0.43 [3.5%]	±0.54 [7.7%]	±0.07 [0.69%]	±0.01 [1.9%]
alpha_WZ_qsf_syst	±0.30 [1.1%]	±0.36 [2.9%]	±0.14 [2.0%]	±0.24 [2.3%]	±0.01 [1.7%]
alpha_JER3	±0.30 [1.1%]	±0.29 [2.3%]	±0.75 [10.7%]	±0.04 [0.42%]	±0.02 [3.1%]
alpha_JER7	±0.27 [1.0%]	±0.34 [2.8%]	±0.64 [9.1%]	±0.46 [4.4%]	±0.01 [1.2%]
alpha_bTag_C	±0.25 [0.94%]	±0.14 [1.2%]	±0.17 [2.5%]	±0.16 [1.6%]	±0.01 [1.2%]
alpha_bTag_L	±0.23 [0.85%]	±0.39 [3.1%]	±0.13 [1.8%]	±0.21 [2.0%]	±0.03 [4.2%]
alpha_JES5	±0.19 [0.70%]	±0.01 [0.05%]	±0.00 [0.00%]	±0.01 [0.14%]	±0.00 [0.01%]
alpha_JER4	±0.17 [0.65%]	±0.44 [3.6%]	±0.65 [9.2%]	±0.23 [2.2%]	±0.03 [3.7%]
alpha_JER6	±0.16 [0.58%]	±0.36 [2.9%]	±1.47 [21.0%]	±0.04 [0.36%]	±0.04 [5.1%]
alpha_bTag_B	±0.14 [0.52%]	±0.11 [0.91%]	±0.07 [0.97%]	±0.11 [1.0%]	±0.01 [1.9%]
alpha_JES2	±0.13 [0.48%]	±0.25 [2.0%]	±0.00 [0.01%]	±0.14 [1.3%]	±0.01 [1.1%]
alpha_bTag_extrapol_charm	±0.06 [0.24%]	±0.08 [0.67%]	±0.06 [0.82%]	±0.06 [0.58%]	±0.00 [0.08%]
alpha_diboson_syst_CR_Gbb_2800_1400	±0.06 [0.20%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JVT	±0.04 [0.14%]	±0.03 [0.23%]	±0.02 [0.30%]	±0.03 [0.26%]	±0.00 [0.16%]
alpha_bTag_extrapol	±0.03 [0.10%]	±0.23 [1.9%]	±0.02 [0.34%]	±0.02 [0.22%]	±0.02 [2.1%]
alpha_JES4	±0.00 [0.02%]	±0.00 [0.01%]	±0.01 [0.07%]	±0.01 [0.05%]	±0.00 [0.02%]
alpha_JES3	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_topEW_syst_VR3_Gbb_2800_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.22 [2.1%]	±0.00 [0.00%]
alpha_ttbb_syst	±0.00 [0.00%]	±0.02 [0.13%]	±0.07 [0.95%]	±0.01 [0.06%]	±0.00 [0.33%]
gamma_stat_VR3_Gbb_2800_1400_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±1.70 [16.3%]	±0.00 [0.00%]
alpha_ttbar_syst_VR2_Gbb_2800_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.94 [13.4%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_SR_Gbb_2800_1400_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.31 [40.7%]
alpha_st_syst_VR2_Gbb_2800_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.90 [12.9%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_st_syst_VR3_Gbb_2800_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±1.18 [11.3%]	±0.00 [0.00%]
alpha_ttcc_syst	±0.00 [0.00%]	±0.03 [0.26%]	±0.02 [0.35%]	±0.01 [0.14%]	±0.00 [0.27%]
alpha_topEW_syst_VR2_Gbb_2800_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.22 [3.1%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_diboson_syst_VR3_Gbb_2800_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.17 [1.6%]	±0.00 [0.00%]
alpha_ttbar_syst_VR1_Gbb_2800_1400	±0.00 [0.00%]	±1.85 [15.1%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_VR2_Gbb_2800_1400_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±1.12 [16.0%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_st_syst_SR_Gbb_2800_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.11 [13.8%]
alpha_topEW_syst_SR_Gbb_2800_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.16%]
alpha_kin_RW	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_VR1_Gbb_2800_1400_cuts_bin_0	±0.00 [0.00%]	±1.99 [16.2%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_SR_Gbb_2800_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.07 [9.3%]
alpha_topEW_syst_VR1_Gbb_2800_1400	±0.00 [0.00%]	±0.27 [2.2%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_VR3_Gbb_2800_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.90 [8.7%]	±0.00 [0.00%]
alpha_diboson_syst_VR1_Gbb_2800_1400	±0.00 [0.00%]	±0.09 [0.77%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_st_syst_VR1_Gbb_2800_1400	±0.00 [0.00%]	±1.17 [9.5%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]

Table 16 – Breakdown of the dominant systematic uncertainties on background estimates for region Gbb_2800_1400

	CR_Gbb_2300_1000	VR1	VR2	VR3	SR
Total background expectation	29.00	15.57	12.22	8.26	0.75
Total statistical ($\sqrt{N_{\text{exp}}}$)	±5.39	±3.95	±3.50	±2.87	±0.87
Total background systematic	±5.86 [20.21%]	±5.32 [34.19%]	±5.91 [48.36%]	±5.43 [65.69%]	±0.79 [105.18%]
mu_ttbar_Gbb_2300_1000	±10.97 [37.8%]	±3.29 [21.1%]	±2.83 [23.2%]	±0.29 [3.5%]	±0.08 [10.5%]
alpha_QCDHundred	±6.55 [22.6%]	±3.92 [25.2%]	±1.57 [12.8%]	±0.61 [7.3%]	±0.10 [13.1%]
alpha_WZ_ckkw_syst	±3.48 [12.0%]	±1.75 [11.2%]	±2.60 [21.3%]	±1.82 [22.0%]	±0.44 [58.7%]
gamma_stat_CR_Gbb_2300_1000_cuts_bin_0	±2.60 [9.0%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JER1	±1.81 [6.2%]	±1.12 [7.2%]	±1.99 [16.3%]	±3.01 [36.5%]	±0.16 [20.7%]
alpha_JER0	±1.44 [5.0%]	±0.55 [3.5%]	±1.48 [12.1%]	±0.45 [5.4%]	±0.13 [17.7%]
alpha_WZ_renorm_syst	±1.41 [4.9%]	±1.46 [9.4%]	±0.82 [6.7%]	±2.48 [30.0%]	±0.03 [4.5%]
alpha_JER2	±1.39 [4.8%]	±0.83 [5.4%]	±1.16 [9.5%]	±0.27 [3.3%]	±0.12 [16.2%]
alpha_WZ_fac_syst	±1.27 [4.4%]	±0.29 [1.8%]	±1.06 [8.7%]	±0.16 [2.0%]	±0.23 [30.9%]
alpha_JES6	±0.99 [3.4%]	±0.26 [1.7%]	±0.06 [0.48%]	±0.50 [6.1%]	±0.12 [15.5%]
alpha_JES0	±0.91 [3.1%]	±0.34 [2.2%]	±0.07 [0.58%]	±0.57 [7.0%]	±0.11 [14.8%]
alpha_JER6	±0.86 [3.0%]	±0.85 [5.5%]	±1.60 [13.1%]	±0.55 [6.7%]	±0.14 [19.1%]
alpha_JES1	±0.84 [2.9%]	±0.40 [2.5%]	±0.14 [1.1%]	±0.85 [10.2%]	±0.13 [16.8%]
alpha_JER3	±0.71 [2.4%]	±0.67 [4.3%]	±0.51 [4.2%]	±0.27 [3.3%]	±0.17 [22.8%]
alpha_topEW_syst_CR_Gbb_2300_1000	±0.68 [2.4%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JER5	±0.59 [2.0%]	±0.87 [5.6%]	±0.37 [3.1%]	±0.03 [0.32%]	±0.22 [29.6%]
alpha_JES2	±0.51 [1.8%]	±0.33 [2.1%]	±0.31 [2.6%]	±0.02 [0.29%]	±0.02 [2.5%]
alpha_JER4	±0.50 [1.7%]	±0.74 [4.8%]	±1.32 [10.8%]	±0.40 [4.9%]	±0.15 [20.4%]
Lumi	±0.46 [1.6%]	±0.34 [2.2%]	±0.26 [2.1%]	±0.23 [2.8%]	±0.02 [2.5%]
alpha_WZ_qsf_syst	±0.44 [1.5%]	±0.60 [3.8%]	±0.21 [1.8%]	±1.07 [12.9%]	±0.01 [1.3%]
alpha_bTag_L	±0.40 [1.4%]	±0.35 [2.3%]	±0.13 [1.1%]	±0.14 [1.7%]	±0.03 [3.7%]
alpha_bTag_C	±0.28 [0.96%]	±0.23 [1.4%]	±0.32 [2.6%]	±0.08 [0.99%]	±0.02 [2.4%]
alpha_JES5	±0.18 [0.62%]	±0.01 [0.08%]	±0.00 [0.03%]	±0.00 [0.03%]	±0.00 [0.03%]
alpha_bTag_B	±0.17 [0.58%]	±0.17 [1.1%]	±0.12 [1.0%]	±0.05 [0.62%]	±0.01 [2.0%]
alpha_JER7	±0.17 [0.57%]	±0.13 [0.86%]	±0.98 [8.1%]	±0.44 [5.3%]	±0.16 [21.0%]
alpha_bTag_extrapol_charm	±0.12 [0.42%]	±0.10 [0.61%]	±0.09 [0.76%]	±0.01 [0.12%]	±0.00 [0.37%]
alpha_JVT	±0.05 [0.17%]	±0.04 [0.23%]	±0.04 [0.37%]	±0.04 [0.45%]	±0.00 [0.18%]
alpha_bTag_extrapol	±0.02 [0.06%]	±0.30 [1.9%]	±0.02 [0.15%]	±0.02 [0.26%]	±0.02 [2.3%]
alpha_JES4	±0.00 [0.02%]	±0.00 [0.01%]	±0.00 [0.00%]	±0.00 [0.03%]	±0.00 [0.00%]
alpha_JES3	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_st_syst_SR_Gbb_2300_1000	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.13 [16.7%]
alpha_ttbb_syst	±0.00 [0.00%]	±0.03 [0.21%]	±0.07 [0.53%]	±0.03 [0.31%]	±0.01 [0.83%]
gamma_stat_VR1_Gbb_2300_1000_cuts_bin_0	±0.00 [0.00%]	±2.06 [13.3%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_st_syst_VR1_Gbb_2300_1000	±0.00 [0.00%]	±1.59 [10.2%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_VR3_Gbb_2300_1000_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±2.74 [33.2%]	±0.00 [0.00%]
alpha_diboson_syst_VR3_Gbb_2300_1000	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.12 [1.4%]	±0.00 [0.00%]
alpha_ttbar_syst_VR3_Gbb_2300_1000	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.17 [2.1%]	±0.00 [0.00%]
alpha_st_syst_VR3_Gbb_2300_1000	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.66 [8.0%]	±0.00 [0.00%]
alpha_topEW_syst_SR_Gbb_2300_1000	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.16%]
alpha_ttcc_syst	±0.00 [0.00%]	±0.04 [0.26%]	±0.03 [0.26%]	±0.02 [0.26%]	±0.01 [0.98%]
alpha_ttbar_syst_SR_Gbb_2300_1000	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.05 [6.2%]
alpha_ttbar_syst_VR1_Gbb_2300_1000	±0.00 [0.00%]	±1.95 [12.5%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_topEW_syst_VR3_Gbb_2300_1000	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.07 [0.85%]	±0.00 [0.00%]
alpha_topEW_syst_VR2_Gbb_2300_1000	±0.00 [0.00%]	±0.00 [0.00%]	±0.36 [2.9%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_VR2_Gbb_2300_1000_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±1.42 [11.7%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_SR_Gbb_2300_1000_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.33 [43.5%]
alpha_diboson_syst_VR2_Gbb_2300_1000	±0.00 [0.00%]	±0.00 [0.00%]	±0.06 [0.51%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_kin_RW	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_VR2_Gbb_2300_1000	±0.00 [0.00%]	±0.00 [0.00%]	±1.68 [13.7%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_topEW_syst_VR1_Gbb_2300_1000	±0.00 [0.00%]	±0.37 [2.4%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_st_syst_VR2_Gbb_2300_1000	±0.00 [0.00%]	±0.00 [0.00%]	±1.54 [12.6%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_diboson_syst_VR1_Gbb_2300_1000	±0.00 [0.00%]	±0.09 [0.61%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]

Table 17 – Breakdown of the dominant systematic uncertainties on background estimates for region Gbb_2300_1000

	CR_Gbb_2100_1600	VR1	VR2	VR3	SR
Total background expectation	48.07	810.65	33.91	17.07	1.91
Total statistical ($\sqrt{N_{\text{exp}}}$)	±6.93	±28.47	±5.82	±4.13	±1.38
Total background systematic	±7.29 [15.16%]	±329.95 [40.70%]	±16.10 [47.49%]	±11.18 [65.51%]	±1.76 [92.11%]
mu_ttbar_Gbb_2100_1600	±9.62 [20.0%]	±150.02 [18.5%]	±6.62 [19.5%]	±2.10 [12.3%]	±0.11 [5.5%]
alpha_WZ_ckkw_syst	±4.59 [9.5%]	±75.64 [9.3%]	±1.67 [4.9%]	±4.70 [27.5%]	±1.34 [70.0%]
gamma_stat_CR_Gbb_2100_1600_cuts_bin_0	±3.64 [7.6%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDHundred	±2.74 [5.7%]	±44.73 [5.5%]	±0.09 [0.27%]	±0.08 [0.48%]	±0.13 [6.9%]
alpha_WZ_fac_syst	±2.01 [4.2%]	±29.71 [3.7%]	±0.24 [0.72%]	±1.98 [11.6%]	±0.71 [37.2%]
alpha_JER1	±1.67 [3.5%]	±7.57 [0.93%]	±0.84 [2.5%]	±5.03 [29.5%]	±0.11 [5.9%]
alpha_JER4	±1.67 [3.5%]	±3.24 [0.40%]	±1.69 [5.0%]	±1.93 [11.3%]	±0.02 [1.1%]
alpha_WZ_renorm_syst	±1.21 [2.5%]	±26.17 [3.2%]	±1.32 [3.9%]	±1.39 [8.2%]	±0.10 [5.0%]
alpha_JER0	±0.95 [2.0%]	±44.81 [5.5%]	±1.04 [3.1%]	±4.32 [25.3%]	±0.35 [18.4%]
alpha_JER3	±0.93 [1.9%]	±39.12 [4.8%]	±3.73 [11.0%]	±2.70 [15.8%]	±0.11 [5.9%]
alpha_JER7	±0.88 [1.8%]	±0.80 [0.10%]	±0.36 [1.1%]	±2.71 [15.9%]	±0.11 [5.6%]
alpha_JER5	±0.87 [1.8%]	±6.62 [0.82%]	±1.16 [3.4%]	±1.92 [11.2%]	±0.02 [0.97%]
alpha_JES0	±0.74 [1.5%]	±30.65 [3.8%]	±4.09 [12.1%]	±0.68 [4.0%]	±0.18 [9.2%]
alpha_JES1	±0.56 [1.2%]	±17.40 [2.1%]	±2.06 [6.1%]	±0.24 [1.4%]	±0.13 [7.0%]
alpha_JES6	±0.42 [0.87%]	±42.44 [5.2%]	±2.84 [8.4%]	±0.25 [1.5%]	±0.05 [2.8%]
alpha_topEW_syst_CR_Gbb_2100_1600	±0.38 [0.78%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JER2	±0.35 [0.73%]	±22.64 [2.8%]	±1.14 [3.4%]	±2.78 [16.3%]	±0.36 [18.9%]
Lumi	±0.34 [0.71%]	±7.12 [0.88%]	±0.26 [0.77%]	±0.26 [1.5%]	±0.04 [2.3%]
alpha_WZ_qsf_syst	±0.26 [0.54%]	±7.14 [0.88%]	±0.49 [1.4%]	±0.32 [1.9%]	±0.05 [2.6%]
alpha_bTag_C	±0.24 [0.51%]	±3.95 [0.49%]	±0.16 [0.48%]	±0.05 [0.31%]	±0.05 [2.8%]
alpha_diboson_syst_CR_Gbb_2100_1600	±0.18 [0.38%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_bTag_L	±0.14 [0.28%]	±8.54 [1.1%]	±0.33 [0.97%]	±0.49 [2.9%]	±0.06 [3.0%]
alpha_bTag_B	±0.13 [0.26%]	±3.11 [0.38%]	±0.18 [0.54%]	±0.11 [0.66%]	±0.01 [0.77%]
alpha_bTag_extrapol_charm	±0.07 [0.15%]	±0.21 [0.03%]	±0.13 [0.40%]	±0.05 [0.31%]	±0.00 [0.20%]
alpha_JER6	±0.05 [0.10%]	±46.89 [5.8%]	±2.60 [7.7%]	±3.65 [21.4%]	±0.36 [19.0%]
alpha_JVT	±0.04 [0.09%]	±0.89 [0.11%]	±0.08 [0.24%]	±0.04 [0.22%]	±0.01 [0.39%]
alpha_JES5	±0.04 [0.07%]	±0.84 [0.10%]	±0.22 [0.66%]	±0.01 [0.09%]	±0.02 [0.82%]
alpha_JES4	±0.03 [0.07%]	±2.65 [0.33%]	±0.10 [0.29%]	±0.07 [0.43%]	±0.00 [0.09%]
alpha_bTag_extrapol	±0.02 [0.05%]	±0.13 [0.02%]	±0.05 [0.15%]	±0.01 [0.07%]	±0.01 [0.47%]
alpha_JES2	±0.01 [0.02%]	±0.52 [0.06%]	±0.15 [0.46%]	±0.43 [2.5%]	±0.03 [1.4%]
alpha_JES3	±0.00 [0.00%]	±0.11 [0.01%]	±0.00 [0.01%]	±0.00 [0.01%]	±0.00 [0.01%]
alpha_st_syst_VR1_Gbb_2100_1600	±0.00 [0.00%]	±33.95 [4.2%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbb_syst	±0.00 [0.00%]	±7.02 [0.87%]	±0.12 [0.34%]	±0.07 [0.42%]	±0.00 [0.22%]
gamma_stat_VR3_Gbb_2100_1600_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±2.39 [14.0%]	±0.00 [0.00%]
alpha_diboson_syst_VR1_Gbb_2100_1600	±0.00 [0.00%]	±4.02 [0.50%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_topEW_syst_VR1_Gbb_2100_1600	±0.00 [0.00%]	±11.62 [1.4%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_VR1_Gbb_2100_1600	±0.00 [0.00%]	±282.04 [34.8%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttcc_syst	±0.00 [0.00%]	±5.63 [0.69%]	±0.05 [0.14%]	±0.46 [2.7%]	±0.01 [0.34%]
alpha_st_syst_SR_Gbb_2100_1600	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.19 [9.8%]
alpha_diboson_syst_VR2_Gbb_2100_1600	±0.00 [0.00%]	±0.00 [0.00%]	±0.11 [0.32%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_topEW_syst_VR2_Gbb_2100_1600	±0.00 [0.00%]	±0.00 [0.00%]	±0.76 [2.2%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_VR2_Gbb_2100_1600_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±2.36 [7.0%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_st_syst_VR2_Gbb_2100_1600	±0.00 [0.00%]	±0.00 [0.00%]	±1.28 [3.8%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_VR2_Gbb_2100_1600	±0.00 [0.00%]	±0.00 [0.00%]	±12.45 [36.7%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_topEW_syst_SR_Gbb_2100_1600	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.04 [2.3%]
alpha_st_syst_VR3_Gbb_2100_1600	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.71 [4.1%]	±0.00 [0.00%]
alpha_kin_RW	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_SR_Gbb_2100_1600	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.20 [10.4%]
alpha_diboson_syst_VR3_Gbb_2100_1600	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.48 [2.8%]	±0.00 [0.00%]
alpha_diboson_syst_SR_Gbb_2100_1600	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.07 [3.9%]
gamma_stat_SR_Gbb_2100_1600_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.55 [28.7%]
alpha_topEW_syst_VR3_Gbb_2100_1600	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.05 [0.30%]	±0.00 [0.00%]
alpha_ttbar_syst_VR3_Gbb_2100_1600	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±3.95 [23.2%]	±0.00 [0.00%]

Table 18 – Breakdown of the dominant systematic uncertainties on background estimates for region Gbb_2100_1600

	CR_Gbb_2000_1800	VR1	VR2	VR3	SR
Total background expectation	29.04	2123.48	119.74	12.90	1.75
Total statistical ($\sqrt{N_{\text{exp}}}$)	±5.39	±46.08	±10.94	±3.59	±1.32
Total background systematic	±5.44 [18.73%]	±1078.83 [50.80%]	±61.78 [51.59%]	±6.50 [50.39%]	±1.30 [74.43%]
mu_ttbar_Gbb_2000_1800	±6.19 [21.3%]	±452.92 [21.3%]	±25.88 [21.6%]	±1.38 [10.7%]	±0.18 [10.2%]
gamma_stat_CR_Gbb_2000_1800_cuts_bin_0	±2.22 [7.7%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_WZ_ckkw_syst	±1.29 [4.4%]	±52.25 [2.5%]	±2.90 [2.4%]	±0.86 [6.7%]	±0.06 [3.6%]
alpha_QCDHundred	±0.85 [2.9%]	±15.32 [0.72%]	±0.59 [0.49%]	±0.04 [0.33%]	±0.04 [2.4%]
alpha_JER0	±0.73 [2.5%]	±109.81 [5.2%]	±6.42 [5.4%]	±2.70 [20.9%]	±0.53 [30.2%]
alpha_WZ_renorm_syst	±0.67 [2.3%]	±17.55 [0.83%]	±0.90 [0.76%]	±0.02 [0.18%]	±0.16 [9.3%]
alpha_JER1	±0.64 [2.2%]	±73.41 [3.5%]	±2.07 [1.7%]	±3.05 [23.6%]	±0.43 [24.4%]
alpha_JER7	±0.62 [2.1%]	±81.92 [3.9%]	±3.46 [2.9%]	±1.50 [11.6%]	±0.18 [10.4%]
alpha_JER2	±0.47 [1.6%]	±46.30 [2.2%]	±1.44 [1.2%]	±2.15 [16.7%]	±0.39 [22.2%]
alpha_JER3	±0.46 [1.6%]	±157.04 [7.4%]	±9.14 [7.6%]	±0.19 [1.5%]	±0.48 [27.2%]
alpha_JER4	±0.39 [1.4%]	±199.56 [9.4%]	±11.33 [9.5%]	±0.59 [4.6%]	±0.38 [21.5%]
alpha_WZ_fac_syst	±0.37 [1.3%]	±18.29 [0.86%]	±1.06 [0.89%]	±0.38 [3.0%]	±0.05 [2.6%]
alpha_JES0	±0.35 [1.2%]	±43.82 [2.1%]	±1.48 [1.2%]	±0.32 [2.5%]	±0.01 [0.75%]
alpha_JER6	±0.34 [1.2%]	±22.25 [1.0%]	±0.24 [0.20%]	±2.25 [17.4%]	±0.25 [14.2%]
alpha_JES6	±0.32 [1.1%]	±81.48 [3.8%]	±1.28 [1.1%]	±0.28 [2.2%]	±0.25 [14.3%]
alpha_WZ_qsf_syst	±0.28 [0.98%]	±2.10 [0.10%]	±0.08 [0.07%]	±0.30 [2.3%]	±0.07 [4.3%]
alpha_JER5	±0.19 [0.67%]	±170.76 [8.0%]	±12.11 [10.1%]	±0.33 [2.6%]	±0.15 [8.5%]
alpha_topEW_syst_CR_Gbb_2000_1800	±0.15 [0.51%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_bTag_C	±0.13 [0.44%]	±4.52 [0.21%]	±0.38 [0.32%]	±0.08 [0.63%]	±0.04 [2.0%]
alpha_JES2	±0.13 [0.44%]	±1.62 [0.08%]	±1.93 [1.6%]	±0.13 [1.0%]	±0.22 [12.7%]
Lumi	±0.11 [0.39%]	±8.29 [0.39%]	±0.43 [0.36%]	±0.21 [1.6%]	±0.03 [1.7%]
alpha_bTag_L	±0.04 [0.15%]	±23.74 [1.1%]	±1.01 [0.84%]	±0.31 [2.4%]	±0.02 [1.1%]
alpha_JES1	±0.04 [0.13%]	±56.08 [2.6%]	±6.92 [5.8%]	±0.17 [1.3%]	±0.02 [0.98%]
alpha_bTag_B	±0.04 [0.12%]	±15.40 [0.72%]	±1.11 [0.93%]	±0.15 [1.2%]	±0.01 [0.51%]
alpha_bTag_extrapol_charm	±0.03 [0.09%]	±7.58 [0.36%]	±0.39 [0.33%]	±0.07 [0.53%]	±0.04 [2.0%]
alpha_JVT	±0.03 [0.09%]	±2.25 [0.11%]	±0.20 [0.17%]	±0.05 [0.35%]	±0.01 [0.84%]
alpha_bTag_extrapol	±0.01 [0.02%]	±3.29 [0.15%]	±0.22 [0.19%]	±0.00 [0.02%]	±0.00 [0.10%]
alpha_JES5	±0.00 [0.02%]	±11.53 [0.54%]	±0.58 [0.49%]	±0.10 [0.79%]	±0.00 [0.26%]
alpha_JES4	±0.00 [0.00%]	±6.08 [0.29%]	±0.26 [0.21%]	±0.02 [0.12%]	±0.00 [0.12%]
alpha_JES3	±0.00 [0.00%]	±0.01 [0.00%]	±0.08 [0.06%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_st_syst_SR_Gbb_2000_1800	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.01 [0.85%]
alpha_ttbb_syst	±0.00 [0.00%]	±23.95 [1.1%]	±0.96 [0.80%]	±0.01 [0.06%]	±0.00 [0.04%]
alpha_topEW_syst_SR_Gbb_2000_1800	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.03 [1.4%]
alpha_ttbar_syst_VR2_Gbb_2000_1800	±0.00 [0.00%]	±0.00 [0.00%]	±52.47 [43.8%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_topEW_syst_VR1_Gbb_2000_1800	±0.00 [0.00%]	±17.14 [0.81%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttcc_syst	±0.00 [0.00%]	±12.52 [0.59%]	±0.79 [0.66%]	±0.28 [2.2%]	±0.05 [3.0%]
alpha_diboson_syst_VR1_Gbb_2000_1800	±0.00 [0.00%]	±5.57 [0.26%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_SR_Gbb_2000_1800_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.59 [33.6%]
alpha_topEW_syst_VR2_Gbb_2000_1800	±0.00 [0.00%]	±0.00 [0.00%]	±1.00 [0.83%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_st_syst_VR3_Gbb_2000_1800	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.56 [4.3%]	±0.00 [0.00%]
alpha_diboson_syst_SR_Gbb_2000_1800	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.07 [3.9%]
alpha_ttbar_syst_VR3_Gbb_2000_1800	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±2.79 [21.6%]	±0.00 [0.00%]
gamma_stat_VR3_Gbb_2000_1800_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±2.20 [17.0%]	±0.00 [0.00%]
alpha_kin_RW	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_diboson_syst_VR2_Gbb_2000_1800	±0.00 [0.00%]	±0.00 [0.00%]	±0.25 [0.21%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_SR_Gbb_2000_1800	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.36 [20.6%]
alpha_st_syst_VR2_Gbb_2000_1800	±0.00 [0.00%]	±0.00 [0.00%]	±2.19 [1.8%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_VR1_Gbb_2000_1800	±0.00 [0.00%]	±918.21 [43.2%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_st_syst_VR1_Gbb_2000_1800	±0.00 [0.00%]	±40.02 [1.9%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_topEW_syst_VR3_Gbb_2000_1800	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.05 [0.42%]	±0.00 [0.00%]

Table 19 – Breakdown of the dominant systematic uncertainties on background estimates for region Gbb_2000_1800

	CR_Gtt_2100_1	VR1	VR2	SR
Total background expectation	35.00	77.61	41.62	0.54
Total statistical ($\sqrt{N_{\rm exp}}$)	±5.92	±8.81	±6.45	±0.73
Total background systematic	±5.92 [16.92%]	±24.41 [31.46%]	±14.45 [34.73%]	±0.23 [42.77%]
mu_ttbar_Gtt_2100_1	±6.60 [18.9%]	±10.27 [13.2%]	±5.21 [12.5%]	±0.05 [9.5%]
alpha_JES6	±1.16 [3.3%]	±0.20 [0.26%]	±0.63 [1.5%]	±0.02 [3.4%]
alpha_JES1	±1.09 [3.1%]	±1.00 [1.3%]	±0.98 [2.4%]	±0.01 [1.8%]
alpha_JES0	±1.00 [2.9%]	±0.67 [0.87%]	±0.80 [1.9%]	±0.02 [4.5%]
alpha_WZ_renorm_syst	±0.96 [2.7%]	±3.61 [4.7%]	±1.67 [4.0%]	±0.00 [0.74%]
alpha_topEW_syst_CR_Gtt_2100_1	±0.89 [2.6%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_kin_RW	±0.73 [2.1%]	±1.56 [2.0%]	±2.41 [5.8%]	±0.05 [8.7%]
alpha_WZ_ckkw_syst	±0.59 [1.7%]	±3.98 [5.1%]	±2.98 [7.2%]	±0.06 [10.6%]
alpha_JER7	±0.58 [1.6%]	±1.22 [1.6%]	±1.37 [3.3%]	±0.02 [3.6%]
alpha_QCDHundred	±0.55 [1.6%]	±2.11 [2.7%]	±0.79 [1.9%]	±0.04 [8.0%]
alpha_JER2	±0.46 [1.3%]	±0.33 [0.43%]	±2.47 [5.9%]	±0.05 [9.3%]
alpha_bTag_L	±0.44 [1.3%]	±1.38 [1.8%]	±0.86 [2.1%]	±0.02 [4.4%]
alpha_WZ_qsf_syst	±0.43 [1.2%]	±1.55 [2.0%]	±0.62 [1.5%]	±0.00 [0.39%]
alpha_JER1	±0.39 [1.1%]	±0.25 [0.32%]	±2.66 [6.4%]	±0.05 [10.2%]
alpha_JER6	±0.35 [0.99%]	±0.61 [0.79%]	±1.66 [4.0%]	±0.02 [4.2%]
alpha_JER4	±0.34 [0.97%]	±1.29 [1.7%]	±1.32 [3.2%]	±0.02 [3.5%]
alpha_bTag_C	±0.29 [0.83%]	±1.05 [1.3%]	±0.63 [1.5%]	±0.02 [4.0%]
Lumi	±0.27 [0.78%]	±1.09 [1.4%]	±0.62 [1.5%]	±0.01 [1.8%]
alpha_JES2	±0.24 [0.69%]	±1.13 [1.5%]	±0.23 [0.55%]	±0.00 [0.22%]
alpha_bTag_B	±0.23 [0.66%]	±0.79 [1.0%]	±0.47 [1.1%]	±0.01 [1.2%]
alpha_JES5	±0.19 [0.53%]	±0.03 [0.04%]	±0.09 [0.21%]	±0.00 [0.03%]
alpha_diboson_syst_CR_Gtt_2100_1	±0.17 [0.49%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_WZ_fac_syst	±0.13 [0.36%]	±0.51 [0.65%]	±0.86 [2.1%]	±0.03 [5.6%]
alpha_bTag_extrapol_charm	±0.11 [0.32%]	±0.26 [0.33%]	±0.14 [0.33%]	±0.00 [0.69%]
alpha_JER5	±0.08 [0.24%]	±0.78 [1.0%]	±1.26 [3.0%]	±0.01 [1.7%]
alpha_bTag_extrapol	±0.07 [0.20%]	±0.75 [0.97%]	±0.10 [0.23%]	±0.01 [2.0%]
alpha_JER0	±0.07 [0.19%]	±0.79 [1.0%]	±3.23 [7.8%]	±0.04 [7.3%]
alpha_JVT	±0.05 [0.13%]	±0.25 [0.32%]	±0.17 [0.40%]	±0.00 [0.28%]
alpha_JES4	±0.05 [0.13%]	±0.04 [0.05%]	±0.02 [0.04%]	±0.00 [0.01%]
alpha_JER3	±0.03 [0.09%]	±1.54 [2.0%]	±0.92 [2.2%]	±0.01 [2.3%]
alpha_JES3	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbb_syst	±0.00 [0.00%]	±0.30 [0.39%]	±0.37 [0.90%]	±0.00 [0.77%]
alpha_ttbar_syst_VR1_Gtt_2100_1	±0.00 [0.00%]	±19.91 [25.7%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_SR_Gtt_2100_1	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.10 [18.4%]
alpha_st_syst_VR2_Gtt_2100_1	±0.00 [0.00%]	±0.00 [0.00%]	±4.53 [10.9%]	±0.00 [0.00%]
alpha_ttcc_syst	±0.00 [0.00%]	±0.04 [0.05%]	±0.17 [0.42%]	±0.01 [1.1%]
alpha_st_syst_SR_Gtt_2100_1	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.07 [13.7%]
alpha_topEW_syst_SR_Gtt_2100_1	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.07 [13.7%]
gamma_stat_SR_Gtt_2100_1_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.11 [19.8%]
alpha_diboson_syst_VR1_Gtt_2100_1	±0.00 [0.00%]	±1.29 [1.7%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_VR2_Gtt_2100_1	±0.00 [0.00%]	±0.00 [0.00%]	±10.09 [24.2%]	±0.00 [0.00%]
alpha_st_syst_VR1_Gtt_2100_1	±0.00 [0.00%]	±7.10 [9.1%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_topEW_syst_VR2_Gtt_2100_1	±0.00 [0.00%]	±0.00 [0.00%]	±3.26 [7.8%]	±0.00 [0.00%]
alpha_diboson_syst_VR2_Gtt_2100_1	±0.00 [0.00%]	±0.00 [0.00%]	±0.09 [0.23%]	±0.00 [0.00%]
alpha_topEW_syst_VR1_Gtt_2100_1	±0.00 [0.00%]	±4.71 [6.1%]	±0.00 [0.00%]	±0.00 [0.00%]

Table 20 – Breakdown of the dominant systematic uncertainties on background estimates for region Gtt_2100_1

	CR_Gtt_1800_1	VR1	VR2	SR
Total background expectation	31.00	107.44	28.24	1.07
Total statistical ($\sqrt{N_{\text{exp}}}$)	±5.57	±10.37	±5.31	±1.03
Total background systematic	±5.57 [17.97%]	±34.76 [32.36%]	±9.35 [33.10%]	±0.41 [38.71%]
mu_ttbar_Gtt_1800_1	±6.60 [21.3%]	±16.20 [15.1%]	±3.74 [13.3%]	±0.10 [8.9%]
alpha_JER1	±1.66 [5.4%]	±0.52 [0.48%]	±0.59 [2.1%]	±0.03 [2.9%]
alpha_JER2	±1.38 [4.4%]	±1.02 [0.95%]	±0.78 [2.8%]	±0.12 [10.9%]
alpha_JER0	±1.20 [3.9%]	±0.93 [0.87%]	±0.64 [2.3%]	±0.07 [6.6%]
alpha_topEW_syst_CR_Gtt_1800_1	±1.09 [3.5%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JER6	±1.00 [3.2%]	±0.22 [0.21%]	±0.64 [2.3%]	±0.05 [4.5%]
alpha_QCDHundred	±0.91 [2.9%]	±2.24 [2.1%]	±0.74 [2.6%]	±0.04 [4.2%]
alpha_kin_RW	±0.74 [2.4%]	±2.78 [2.6%]	±1.60 [5.7%]	±0.07 [6.3%]
alpha_JER3	±0.74 [2.4%]	±1.77 [1.6%]	±0.60 [2.1%]	±0.00 [0.17%]
alpha_JER7	±0.73 [2.3%]	±1.52 [1.4%]	±0.42 [1.5%]	±0.01 [0.55%]
alpha_WZ_ckkw_syst	±0.66 [2.1%]	±6.68 [6.2%]	±2.09 [7.4%]	±0.03 [3.2%]
alpha_JER4	±0.54 [1.7%]	±0.35 [0.33%]	±0.93 [3.3%]	±0.04 [4.2%]
alpha_JER5	±0.53 [1.7%]	±1.76 [1.6%]	±0.29 [1.0%]	±0.01 [0.61%]
alpha_bTag_L	±0.41 [1.3%]	±1.61 [1.5%]	±0.27 [0.95%]	±0.03 [2.9%]
alpha_JES0	±0.36 [1.2%]	±2.43 [2.3%]	±0.95 [3.4%]	±0.04 [4.1%]
alpha_WZ_renorm_syst	±0.31 [0.99%]	±4.50 [4.2%]	±1.42 [5.0%]	±0.14 [12.9%]
alpha_bTag_C	±0.27 [0.88%]	±1.48 [1.4%]	±0.57 [2.0%]	±0.03 [2.5%]
Lumi	±0.25 [0.81%]	±1.52 [1.4%]	±0.45 [1.6%]	±0.02 [2.0%]
alpha_bTag_B	±0.24 [0.76%]	±1.09 [1.0%]	±0.31 [1.1%]	±0.01 [1.4%]
alpha_WZ_fac_syst	±0.22 [0.72%]	±1.59 [1.5%]	±0.49 [1.7%]	±0.05 [4.4%]
alpha_JES1	±0.14 [0.44%]	±2.79 [2.6%]	±0.41 [1.4%]	±0.02 [2.1%]
alpha_bTag_extrapol_charm	±0.13 [0.41%]	±0.35 [0.32%]	±0.14 [0.48%]	±0.00 [0.31%]
alpha_JES2	±0.12 [0.37%]	±0.91 [0.84%]	±0.36 [1.3%]	±0.03 [3.0%]
alpha_WZ_qsf_syst	±0.11 [0.34%]	±1.80 [1.7%]	±0.56 [2.0%]	±0.06 [5.9%]
alpha_JES6	±0.08 [0.25%]	±1.89 [1.8%]	±0.54 [1.9%]	±0.01 [1.0%]
alpha_diboson_syst_CR_Gtt_1800_1	±0.07 [0.24%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JES4	±0.05 [0.16%]	±0.06 [0.05%]	±0.12 [0.41%]	±0.01 [1.4%]
alpha_bTag_extrapol	±0.05 [0.15%]	±0.61 [0.57%]	±0.05 [0.18%]	±0.01 [1.0%]
alpha_JVT	±0.03 [0.09%]	±0.31 [0.29%]	±0.09 [0.32%]	±0.00 [0.46%]
alpha_JES5	±0.02 [0.06%]	±0.02 [0.02%]	±0.04 [0.14%]	±0.00 [0.00%]
alpha_JES3	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_st_syst_VR2_Gtt_1800_1	±0.00 [0.00%]	±0.00 [0.00%]	±3.25 [11.5%]	±0.00 [0.00%]
alpha_ttbb_syst	±0.00 [0.00%]	±0.89 [0.83%]	±0.25 [0.87%]	±0.01 [1.2%]
alpha_diboson_syst_VR1_Gtt_1800_1	±0.00 [0.00%]	±2.04 [1.9%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_SR_Gtt_1800_1	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.16 [15.1%]
alpha_ttcc_syst	±0.00 [0.00%]	±0.14 [0.13%]	±0.01 [0.04%]	±0.00 [0.41%]
alpha_ttbar_syst_VR2_Gtt_1800_1	±0.00 [0.00%]	±0.00 [0.00%]	±6.34 [22.5%]	±0.00 [0.00%]
alpha_topEW_syst_VR2_Gtt_1800_1	±0.00 [0.00%]	±0.00 [0.00%]	±2.12 [7.5%]	±0.00 [0.00%]
gamma_stat_VR2_Gtt_1800_1_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±1.44 [5.1%]	±0.00 [0.00%]
alpha_ttbar_syst_VR1_Gtt_1800_1	±0.00 [0.00%]	±27.45 [25.6%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_SR_Gtt_1800_1_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.24 [22.4%]
alpha_st_syst_VR1_Gtt_1800_1	±0.00 [0.00%]	±9.57 [8.9%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_diboson_syst_VR2_Gtt_1800_1	±0.00 [0.00%]	±0.00 [0.00%]	±0.34 [1.2%]	±0.00 [0.00%]
alpha_st_syst_SR_Gtt_1800_1	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.09 [8.8%]
alpha_topEW_syst_VR1_Gtt_1800_1	±0.00 [0.00%]	±7.30 [6.8%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_topEW_syst_SR_Gtt_1800_1	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.15 [14.1%]

Table 21 – Breakdown of the dominant systematic uncertainties on background estimates for region Gtt_1800_1

	CR_Gtt_2300_1200	VR1	VR2	SR
Total background expectation	46.04	182.49	30.57	0.74
Total statistical ($\sqrt{N_{\text{exp}}}$)	±6.79	±13.51	±5.53	±0.86
Total background systematic	±6.79 [14.74%]	±68.56 [37.57%]	±13.93 [45.55%]	±0.36 [48.73%]
mu_ttbar_Gtt_2300_1200	±7.19 [15.6%]	±23.51 [12.9%]	±3.73 [12.2%]	±0.07 [10.0%]
alpha_topEW_syst_CR_Gtt_2300_1200	±1.12 [2.4%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDHundred	±1.06 [2.3%]	±2.23 [1.2%]	±0.42 [1.4%]	±0.04 [5.8%]
alpha_JER1	±0.73 [1.6%]	±6.43 [3.5%]	±3.69 [12.1%]	±0.10 [13.5%]
alpha_kin_RW	±0.64 [1.4%]	±4.38 [2.4%]	±1.28 [4.2%]	±0.03 [4.7%]
alpha_WZ_ckkw_syst	±0.56 [1.2%]	±8.03 [4.4%]	±1.35 [4.4%]	±0.10 [13.2%]
alpha_JER0	±0.50 [1.1%]	±4.03 [2.2%]	±2.88 [9.4%]	±0.03 [4.4%]
alpha_bTag_L	±0.46 [1.0%]	±3.24 [1.8%]	±0.43 [1.4%]	±0.06 [7.9%]
alpha_JER4	±0.45 [0.99%]	±1.79 [0.98%]	±3.11 [10.2%]	±0.04 [5.4%]
alpha_WZ_renorm_syst	±0.44 [0.96%]	±4.73 [2.6%]	±1.14 [3.7%]	±0.01 [0.94%]
alpha_JER6	±0.44 [0.95%]	±1.61 [0.88%]	±2.89 [9.5%]	±0.03 [3.8%]
alpha_JER3	±0.35 [0.77%]	±3.88 [2.1%]	±3.48 [11.4%]	±0.05 [6.8%]
alpha_JER2	±0.29 [0.63%]	±1.54 [0.84%]	±1.99 [6.5%]	±0.11 [15.1%]
alpha_JER7	±0.29 [0.63%]	±2.06 [1.1%]	±0.24 [0.78%]	±0.03 [3.7%]
alpha_bTag_B	±0.26 [0.56%]	±1.17 [0.64%]	±0.21 [0.70%]	±0.01 [1.1%]
Lumi	±0.26 [0.56%]	±1.76 [0.96%]	±0.33 [1.1%]	±0.01 [1.4%]
alpha_bTag_C	±0.24 [0.52%]	±1.40 [0.77%]	±0.57 [1.9%]	±0.01 [1.7%]
alpha_WZ_qsf_syst	±0.20 [0.44%]	±2.09 [1.1%]	±0.51 [1.7%]	±0.00 [0.07%]
alpha_JER5	±0.18 [0.39%]	±1.90 [1.0%]	±2.53 [8.3%]	±0.03 [3.9%]
alpha_JES0	±0.14 [0.31%]	±4.08 [2.2%]	±1.01 [3.3%]	±0.01 [1.7%]
alpha_WZ_fac_syst	±0.11 [0.23%]	±2.30 [1.3%]	±0.22 [0.73%]	±0.05 [7.0%]
alpha_JES5	±0.10 [0.21%]	±0.15 [0.08%]	±0.11 [0.37%]	±0.00 [0.27%]
alpha_JES2	±0.09 [0.19%]	±0.25 [0.14%]	±0.30 [0.98%]	±0.00 [0.62%]
alpha_bTag_extrapol_charm	±0.07 [0.16%]	±0.51 [0.28%]	±0.11 [0.36%]	±0.00 [0.19%]
alpha_JES6	±0.07 [0.15%]	±1.66 [0.91%]	±0.27 [0.87%]	±0.01 [1.0%]
alpha_bTag_extrapol	±0.06 [0.12%]	±0.30 [0.16%]	±0.07 [0.23%]	±0.00 [0.18%]
alpha_JVT	±0.05 [0.10%]	±0.45 [0.24%]	±0.12 [0.38%]	±0.00 [0.45%]
alpha_JES4	±0.03 [0.07%]	±0.41 [0.22%]	±0.07 [0.24%]	±0.00 [0.55%]
alpha_JES1	±0.01 [0.02%]	±2.20 [1.2%]	±1.34 [4.4%]	±0.01 [2.0%]
alpha_JES3	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_VR2_Gtt_2300_1200_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±1.83 [6.0%]	±0.00 [0.00%]
alpha_ttbb_syst	±0.00 [0.00%]	±1.49 [0.81%]	±0.12 [0.39%]	±0.01 [0.75%]
alpha_diboson_syst_VR2_Gtt_2300_1200	±0.00 [0.00%]	±0.00 [0.00%]	±0.09 [0.30%]	±0.00 [0.00%]
alpha_st_syst_VR2_Gtt_2300_1200	±0.00 [0.00%]	±0.00 [0.00%]	±2.38 [7.8%]	±0.00 [0.00%]
alpha_ttcc_syst	±0.00 [0.00%]	±1.46 [0.80%]	±0.06 [0.19%]	±0.01 [1.1%]
alpha_st_syst_VR1_Gtt_2300_1200	±0.00 [0.00%]	±11.39 [6.2%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_diboson_syst_VR1_Gtt_2300_1200	±0.00 [0.00%]	±1.89 [1.0%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_SR_Gtt_2300_1200_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.17 [22.3%]
alpha_st_syst_SR_Gtt_2300_1200	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.03 [4.4%]
alpha_topEW_syst_VR2_Gtt_2300_1200	±0.00 [0.00%]	±0.00 [0.00%]	±1.39 [4.6%]	±0.00 [0.00%]
alpha_ttbar_syst_VR2_Gtt_2300_1200	±0.00 [0.00%]	±0.00 [0.00%]	±9.66 [31.6%]	±0.00 [0.00%]
alpha_topEW_syst_VR1_Gtt_2300_1200	±0.00 [0.00%]	±9.61 [5.3%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_topEW_syst_SR_Gtt_2300_1200	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.12 [15.8%]
alpha_ttbar_syst_VR1_Gtt_2300_1200	±0.00 [0.00%]	±60.79 [33.3%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_SR_Gtt_2300_1200	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.19 [25.8%]

Table 22 – Breakdown of the dominant systematic uncertainties on background estimates for region Gtt_2300_1200

	CR_Gtt_1900_1400	VR1	VR2	SR
Total background expectation	25.99	427.93	40.55	0.84
Total statistical ($\sqrt{N_{\text{exp}}}$)	±5.10	±20.69	±6.37	±0.92
Total background systematic	±5.11 [19.66%]	±204.03 [47.68%]	±20.27 [50.00%]	±0.58 [69.24%]
mu_ttbar_Gtt_1900_1400	±5.63 [21.7%]	±87.20 [20.4%]	±8.69 [21.4%]	±0.17 [20.7%]
gamma_stat_CR_Gtt_1900_1400_cuts_bin_0	±1.44 [5.5%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_topEW_syst_CR_Gtt_1900_1400	±0.83 [3.2%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JER4	±0.80 [3.1%]	±16.79 [3.9%]	±3.05 [7.5%]	±0.11 [12.6%]
alpha_JER6	±0.71 [2.7%]	±11.17 [2.6%]	±0.43 [1.1%]	±0.07 [7.7%]
alpha_QCDHundred	±0.62 [2.4%]	±0.83 [0.19%]	±0.05 [0.12%]	±0.05 [5.6%]
alpha_JER1	±0.54 [2.1%]	±22.43 [5.2%]	±3.86 [9.5%]	±0.23 [26.9%]
alpha_WZ_ckkw_syst	±0.52 [2.0%]	±9.99 [2.3%]	±1.53 [3.8%]	±0.00 [0.00%]
alpha_JER0	±0.46 [1.8%]	±0.36 [0.08%]	±2.64 [6.5%]	±0.05 [5.8%]
alpha_JER3	±0.41 [1.6%]	±4.35 [1.0%]	±2.56 [6.3%]	±0.09 [10.5%]
alpha_JER2	±0.36 [1.4%]	±28.85 [6.7%]	±0.39 [0.96%]	±0.13 [15.1%]
alpha_JER5	±0.31 [1.2%]	±19.38 [4.5%]	±0.93 [2.3%]	±0.02 [2.8%]
alpha_JER7	±0.30 [1.2%]	±14.13 [3.3%]	±1.62 [4.0%]	±0.09 [10.3%]
alpha_JES1	±0.30 [1.1%]	±29.64 [6.9%]	±3.08 [7.6%]	±0.09 [11.2%]
alpha_kin_RW	±0.28 [1.1%]	±6.88 [1.6%]	±0.67 [1.6%]	±0.03 [3.0%]
alpha_WZ_fac_syst	±0.24 [0.93%]	±2.39 [0.56%]	±0.71 [1.8%]	±0.00 [0.00%]
alpha_JES4	±0.15 [0.57%]	±1.32 [0.31%]	±0.23 [0.57%]	±0.00 [0.53%]
Lumi	±0.12 [0.44%]	±2.52 [0.59%]	±0.19 [0.47%]	±0.00 [0.55%]
alpha_JES2	±0.10 [0.39%]	±4.54 [1.1%]	±0.90 [2.2%]	±0.02 [2.6%]
alpha_WZ_renorm_syst	±0.10 [0.37%]	±6.44 [1.5%]	±0.26 [0.64%]	±0.00 [0.00%]
alpha_bTag_C	±0.09 [0.36%]	±1.13 [0.26%]	±0.16 [0.38%]	±0.02 [2.5%]
alpha_bTag_L	±0.09 [0.33%]	±1.60 [0.37%]	±0.40 [0.99%]	±0.03 [3.8%]
alpha_bTag_B	±0.07 [0.28%]	±1.28 [0.30%]	±0.16 [0.39%]	±0.00 [0.08%]
alpha_JES6	±0.03 [0.11%]	±7.26 [1.7%]	±1.55 [3.8%]	±0.00 [0.42%]
alpha_JVT	±0.02 [0.09%]	±1.33 [0.31%]	±0.12 [0.29%]	±0.00 [0.47%]
alpha_JES0	±0.02 [0.08%]	±19.29 [4.5%]	±2.37 [5.8%]	±0.12 [14.1%]
alpha_bTag_extrapol_charm	±0.01 [0.04%]	±1.94 [0.45%]	±0.35 [0.86%]	±0.01 [1.4%]
alpha_WZ_qsf_syst	±0.01 [0.04%]	±1.88 [0.44%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JES5	±0.01 [0.03%]	±0.10 [0.02%]	±0.03 [0.08%]	±0.00 [0.01%]
alpha_bTag_extrapol	±0.01 [0.03%]	±0.56 [0.13%]	±0.06 [0.15%]	±0.00 [0.27%]
alpha_JES3	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_st_syst_SR_Gtt_1900_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.45%]
alpha_ttbb_syst	±0.00 [0.00%]	±1.23 [0.29%]	±0.44 [1.1%]	±0.04 [4.3%]
alpha_topEW_syst_VR2_Gtt_1900_1400	±0.00 [0.00%]	±0.00 [0.00%]	±1.45 [3.6%]	±0.00 [0.00%]
alpha_topEW_syst_SR_Gtt_1900_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.08 [9.0%]
alpha_st_syst_VR1_Gtt_1900_1400	±0.00 [0.00%]	±11.53 [2.7%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttcc_syst	±0.00 [0.00%]	±5.81 [1.4%]	±1.34 [3.3%]	±0.03 [3.3%]
alpha_ttbar_syst_VR2_Gtt_1900_1400	±0.00 [0.00%]	±0.00 [0.00%]	±16.96 [41.8%]	±0.00 [0.00%]
alpha_diboson_syst_VR1_Gtt_1900_1400	±0.00 [0.00%]	±1.71 [0.40%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ttbar_syst_SR_Gtt_1900_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.34 [40.5%]
alpha_st_syst_VR2_Gtt_1900_1400	±0.00 [0.00%]	±0.00 [0.00%]	±1.11 [2.7%]	±0.00 [0.00%]
alpha_diboson_syst_VR2_Gtt_1900_1400	±0.00 [0.00%]	±0.00 [0.00%]	±0.05 [0.12%]	±0.00 [0.00%]
alpha_topEW_syst_VR1_Gtt_1900_1400	±0.00 [0.00%]	±17.93 [4.2%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_VR2_Gtt_1900_1400_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±2.15 [5.3%]	±0.00 [0.00%]
gamma_stat_SR_Gtt_1900_1400_cuts_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.23 [27.3%]
alpha_ttbar_syst_VR1_Gtt_1900_1400	±0.00 [0.00%]	±170.25 [39.8%]	±0.00 [0.00%]	±0.00 [0.00%]

Table 23 – Breakdown of the dominant systematic uncertainties on background estimates for region Gtt_1900_1400