University of Glasgow | Experimental Particle Physics

The Standard Model Higgs Boson Produced in Association with a t-Quark Pair and Decaying to a b-Quark pair from LHC Run-2 proton-proton 13 TeV Collisions in ATLAS

William Dmitri Morgan Breaden Madden

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

Principal Supervisor: Professor Anthony Terence Doyle
Second Supervisor: Professor Craig Buttar

School of Physics and Astronomy
College of Science and Engineering
University of Glasgow

2022

Abstract

Discovery of the Standard Model Higgs boson in 2012 by the ATLAS and CMS Collaborations at the LHC was an extraordinary moment in the advancement of science. In this thesis the author presents the theory, methodology and results of a search for the Standard Model Higgs boson produced in association with a t-quark pair and decaying to a b-quark pair as it was published in 2018 [1]. This search lead to evidence for $t\bar{t}H$ and subsequent discovery. The data used are from proton-proton (pp) collisions in A Toroidal LHC ApparatuS (ATLAS) at CERN during Run-2 of the Large Hadron Collider (LHC) at centre-of-mass energy $\sqrt{s} = 13\,\text{TeV}$ corresponding to an integrated luminosity of $36.1 \pm 0.8\,\text{fb}^{-1}$. Events were selected to contain one electron or one muon from the t-quark decays and the events were then categorised according to the number of jets counted in the event and the number of those jets which were identified as likely b-jets. Many backgrounds were considered in the analysis and all are dominated by $t\bar{t}+\text{jets}$ production. An analysis framework was developed and validated and was successful in the analysis at finding evidence for Higgs bosons in $t\bar{t}H\left(b\bar{b}\right)$, as presented in this thesis, and was subsequently used successfully in observation of $t\bar{t}H$. Multivariate methods were used to discriminate between signal and background events, and this thesis presents both the standard analysis methods and newer multivariate methods of a more tentative nature used for classification of events as signal $t\bar{t}H\left(b\bar{b}\right)$ and background $t\bar{t}b\bar{b}$ in order to investigate the efficacy of these methods. Novel techniques were developed for interpretation of training data and for improvement of both classification efficiency and interpretability of modelling for the purposes of classification.

A highly-parameterised inception convolutional neural network (ICNN) model was compared with the standard analysis boosted decision tree (BDT) model for classification of events as signal and background. The ICNN was observed to classify a fraction of $0.7006 \pm 0.0097\,(\text{stat.})^{+0.0033}_{-0.0028}\,(\text{syst.})$ events correctly while the BDT was observed to classify a fraction of $0.6979 \pm 0.0136\,(\text{stat.})^{+0.0027}_{-0.0031}\,(\text{syst.})$ events correctly. Further, the ICNN and other newer models also made accessible alternative and additional insights and interpretations of their respectively achieved classification modelling. These included the techniques of mean saliency mapping and activation maximisation. Saliency mapping provided not only a means of ranking features in data on a per-class basis, but also a means of providing saliency measures for individual

features in data on a per-event basis, something not obviously possible using conventional models such as decision trees.

In the full combined analysis, for a Standard Model Higgs boson of 125 GeV, the ratio of the observed signal $t\bar{t}H$ cross-section $\sigma_{t\bar{t}H}^{\text{obs.}}$ to the cross-section expectation of the Standard Model $\sigma_{t\bar{t}H}^{\text{SM}}$, $\mu_{t\bar{t}H} = \sigma_{t\bar{t}H}^{\text{obs.}}/\sigma_{t\bar{t}H}^{\text{SM}}$ (signal strength), was found to be

$$\mu_{t\bar{t}H} = 0.84^{+0.64}_{-0.61},$$

with signal strength greater than 2 excluded at the 95 % confidence level and the expected exclusion limit in the absence of signal being 1.2. A 1.4σ excess above the expected background was observed, while an excess of 1.6σ was expected with the existence of the Standard Model Higgs boson.

Evidence of $t\bar{t}H$ was declared in 2018 [1]. The $t\bar{t}H\left(b\bar{b}\right)$ analysis, one of the most complicated ATLAS analyses to date, led to observation of the $t\bar{t}H$ process, enabling direct investigation of the mechanism of electroweak symmetry breaking and the matter-Higgs relationships predicted by electroweak symmetry breaking in the Standard Model. The observation of the $t\bar{t}H$ process used data corresponding to an integrated luminosity of $79.8\,\text{fb}^{-1}$ and a combination of results from Higgs boson decays to $b\bar{b}$, WW^*, $\tau^+\tau^-$, $\gamma\gamma$ and ZZ^* with an observed significance of 5.8σ, compared to an expected significance of 4.9σ [2, 3]. Combined with $t\bar{t}H$ search results using data recorded at centre-of-mass energies of $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$, the observed significance was 6.3σ, compared to an expected significance of 5.1σ. Assuming Standard Model branching fractions, the $t\bar{t}H$ production cross-section at $\sqrt{s} = 13\,\text{TeV}$ was measured as $670 \pm 90\,(\text{stat.})^{+110}_{-100}\,(\text{syst.})\,\text{fb}$, which is in agreement with the Standard Model theoretical prediction of $507^{+35}_{-50}\,\text{fb}$.

Declaration

The research presented in this thesis is the product of the author's own work. Appropriate references are provided when results of third parties are mentioned. The research presented here was not submitted for another degree in any other department or university.

Acknowledgements

I especially thank Tony Doyle, my principal supervisor, for his dedicated academic guidance, encouragement, assistance, patience, and for helpful conversations during my studies. I also thank Craig Buttar for his assistance and support in his capacity as my second supervisor. I am grateful to Andrea Knue and Sarah Boutle for academic guidance, support and detailed technical guidance on the $t\bar{t}H\left(b\bar{b}\right)$ semileptonic analysis. I thank Graeme Stewart for guidance on ATLAS authorship qualification and both Saverio D'Auria and Kerstin Lantzsch for guidance on Inner Detector ATLAS Control Room shift work and thoughts on IBL commissioning. Special mention must be made of my former supervisor Rick St. Denis who died during the course of this Ph.D. His guidance, insight and humour will not be forgotten.

I acknowledge outstanding comrades Sam Crawley, Azzah Al-Shehri, Johnny Raine, Ben Sowden, Diane Cinca, Rachik Soualah, and other members of the HTop group, including those who devised and validated the analysis and came up with ways to squeeze every last bit of information from data recorded, and the TTHbbLeptonic development team who made the Run-2 $t\bar{t}H\left(b\bar{b}\right)$ analysis framework (and Lukas Heinrich for collaboration in containerisation of same), which helped lead to significant advancement in fundamental physics.

I acknowledge Carmel Breaden, Liam Madden, Tony Doyle, Michael Fenton, Liam Moore, Stephen Ogilvy and Gavin Kirby for thesis comments and suggestions, Paul Soler and David Colling for thesis assessment, and Liam Moore for advice on computing hardware.

I acknowledge guides and colleagues Chris Collins-Tooth, Mark Owen, David Miller, Andrew Lytle, Christoph Englert, Deepak Kar, Umar Gul, Andy Buckley, John Nugent, Michael Alexander, Dima Maneuski, Chris Bouchard, Val O'Shea, Ian Connelly, James Ferrando, Chris White, Paul Soler, Chris Pollard, Federica Fabbri, Giuseppe Callea and Aidan Robson, and I salute comrades John B., Dan S., Bijan H., Anna D., Cameron D., Fionn B., Fiona-Rose S., Patrik H., Tim E., Stephen J., Michael F., Chloe G., Xanthe H., Emma B., Ellis K., Rob H., Ben S., Euan Mcl., Dan H., Dan S., Harry M., Iwan S., Greg F.-S., Diane C., Anita N., Duncan L., Rose A., Dwayne S., Jonathan J., Morag W., Murdo T., Donatas Z., Donny Q., Michael W., Karl N., Sarah K., John McD., Abby P., Roddy McC., Nicolas G., Paul M., Becky D., Iain L., Rick F., Valerie F., Brian C., Rachel D., Bipasha C., Lies M., Tom Q., Euan N. C., Jon B., Tom R., Nick E., Sean L., Will F., John A., Tom B., Joe E., Stan T., Bruno B., Adam R., David F., Laurynas M., Charlie R., Puwen S., Elien M., Dana S. B., Peter S., Hiu F.-W., Bill H., Yann LeC., Pedro Carolino, Glasgow PPE and PPT, the LTA Self-Annihilation Working Group, and all others too numerous to list. CERN is to be commended for the achievements of the LHC, as is everyone involved in the ATLAS Collaboration [4]. For general support, opportunities and financial support I acknowledge my family, the STFC, taxpayers, Glasgow PPE, CERN, anyone who helped, and Bitcoin volatility.

In particular, Gavin Kirby for being a standardised guiding light, Stephen Ogilvy for an arm around a shoulder, Lauren Douglas for team spirit, Liam Moore for smoke on the water of Montreux – thank you.

My family have supported me through this journey – I couldn't have asked for better; Carmel, Liam, Gerard – go raibh míle maith agaibh! I love you all!

When all else fails manipulate the data.

Quote from a Rick St. Denis jumper

Figure 1: Some members of the TTHbbLeptonic development team.

Figure 2: Some members of the ATLAS HTop Higgs sub-group.

Figure 3: Some members of the School of Physics and Astronomy at the University of Glasgow.

For Carmel, Liam and Gerard.

I love you all dearly.

Contents

Chapter 1

Preface

> The gross and net result of it is that people who spent most of their natural lives riding iron bicycles over the rocky roadsteads of this parish get their personalities mixed up with the personalities of their bicycle as a result of the interchanging of the atoms of each of them and you would be surprised at the number of people in these parts who nearly are half people and half bicycles.
>
> Sergeant Pluck in *The Third Policeman* (1940)

The Higgs boson was the last particle predicted by the Standard Model of particle physics to be discovered [5, 6, 7, 8, 9, 10]. It was observed in 2012 by the ATLAS and CMS collaborations [11, 12]. Since then, there has been intense research into its characteristics, particularly its couplings, which serve as a test of the precision of Standard Model predictions [13, 14, 15]. While significant progress has been made on understanding the Higgs boson, further efforts are underway in analysis, in developments of new techniques for analysis and computation, and in upgrades to the LHC and ATLAS in preparation for Run-3, in order to measure properties of the Higgs boson more precisely, and in general to further test predictions of theories of particle physics [16].

1.1 Contributions and publications

This thesis takes a snapshot of the search for $t\bar{t}H$ in ATLAS using LHC Run-2 proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$ corresponding to an integrated luminosity of $36.1 \pm 0.8\,\text{fb}^{-1}$, recorded in 2015 and 2016 and processed by the $t\bar{t}H\left(b\bar{b}\right)$ analysis infrastructure. Contributions were made by the author as part of the ATLAS Collaboration, the ATLAS HTop Higgs sub-group and the Experimental Particle Physics research group of the School of Physics and Astronomy at the University of Glasgow. The main search of this analysis was approved by the ATLAS Collaboration and was published in *Physical Review Letters D* in 2018 [1] (earlier work having been contributed to Reference [17]).

The author created and led development and coordination of the combined $t\bar{t}H\left(b\bar{b}\right)$ analysis derivation framework TTHbbLeptonic. The author was contact person for the analysis framework and directed merging of various related analyses and tools into the analysis framework, with various physics and framework validations and cut-flow checks of the infrastructure. The TTHbbLeptonic development team worked sufficiently closely that attributing its achievements in detail to individuals would not be practical, however an internal note details contributions by the author to the physics analysis and to lead development of the analysis derivation framework and its physics validations, and is referenced for the purposes of ATLAS Collaboration members [18]. The work of the author also included framework virtualisation, contributions to the fitting framework TRExFitter, development of various validated tools and prototype tools, such as grid production code and a prototype tag-and-probe lepton fakes method, which led to a CERN Note [19].

The author developed, wrote, investigated and validated a variety of modern multivariate methods, within a developed framework referred to as Abstraction, for classification of signal and background in high-energy physics analyses, and also novel techniques for interpretation of classification modelling achieved and of input data. The multivariate methods and the novel techniques of interpretation were investigated and validated using "The SUSY Data Set" and were then applied to data of the $t\bar{t}H\left(b\bar{b}\right)$ analysis [20].

The author contributed to the commissioning of the Insertable B-Layer of ATLAS (the innermost layer of the detector) as part of ATLAS Control Room shift work, and developed and provided support for new grid computing software for parallel ATLAS data transformations as part of authorship qualification tasks, which were published in *Journal of Physics: Conference Series* [21].

In addition, the author contributed to other related projects. The author provided guidance on development of an online interface for the control system of the nuclear reactor of the SoLid experiment (Search for Oscillations with Lithium-6 detector at the SCK-CEN BR2 reactor), an experiment involved in investigation of hypothetical sterile neutrinos, collaborating with Daniel Saunders [22, 23, 24]. The author

assisted Johnny Raine with TDAQ HLT online monitoring work. The author assisted Gavin Kirby with limit setting for a Run-1 $t\bar{t}H\left(b\bar{b}\right)$ analysis [25]. The author engaged in outreach as part of an exhibit of the Large Hadron Collider at the Scottish Parliament [26]. At the request of the ATLAS spokesperson, the author also engaged with representatives of the Irish government in order to advance of Ireland's membership of CERN. Joining CERN is now the endorsed position of Oireachtas Éireann, and welcome advancements towards membership are underway [27, 28, 29].

A visual representation of author contributions to analysis infrastructure and multivariate investigations is shown in Figure 1.1.

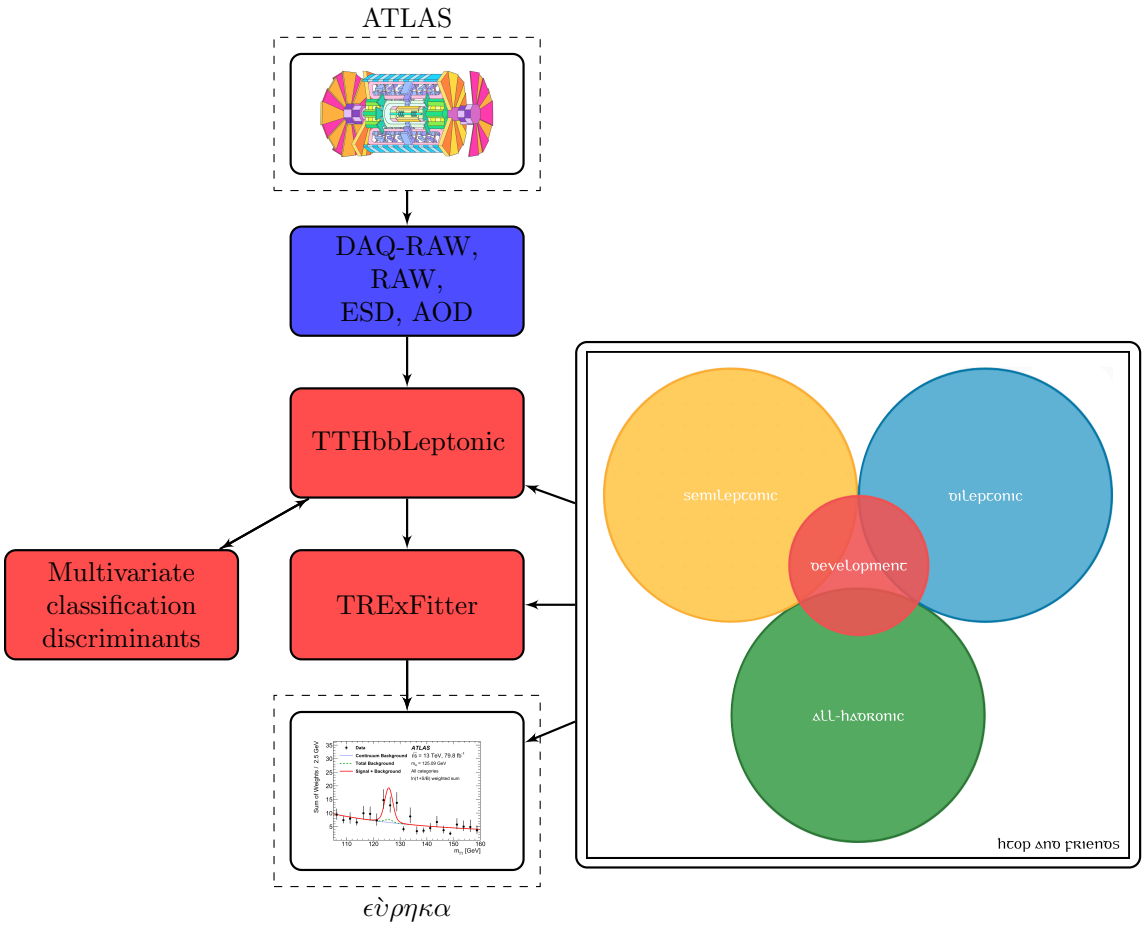


Figure 1.1: Contributions overview [30, 2]. Data flows from ATLAS through various computational infrastructures, ultimately to be used to compare data observed with theoretical predictions. Significant contributions by the author were made particularly to the semileptonic $t\bar{t}H$ $(b\bar{b})$ analysis group (shown in yellow), to the HTop software development group (shown in red), to the TTHbbLeptonic framework (shown in red) and to TRExFitter (shown in red). Significant contributions developed solely by the author to multivariate classification discriminants are shown also (in red).

TTHbbLeptonic was developed by a team of very active contributors. For a video animation of the development of TTHbbLeptonic through 2015, the reader may view Reference [31]. In this video, the author is represented as a red icon labelled 'wbreaden', the root directory is shown at centre, directories are represented by line branches, files are represented by circle leaves, and brief projections from user icons to files and directories represent repository commit actions.

The $t\bar{t}H$ analysis is one of the most challenging analyses of the LHC physics programme as $t\bar{t}H$ production is only a very tiny, and very hard to observe, fraction of the Higgs production of the LHC – approximately 1 %, a very rare process – so, in order to observe the process, it was necessary to record vast datasets, to combine many decay channels and to unite the efforts of hundreds of people working in physics analyses, and thousands more people involved with the LHC besides.

The findings of the main $t\bar{t}H\left(b\bar{b}\right)$ analysis of this thesis were approved and published by the ATLAS Collaboration in 2018 [1, 32]. The analysis was subsequently combined with three other analyses of $t\bar{t}H$ using the same 2015–2016 Run-2 dataset, establishing evidence for the existence for the associated production of the Higgs boson and a t-quark pair [1, 33]. Corresponding findings for a $t\bar{t}H\left(b\bar{b}\right)$ analysis were published by the CMS Collaboration also in 2018 [34].

For the last few decades, the hypothesis of the Standard Model for the origin of quark masses is that the masses arise through the interaction of quarks with the Higgs field. While the discovery of the Higgs boson in 2012 was a major advancement, the observation, with a statistical significance of more than 5σ, of the *direct* coupling of the Higgs boson to the t-quark, the most massive particle of the Standard Model, is a particularly important advancement in high-energy physics because it effectively establishes *directly* that this hypothesis is accurate. The observation of $t\bar{t}H$ is notable also as being the observation of the interaction of the two most massive fundamental particles known.

Due to the nature of contemporary high-energy physics collaborations, the results presented in this thesis rely on the combined efforts of many thousands of people, and on the work of others that help to support those efforts. The contributions to finding evidence for $t\bar{t}H$, which led to its subsequent discovery, and the contributions listed above, including the multivariate methods for classification and modelling and techniques for interpretation of modelling, are original contributions by the author to the scientific knowledge base. The author emphasises that the ideas in this thesis and many of its referenced publications should be seen as hopeful, speculative, and constrained by the limitations of their time.

1.2 Structure of thesis

Chapter 2 is an introduction to the theory and motivation for the $t\bar{t}H\left(b\bar{b}\right)$ analysis, providing a broad overview of the Standard Model, electroweak symmetry breaking, the Higgs boson, the importance of $t\bar{t}H$ in probing the t-Higgs Yukawa coupling, with details on past searches, measurements and analyses, the current experimental status of the Higgs boson, and production processes and decay modes of the Higgs boson in the LHC.

Chapter 3 is a description of the experimental apparatus of the LHC and A Toroidal LHC ApparatuS (ATLAS), the various sub-detectors and components of ATLAS, and data propagation from the detector to the global computing infrastructure used for the analysis, with details on the networks of humans in the ATLAS Collaboration, commissioning of the Insertable B-Layer (IBL) (the innermost detection layer of ATLAS), and grid-based parallel data processing (which was the ATLAS authorship qualification task of the author).

Chapter 4 is an overview of simulation of physics events in ATLAS, as predicted by the Standard Model, and reconstruction of various physics objects, such as leptons and jets, which are used to compare data recorded by ATLAS with simulated data.

Chapter 5 details the analysis strategy and methodology in the search for $t\bar{t}H\left(b\bar{b}\right)$ using a snapshot of data corresponding to an integrated luminosity of $36.1 \pm 0.8\,\text{fb}^{-1}$ recorded by ATLAS in Run-2.

Chapter 6 is a description of the structure, development and continuous validation and testing of the computing software framework used in the analysis.

Chapter 7 presents the results of the search for $t\bar{t}H\left(b\bar{b}\right)$.

Chapter 8 opens with an overview of multivariate methods in the analysis for physics event classification, including decision trees and a variety of neural network models, including some models under recent and current research and development. Also presented are some techniques for interpreting what is modelled by such multivariate methods. The chapter then details the application of the multivariate methods and techniques of interpretation thereof to the classification of the signal $t\bar{t}H\left(b\bar{b}\right)$ and background $t\bar{t}b\bar{b}$. Results are presented and compared, and a variety of interpretations of the data and the modelling is provided.

Chapter 9 comprises conclusions, and summarises the results of the search for the Standard Model Higgs boson in the $t\bar{t}H\left(b\bar{b}\right)$ analysis, the success of the analysis framework, and details on the various new multivariate studies.

Chapter 10 is an outlook, commenting on plans, speculations, educated guesses and challenges for the

future of the exploration of particle physics.

A series of three Institute of Physics (IOP) talks by the author, Ben Sowden and Sam Crawley, respectively, serves as an executive summary of the $t\bar{t}H\left(b\bar{b}\right)$ analysis, a snapshot within its historical context, and can be viewed at References [35, 36, 37].

Chapter 2

Theoretical framework and motivation: Standard Model Higgs bosons in $t\bar{t}H$ $(b\bar{b})$

> When the laws are obviously in the way of scientific development, they should be canceled.
>
> Alexandra Elbakyan

This chapter provides an overview of some terminology, the Standard Model, the Higgs field, the Higgs boson, and details on relevant searches. An effort has been made to keep the overview within the constraints of the perspective of an experimentalist. A full pedagogical introduction to quantum field theories and the Standard Model of particle physics is provided by the classic Peskin and Schroeder (which has informed various derivations described in this chapter), Reference [38]. An introduction to mathematics relevant to quantum field theories, such as group theory, is provided by Georgi, Reference [39]. A general reference and extensive overview of mathematics is provided by *The Princeton Companion to Mathematics*, Reference [40].

2.1 Overview of the Standard Model of particle physics

2.1.1 Quantum chromodynamics

A successful description of most apparent aspects of hadronic final states in high-energy particle collision processes over a wide range of centre-of-mass energies is provided by quantum chromodynamics (QCD) [41]. The force is strong in this one, with QCD providing the theory of strong interactions of

quarks and gluons. A coupling constant (or gauge coupling parameter) is a number that expresses the strength of the force exerted in an interaction. In QCD, the strong coupling constant α_s, the only free parameter in the theory, can be measured to an accuracy of a few percent [42]. A variety of measurements of α_s are shown in Figure 2.1.

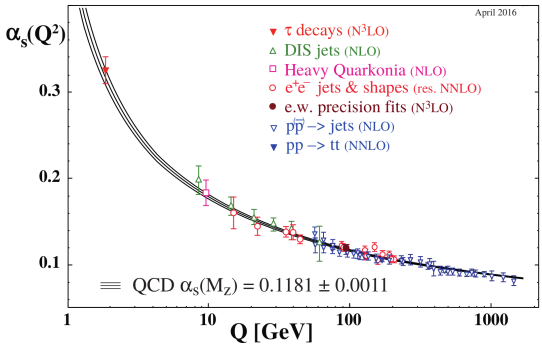


Figure 2.1: Summary of measurements of the strong coupling constant α_s as a function of energy scale Q, with the respective degree of the QCD perturbation theory used in the extraction of the value indicated in brackets (NLO: next-to-leading-order; NNLO: next-to-next-to-leading-order; N^3LO: next-to-NNLO) [43].

Quantum electrodynamics (QED), the theory of the electron and the photon, was the first major example of a successful quantum field theory. In QED, electrically charged fermion particles interact via the exchange of electrically neutral photons, which are the manifestations of local $U(1)_Y$ gauge symmetry. QCD is an analogous theory with the crucial difference that the gauge symmetry group is the non-abelian (non-commutative) group $SU(3)_C$, and belongs to a general family of $SU(N)_L$ gauge theories referred to as Yang-Mills theories. Here, particles carry a type of charge referred to as colour charge, and gauge bosons referred to as "gluons" (analogous to photons) themselves carry this charge. Standard QCD features three distinct colour charges carried by quarks which obey the $SU(3)_C$ symmetry group [44]. Requiring local gauge invariance of the theory generates eight massless vector fields, which correspond to the gluons, with each carrying a colour and an anticolour charge. As the gluons are colour-charged, QCD has more fundamental vertices than QED, for which the gauge boson, the photon, is electrically neutral. The fundamental QCD vertices are

- the radiation of a gluon from a quark line (gluon bremsstrahlung),
- the interaction of a gluon with a quark-antiquark pair (quark pair production),
- the pointwise interaction of three gluons (triple gluon vertex) and
- the pointwise interaction of two gluons (quartic gluon vertex).

2.1.2 The Standard Model

The Standard Model of particle physics (SM) is a quantum field theory based on gauge symmetries that describe fundamental particles and their interactions. The Standard Model is a gauge theory, so the Lagrangian of the theory is invariant locally as well as globally. The Lagrangian dictates the equations of motion of a system, and invariance of this under transformations gives rise to conserved quantities. As the Standard Model is a quantum field theory, its fundamental constituent mathematical objects are quantum fields which are defined at all points in space-time, specifically being operator-valued and acting on quantum states. The Standard Model quantum fields are the fermion field ψ, the electroweak boson fields W_1, W_2, W_3 and B, the gluon field G_a and the Higgs field ϕ. The Standard Model fundamental particles, which are excitations of its constituent quantum fields, consist of matter particles called fermions, interaction mediators called bosons, and the Higgs boson, which is the smallest possible excitation of the Higgs field. Together the standard models of particle physics and cosmology appear to require around 31 to 37 dimensionless physical free parameters [45].

In the Standard Model, there are a number of interactions which are treated as distinct and not unified, though theorists are often of the view that the interactions are likely to be different manifestations of the same underlying interaction. Particles with an electric charge are influenced by the electromagnetic interaction (QED), particles with a weak charge are influenced by the weak interaction (quantum flavour dynamics, QFT) and particles with a colour charge are influenced by the strong interaction (QCD). The Higgs field is scalar and so induces spontaneous symmetry-breaking, which gives mass to particles with which it interacts (without introducing unworkable infinities into calculations of quantum flavour dynamics), and the Higgs boson couples to massive particles, including itself [46]. The character of the Standard Model is probabilistic; it provides the rates at which processes involving the fundamental particles occur. It is a candidate for being the most consistent and precisely predictive theory of reality known, and one of the most thoroughly tested, indeed to the degree that it has been highly successful at describing physical phenomena accessible to collider experiments. It has been shown to describe observed phenomena to a high degree of precision. It predicted the existence of the t-quark, which was discovered by the CDF and DØ collaborations at the Tevatron proton-antiproton ($p\bar{p}$) collider in 1995, the tauon-neutrino, which was discovered by the DONUT Collaboration in 2000, and the Higgs boson, which was discovered by the ATLAS and CMS collaborations at the LHC in 2012 [47, 48, 49, 11, 12].

The Standard Model is based on the gauge group

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

where C is colour, L is left-handed isospin and Y is hypercharge. By Noether's theorem, each symmetry

has a corresponding conserved quantity. In a formal sentence, the Standard Model may be described as the simplest four-dimensional low energy quantum effective field theory description consistent with the known degrees of freedom and their interactions (except gravity) and all experimental data.

The three constituent groups of the Standard Model gauge group correspond to the three fundamental interactions described by the Standard Model. The non-abelian symmetry group $\mathrm{SU}(3)_{\mathrm{C}}$ describes the strong interaction formalised in QCD [38]. The $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$ part of the gauge group describes the unified electroweak interactions, known as the Glashow-Salam-Weinberg theory, and this is broken to $\mathrm{U}(1)_{\mathrm{Y}}$ electromagnetism via the Brout-Englert-Higgs-Guralnik-Hagen-Kibble (BEHGHK) mechanism [38]. The fermionic matter fields can be divided into two classes. The quarks u_R, d_R, and q_L, are distinguished by their transforming under the strong gauge group $\mathrm{SU}(3)_{\mathrm{C}}$, and thus interact with gluons. These also transform under the symmetry $\mathrm{U}(1)_{\mathrm{Y}}$ mediated by the B boson of weak hypercharge, while the latter additionally transforms under $\mathrm{SU}(2)_{\mathrm{L}}$ mediated by the W_1, W_2 and W_3 bosons of weak isospin. The lepton fields e_R and l_L are also both charged under $\mathrm{U}(1)_{\mathrm{Y}}$, while the latter transforms also under $\mathrm{SU}(2)_{\mathrm{L}}$. The lone scalar Higgs field $V(\phi)$ transforms under both $\mathrm{SU}(2)_{\mathrm{L}}$ and $\mathrm{U}(1)_{\mathrm{Y}}$. The fermion field ψ transforms under these symmetries, although all of them leave some parts of it unchanged.

The Standard Model is remarkably successful and also glaringly incomplete. It contains a large number of kinds of elementary particles and a number of interactions between them, but provides no explanation for all that variety. It features a number of apparently arbitrary constants describing the interactions, including constants that yield masses for the different kinds of particles, and theorists often find it hard to accept as a fundamental theory given this number of important numbers apparently incalculable in principle. Gravity appears to act on all known particles, however gravitational interactions are not included in the Standard Model as it is not known how to incorporate a quantum theory of gravitation into the Standard Model. Some theories beyond the Standard Model postulate an exchange particle for this interaction called the graviton. While theoretical bridges between gravitation and quantum physics have been developed, such as Hawking radiation, there are no consistent theories that combine general relativity and quantum mechanics that are widely accepted. Renormalisation is a collection of techniques used in quantum field theory that are used to address infinities which arise in calculated quantities by altering the values of these quantities in order to compensate for the effects of their self-interactions. Theories that attempt to combine the graviton with the strong and electroweak interactions lead to fundamental difficulties and are not renormalisable. In the context of the high-energy physics of this thesis, gravitational interactions are considered negligible at the energies encountered normally in the current generation of particle colliders, and it is possible that collision energies corresponding to the Planck mass would be required in order to probe gravity using particle colliders of a familiar form.

Particles of the Standard Model are classified by spin, which may be described as an intrinsic angular

momentum of particles. The spin of a particle is used to classify it into fermionic or bosonic matter. Fermions have half-integer spin and are described by Fermi-Dirac statistics and bosons have integer spin and are described by Bose-Einstein statistics.

Interactions of the Standard Model are considered in terms of the quantum fields corresponding to the half-integer spin fermions with the integer spin bosons. The fermions are comprised of the colourless leptons and the colour charged quarks, and are categorised into three generations, with each generation having identical properties except for fermion mass. Bosons mediate the fundamental interactions of the Standard Model. The photon, the $W^{\pm}$ bosons, the Z^0-boson, the eight gluons and the Higgs boson make up the thirteen bosons of the Standard Model. Within the current Standard Model, electrons and muons have identical interaction strengths, however recent hints from the Large Hadron Collider Beauty (LHCb) experiment suggest that the decay rate of B-mesons to $\mu^+\mu^-$ is not equal to the decay rate to e^+e^-, and this might mean that lepton flavour universality is undermined – perhaps a hint of new, beyond the Standard Model physics [50]. There has been some recent apparent evidence from the Muon $g-2$ Experiment at Fermilab that the muon is sensitive to something beyond the Standard Model, specifically via the measurement of the anomalous magnetic dipole moment of the muon [51].

Standard Model fundamental particles are summarised in Figure 2.2.

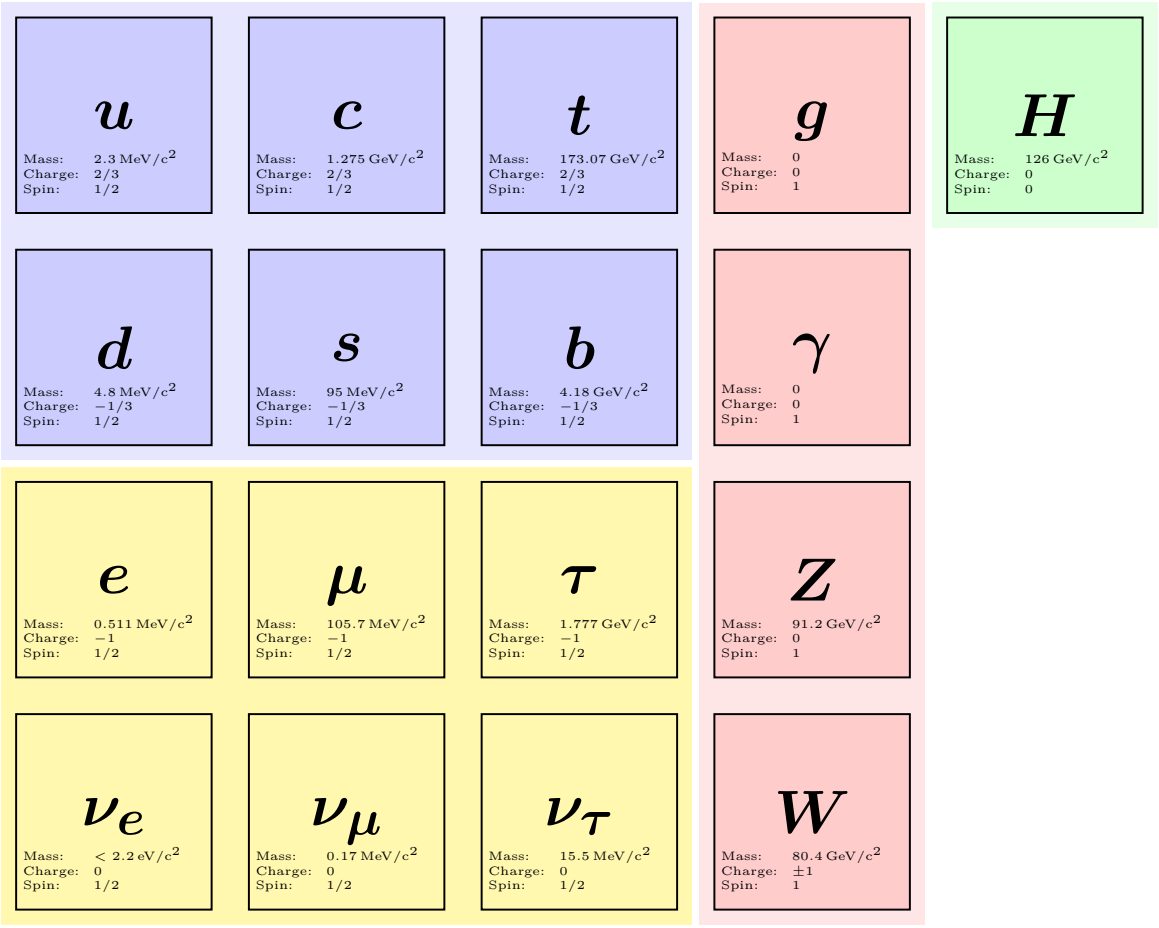


Figure 2.2: Standard Model fundamental particles in their mass eigenstates. Three fermion matter generations are distinguished by column at left, with quarks in blue and leptons in yellow, then gauge bosons are in red and then the Higgs boson is in green at right. This visualisation was composed with reference to the Particle Data Group review of 2018 [43].

As mentioned, bosons mediate all of the Standard Model interactions. The photon γ mediates the electromagnetic interactions between electrically charged particles, the $W^{\pm}$ and Z^0 bosons mediate the weak charged and weak neutral interactions between particles of different flavours (quarks and leptons) and the gluons mediate the strong interactions between colour-charged particles (quarks). Bosons with non-zero spin are termed vector gauge bosons, which include the photons, the $W^{\pm}$ and Z^0 bosons, and gluons. The only boson with zero spin (a scalar boson) is the Higgs boson. In a sense, this makes it the simplest particle of the Standard Model (the half-integer spin fields are described by the Dirac equation

and the integer spin fields are described by the Maxwell or Yang-Mills equations, while spin zero fields are described by the comparably simple Klein-Gordon equation), and also the way by which it interacts with other fields is most intricate. The Higgs boson was postulated in an attempt to theoretically reconcile the existence of the massless photon with the massive $W^{\pm}$ and Z^0 bosons. Since weak interactions are of short range and are very weak at low energies, it was clear that if they were mediated by an intermediate boson $W^{\pm}$ that it must have have a large mass.

> This summer I have discovered something totally useless.
>
> Peter Higgs [52]

A mechanism was required to give mass to the $W^{\pm}$ bosons while leaving the photon massless. In analogy with QED, an electroweak theory might be expected to be locally gauge invariant and to contain massless gauge bosons analogous to the photon. The relatively large masses of the $W^{\pm}$ and Z^0 are accounted for, together with the scalar Higgs boson, by this mechanism. It arose from work by Anderson[1], Brout, Englert, Higgs, Guralnik, Hagen and Kibble [54, 5, 6, 7, 8]. This BEHGHK mechanism could be described as essentially a relativistic version of a mechanism that operates in a superconductor that was described by Anderson [54]. The Higgs mechanism gives mass to fermions when they interact with the Higgs field generated by spontaneous symmetry breaking. Spontaneous symmetry breaking occurs when the Lagrangian of a system is invariant under the symmetry group, but the ground state (or vacuum state) breaks the symmetry. The mass values remain arbitrary parameters, as the masses are determined by coupling constants (the t-quark is very massive because it interacts strongly with the Higgs field). Mass in normal, tangible matter arises from protons and neutrons, which are given mass from the binding energy of quarks and gluons in QCD. The Higgs boson is so called because Higgs was the first to explicitly express that the BEHGHK mechanism necessitated a massive scalar particle: "... an essential feature of [this] type of theory ... is the prediction of incomplete multiplets of scalar and vector bosons" [7].

The photon is the mediator of the electromagnetic interaction and is massless and chargeless. The electric charge of a particle is the magnitude of the coupling that the particle has to the electromagnetic interaction. The $W^{\pm}$ and Z^0 bosons are the mediators of the weak interaction. The weak isospin of a particle is the magnitude of the coupling that the particle has to the weak interaction. Left chiral u-type quarks and neutrinos have $+\frac{1}{2}$ weak isospin. d-type quarks, charged leptons and the Higgs boson have $-\frac{1}{2}$ weak isospin. The $W^{\pm}$ bosons have a weak isospin equal to their electric charge. Right chiral fermions, the Z^0-boson and the photon have zero weak isospin and so do not interact via the weak interaction. The weak interaction mediators are massive and this results in the range of the weak interaction being relatively small. This means that the weak interaction has a small effect at low energies. Flavour-changing interactions happen via the weak interaction, such as the quark flavour transformation from a d quark

[1] ... who had something of an opposition to the construction of new particle accelerators [46, 53]...

to a u quark via a $W^{\pm}$-boson observed in the beta decay of a neutron to a proton.

Quarks interact via the four fundamental interactions and can be categorised into families of u-type quarks, which have a positive electric charge $\frac{2}{3}e$, and d-type quarks, which have a negative electric charge $\frac{1}{3}e$, where e is the elementary positive electric charge, and all members of each of the u and d families share the same quantum numbers, but they do not share their generation number or mass. Quarks also carry colour charge, the charge associated with the strong interaction, and can be in states of R, B or G, with corresponding anti-colours. While "colour" in this context is largely meaningless, it serves as a basic metaphor with the mixing of colours of light. A combination of R, B and G quarks together, or one coloured quark and a corresponding anti-coloured quark, results in a colour-neutral particle. So, the colour 'combination' results in neutrons and protons averaging out as white. Quarks have the property of being confined in the 'white' particles, in the sense that it is only the 'white' particles that are observed directly. Similarly, observable objects have electric charge of only whole numbers, with fractionally charge particles existing only inside. For quarks to be confined in this way, the interactions between them are quite different to interactions such as those in electromagnetism. Just as the electromagnetic interaction between electrons is generated by the virtual exchange of photons, the quarks are bound together by the exchange of other quanta called gluons. Gluons carry a combination of colour and anti-colour charge in order to participate in strong interactions. That leads to nine possible combinations, however linear combinations like $R\bar{R}$, $G\bar{G}$ and $B\bar{B}$ correspond to colourless states and are held not to exist, while linear combinations such as $R\bar{R} + B\bar{B} + G\bar{G}$ cannot change the colour of a quark in an interaction, so also are held not to exist. Colour-neutral states are held as not possible for gluons. Unbalanced linear combinations are possible, though only two combinations are shown to be linearly independent from QCD. So, there are eight possible gluons, the colour octet: $R\bar{G}$, $R\bar{B}$, $G\bar{B}$, $G\bar{R}$, $B\bar{R}$, $B\bar{G}$, $\frac{1}{\sqrt{2}}\left(R\bar{R} - G\bar{G}\right)$ and $\frac{1}{\sqrt{6}}\left(R\bar{R} + G\bar{G} - 2B\bar{B}\right)$ (where the last two can be formed differently also). Formally, the non-abelian symmetry group $\mathrm{SU}(3)_{\mathrm{C}}$ expresses QCD, with quarks in fundamental representation of $\mathrm{SU}(3)_{\mathrm{C}}$ and gluons in adjoint representation of $\mathrm{SU}(3)_{\mathrm{C}}$, leading to the colour octet gluons. A sort of dictionary can be constructed to compare QED with QCD, as shown in Figure 2.1. Electric and colour charges are conserved; neither electric charge nor colour charge can be created or destroyed. However, in QED, the photon is electrically neutral while the gluons are colourful. Consequentially, the gluons self-interact in a way that photons do not, and that corresponds to terms in QCD equations that have no analogue in QED. This results in the colour interaction not diminishing at large distances in the same way that electromagnetic interactions diminish. While confinement limits direct observation of quarks in isolation, they were in a sense observed at the Stanford Linear Accelerator (SLAC), with the accelerator used effectively as a large electron microscope to probe nucleon structure via deep inelastic scattering, as depicted in Figure 2.3. From 1967 to about 1975 a series of measurements of deep inelastic electron scattering from the proton and neutron resulted in experimental evidence that the proton and neutron

were made up of quarks, with the "nucleon resonances" shown in Figure 2.4 [55].

A number of laudable visualisations of the proton were produced in a 2022 publication by the Thomas Jefferson National Accelerator Facility (TJNAF, Jefferson Lab) [56]. The visualisations highlight the dynamical systems of quarks and gluons that make up protons. What a proton looks like depends on both the momentum at which it is probed and the fraction of the its momentum that is carried by each of its quark or gluon constituents. Counter to the common outreach presentation of protons, it is not the case that each proton consists of precisely three quarks. A more accurate description would be to describe a quark as consisting of *at least* three quarks. Gluons split into quark-antiquark pairs, and for a smaller fraction of momentum carried by each quark, the proton appears more as a cloud of quarks and gluons, while for a higher momentum carried by each quark, the proton appears such that the total momentum of the proton is carried by a minimal number of particles, making it appear consistent with the usual three-quark outreach illustration. Protons also are not always spherical; with a change to the momentum, the shape of the proton changes consistent with Lorentz length contraction. "Looking" at a proton at these energies means colliding it with another particle such as an electron.

	QED	QCD		
Fermions	e^-	$u_R^{+\frac{2}{3}}$	$d_R^{-\frac{1}{3}}$	...
		$u_G^{+\frac{2}{3}}$	$d_G^{-\frac{1}{3}}$	...
		$u_B^{+\frac{2}{3}}$	$d_B^{-\frac{1}{3}}$	...
Bosons	γ^0	g^0		
Fermions	e^+	$\bar{u}_R^{-\frac{2}{3}}$	$\bar{d}_R^{+\frac{1}{3}}$	...
		$\bar{u}_G^{-\frac{2}{3}}$	$\bar{d}_G^{+\frac{1}{3}}$	...
		$\bar{u}_B^{-\frac{2}{3}}$	$\bar{d}_B^{+\frac{1}{3}}$	...

Table 2.1: Dictionary comparison of QED and QCD. Quarks and antiquarks are coupled to the photon through their electric charges while the electron and positron are not coupled to the gluons [46].

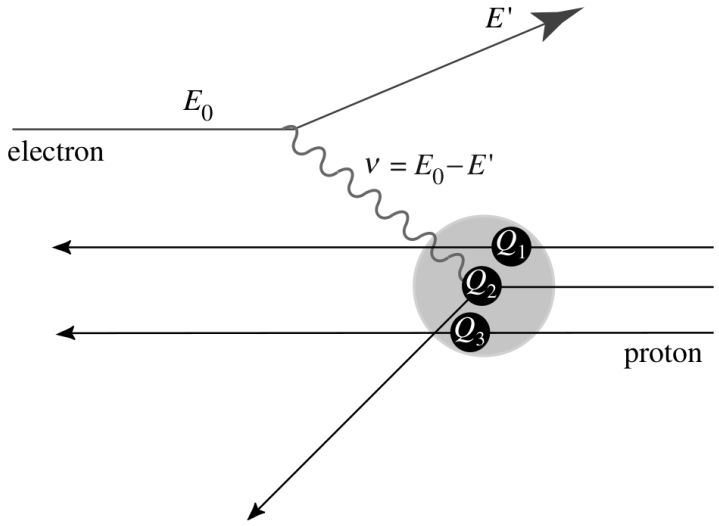


Figure 2.3: Deep inelastic scattering of an electron from a proton, with inelastic scattering treated as elastic scattering from small parts of the proton. The fraction of the proton momentum carried by the 'struck' parton is given by $\frac{Q^2}{2M\nu}$, with $Q^2 = 4E_0E' \sin^2 \frac{\theta}{2}$ [57].

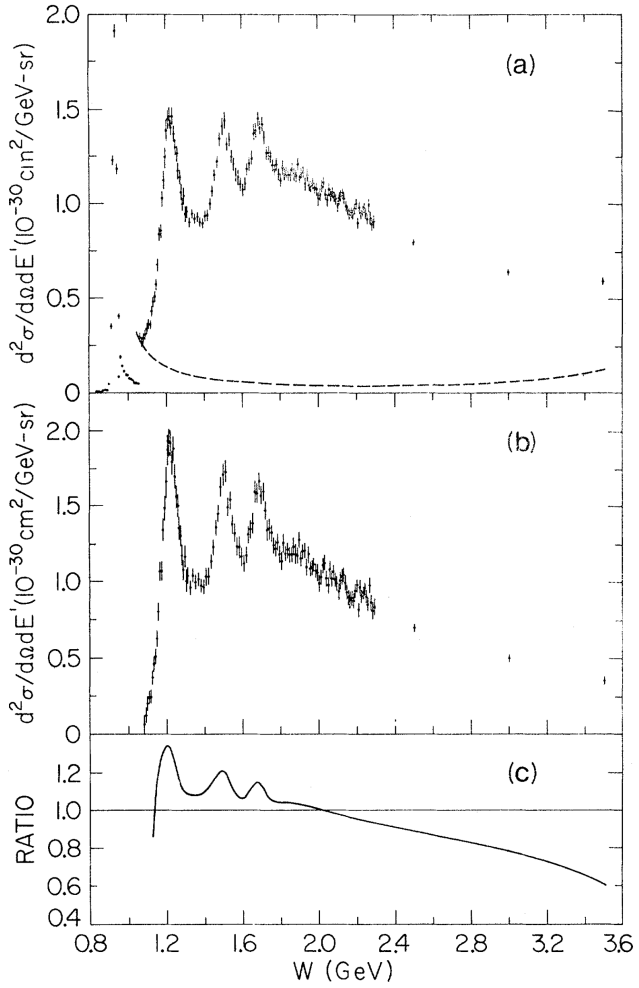


Figure 2.4: "Nucleon resonances": spectrum observed for deep inelastic scattering of an electron from a proton. The ratio is shown in (c) of the radiatively corrected (b) to the uncorrected (a) double differential cross-sections (with no systematic uncertainties shown), where W is the mass of the unobserved final state [58].

Much of the complexity, and puzzling character, of the Standard Model arises from the fermion sector; it is a sector with a large number of free parameters and there are huge hierarchies between the masses of these fermion particles. For example, the t-quark, with its mass of $173.0 \pm 0.4\,\text{GeV}$ is about 338,553 times the mass of the electron, with its mass of $0.5109989461 \pm 0.0000000031\,\text{MeV}$, and there are the neutrinos, which are much less massive again [43]. The t-quark is the most massive quark by more than

an order of magnitude.

Leptons interact via electromagnetic, weak and gravitational interactions, not strong interactions (so they have no colour charge), and can be categorised into generations of charged leptons, which have an electric charge equal to that of the electron, and neutral leptons (neutrinos), which have no electric charge, and all members of the charged lepton and neutrino families share the same quantum numbers, but they do not share their generation number or mass. The neutrinos are the least massive Standard Model particles and their masses are not well-known.

Quarks and leptons have corresponding antiparticles which have the same mass and opposite quantum numbers, with the operation of converting a particle into its antiparticle referred to as charge conjugation. Leptons can exist as free particles while colour-charged quarks are found in groups of two, three or more, consistent with confinement. These colour-neutral quark groups are called hadrons. Baryons are particles composed of an odd number of valence quarks (at least three quarks or at least three antiquarks) and mesons are particles composed of an equal number of quarks and anti-quarks (and so are often are a quark-antiquark pair). Protons and neutrons are examples of baryons. There is evidence of tetraquarks, consisting of two b-quarks and two light anti-quarks [59, 60]. Pentaquarks have been observed by the LHCb experiment [61]. Three of the leptons, the electrons, the muons and the tauons, carry electromagnetic charge and each has a corresponding neutrino. The electrons and muons are stable at LHC detectors, while the tauons tend to decay before reaching even the innermost sub-detectors, so generally in contemporary experimental particle physics, the term lepton, in a direct sense, refers to electrons and muons.

A visualisation of the Standard Model was developed by Chris Quigg. A modified form of this visualisation is shown in Figure 2.5.

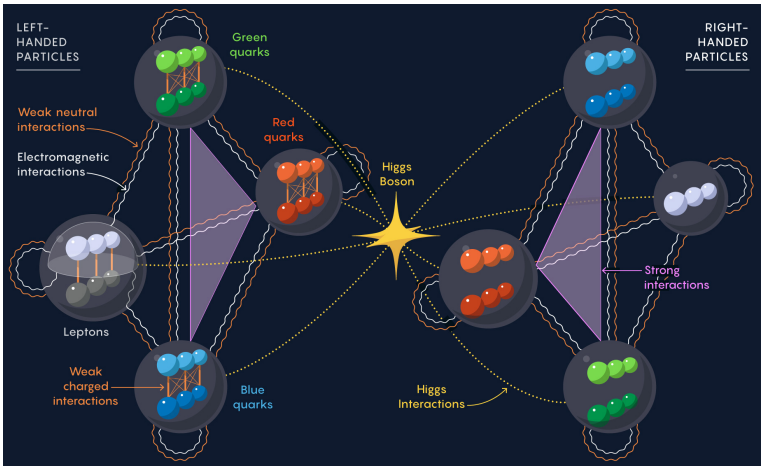


Figure 2.5: A modified form of the "double-simplex" representation of the Standard Model developed by Chris Quigg [62]. Left-handed and right-handed particles each form a simplex. For reasons of clarity, anti-particles are excluded (but would form a similar double-simplex), as are electromagnetic and weak neutral interactions between oppositely-handed particles in each simplex. Note especially that an important distinction between the left-handed and right-handed fields is that there are no right-handed neutrinos.

Left-handed fermions belong to weak isospin doublets while right-handed components transform as weak isospin doublets [38]. Weak decays transform a u-type quark (u, c and t) of charge $+\frac{2}{3}$ to a d-type quark (d, s, b), and vice versa, by exchange of virtual $W^{\pm}$ bosons. Generally a quark decays to the next most massive quark possible, corresponding to a decay chain of

$$t \to b \to c \to s \to u \leftrightarrow d, \tag{2.2}$$

which always involves a change of quark flavour. The virtual $W^{\pm}$-boson can decay to a lepton and a neutrino, which leads to a lepton presence in a meson decay. This is why leptons are present in jets originating from B-hadrons (b-tagged jets). Hadrons have other rare decay modes also. Weak decays are presented in terms of the mixed states d', s' and b' of d-type quarks. The Cabibbo-Kobayashi-Maskawa (CKM) quark-mixing matrix is a 3×3 unitarity matrix which encodes the strength of the flavour-changing weak interaction [38]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\mathrm{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.3}$$

The elements of the matrix represent the strength of the amplitude of the flavour-changing quark transition $i \to j + W$, a transition of quark i to quark j along with a $W^{\pm}$-boson. This is referred to as the weak charged current process (in contrast with the weak neutral processes involving a Z^0-boson). To first order, there are no transitions between quarks of the same charge, while higher-order diagrams predict such transitions, which are highly suppressed. Such transitions are called flavour-changing neutral current and corresponding diagrams are referred to as Penguin diagrams, an example of which is shown in Figure 2.6. Decay routes due to the weak interaction are depicted in Figure 2.7.

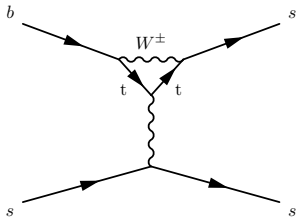


Figure 2.6: An example of a penguin diagram.

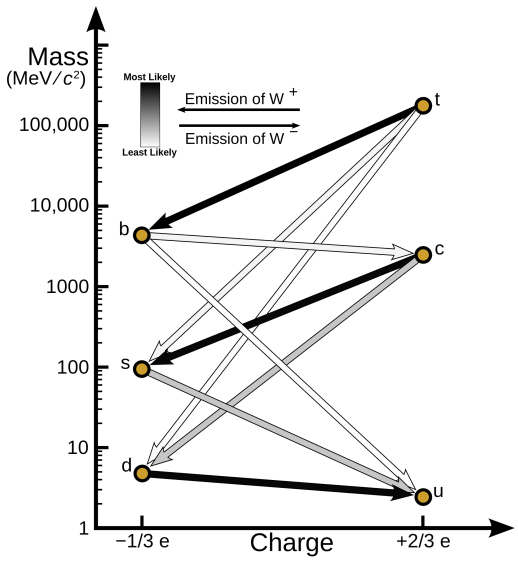


Figure 2.7: Decay routes due to the charged weak interaction and an indication of their probability. The line intensities correspond to CKM parameters [63].

The CKM matrix is described by three angles and one complex phase [38]. The CKM matrix elements are fundamental parameters of the Standard Model, so their precise determination is important. Observed

values for the CKM matrix elements are shown in Equation 2.4 [64]. The V_{CKM} terms represent transitions between the different generations of quarks. For example, $V_{tb} \simeq 1$, indicates that the t-quark decays in nearly all observations to a b-quark and a $W^{\pm}$-boson.

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97407 \pm 0.00011 & 0.22650 \pm 0.00048 & 0.00361^{+0.00011}_{-0.00009} \\ 0.22636 \pm 0.00048 & 0.97320 \pm 0.00011 & 0.04053^{+0.00061}_{-0.00083} \\ 0.00854^{+0.00023}_{-0.00016} & 0.03978^{+0.00082}_{-0.00060} & 0.999172^{+0.000024}_{-0.000035} \end{pmatrix}. \quad (2.4)$$

Predictive use of the Standard Model features computational challenges. Instead of calculating the paths of quarks and gluons through space and time using Feynman diagrams, in practice in computational simulations, each segment of space-time keeps track of how many quarks and gluons it contains and each of these segments is treated as an assembly of interacting subsystems.

Given the Standard Model gauge group $\mathrm{SU}(3)_{\mathrm{C}} \times \mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$, quarks and leptons can be grouped into five separate classes, or multiplets, and can be written as $(C, T)_Y$, where C denotes colour multiplet, T denotes weak isospin multiplet and Y denotes hypercharge value. Then the first generation multiplets may be depicted in the following way [65]. The quark multiplet is given as

$$Q = (3, 2)_{\frac{1}{3}} = \begin{pmatrix} u_R & u_G & u_B \\ d_R & d_G & d_B \end{pmatrix}, \quad (2.5)$$

the lepton doublet is given as

$$L = (1, 2)_{-1} = \begin{pmatrix} \nu_e \\ e \end{pmatrix}, \quad (2.6)$$

the anti-u quarks are given as

$$u^c = (\bar{3}, 1)_{-\frac{4}{3}} = \begin{pmatrix} u^c_{\bar{r}} & u^c_{\bar{g}} & u^c_{\bar{b}} \end{pmatrix}, \quad (2.7)$$

the anti-d quarks are given as

$$d^c = (\bar{3}, 1)_{\frac{2}{3}} = \begin{pmatrix} d^c_{\bar{r}} & d^c_{\bar{g}} & d^c_{\bar{b}} \end{pmatrix} \quad (2.8)$$

and the positron is given as

$$e^{c} = (1,1)_{2}\,. \tag{2.9}$$

The anti-neutrino is a singlet in the Standard Model and so does not transform under the Standard Model gauge group, and so can be written as

$$\nu^{c} = (1,1)_{0}\,. \tag{2.10}$$

The positron is included as the Standard Model 'multiplet' e_{c} as it has non-zero hypercharge and thus participates in the $\mathrm{U}(1)_{\mathrm{Y}}$ interactions, unlike ν^{c}.

For each generation there are these five multiplets; Q_{1} features u and d quarks, Q_{2} features c and s quarks, and so on. Q transforms as a triplet under $\mathrm{SU}(3)_{\mathrm{C}}$, as a doublet under $\mathrm{SU}(2)_{\mathrm{L}}$ and that it belongs to a non-trivial representation under $\mathrm{U}(1)_{\mathrm{Y}}$, meaning, as mentioned, that it has non-zero hypercharge. Thus it interacts via the strong nuclear, weak nuclear and electromagnetic interactions. The strong interactions (mediated by the gluons) can transform particles within a multiplet; the weak interactions (mediated by the $W^{\pm}$ bosons), can transform particles between multiplets; hypercharge interactions (mediated by a mixture of the photons and Z^{0} bosons) leave the particles unchanged, and underlie the electromagnetic interaction.

While the Standard Model has had many successes, it does not account for several experimental observations, such as the non-zero masses of neutrinos, the accelerating expansion of the observed universe, the abundance of matter over antimatter, the evidence for dark matter, and apples falling to the ground [66, 67, 68]. It features unexplained parameters that are taken from observations, such as the Cabbibo and Weinberg angles, the masses of the quarks and leptons and a number of other features. There is no consensus fundamental explanation for why the value of the Higgs field, from which the masses of the $W^{\pm}$ and Z^{0} bosons are derived, is so small in comparison to the Planck scale (which is one statement of the hierarchy problem). If there is some extended physics sector beyond the Standard Model, any extra contribution to the Higgs field from this sector must be balanced by the contribution from the Standard Model. Naturalness is the term used to refer to the aesthetic degree of mismatch acceptable, or how high the scale of the new theory can be.

2.2 The Standard Model formalism and the BEHGHK mechanism

The Lagrangian of the Standard Model has four parts:

$$\mathcal{L}_{SM} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{fermions}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}. \tag{2.11}$$

The first part of the Lagrangian corresponds to the kinematic and self-interaction terms of the various gauge fields. It is formed with the composition of the gauge field B_{μ} of the $\mathrm{U}(1)_{\mathrm{Y}}$ symmetry group, the three gauge fields W^{i}_{μ}, where $i = 1, 2, 3$, of the $\mathrm{SU}(2)_{\mathrm{L}}$ symmetry group and the eight gauge fields $G^{a}_{\mu\nu}$ of the $\mathrm{SU}(3)_{\mathrm{C}}$ symmetry group:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} G^{a}_{\mu\nu} G^{\mu\nu}_{a} - \frac{1}{4} W^{a}_{\mu\nu} W^{\mu\nu}_{a} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{2.12}$$

where, following the notation of Peskin and Schroeder, the gauge fields are defined as

$$\begin{aligned} G^{a}_{\mu\nu} &= \partial_{\mu} G^{a}_{\nu} - \partial_{\nu} G^{a}_{\mu} + g_{3} f^{abc} G^{b}_{\mu} G^{c}_{\nu},\ a \in [1, 8], \\ W^{i}_{\mu\nu} &= \partial_{\mu} W^{i}_{\nu} - \partial_{\nu} W^{i}_{\mu} + g_{2} \epsilon^{ijk} W^{b}_{\mu} W^{c}_{\nu},\ i \in [1, 3], \text{ and,} \\ B_{\mu\nu} &= \partial_{\mu} B_{\nu} - \partial_{\nu} B_{\mu}, \end{aligned} \tag{2.13}$$

where g_{2} and g_{3} are the weak isospin and the strong coupling constant of the $\mathrm{SU}(2)_{\mathrm{L}}$ and $\mathrm{SU}(3)_{\mathrm{C}}$ symmetry groups, respectively [38].

The second part of the Lagrangian corresponds to the fermion gauge boson coupling:

$$\mathcal{L}_{\text{fermions}} = \sum_{\psi_{L}, \psi_{R}} \bar{\psi} i \gamma^{\mu} D_{\mu} \psi, \tag{2.14}$$

where the summation is over both the left L and right R field components of the leptons and quarks. The fields ψ are coupled to the gauge fields through the covariant derivative D_{μ}, which, for the case of quarks, is

$$D_{\mu} \psi = \left(\partial_{\mu} - i g_{3} T_{a} G^{a}_{\mu} - i g_{2} I_{i} W^{i}_{\mu} - i g_{1} \frac{Y}{2} B_{\mu} \right) \psi, \tag{2.15}$$

where W^{i}_{μ} corresponds to the three vector bosons, $W^{\pm}$ and Z^{0} ($i \in \{1, 2, 3\}$), I_{i} corresponds to components of weak isospin, T_{a} corresponds to one of the Gell-Mann matrices, Y corresponds to weak hypercharge and g_{1} is the weak hypercharge coupling constant of the $\mathrm{U}(1)_{\mathrm{Y}}$ symmetry group.

The electroweak part of the gauge fields Lagrangian is as follows:

$$\mathcal{L}_{EW} = -\frac{1}{4} W^{a}_{\mu\nu} W^{\mu\nu}_{a} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}. \tag{2.16}$$

Introducing the Weinberg angle θ_W, which is defined using the ratio of the $\mathrm{SU}(2)_{\mathrm{L}}$ and $\mathrm{U}(1)_{\mathrm{Y}}$ coupling constants, $\tan \frac{g_1}{g_2}$, the following transformations can be defined:

$$\begin{aligned} A_\mu &= \sin\theta_W W^3_\mu + \cos\theta_W B_\mu, \\ Z_\mu &= \cos\theta_W W^3_\mu - \sin\theta_W B_\mu \text{ and} \\ W^{\pm} &= \frac{1}{\sqrt{2}} \left(W^1_\mu \mp i W^2_\mu \right), \end{aligned} \tag{2.17}$$

where A corresponds to the massless neutral photon, Z_μ corresponds to the massive neutral Z^0-boson and $W^{\pm}$ corresponds to the massive charged $W^{\pm}$ bosons. These transformations can reformulate the electroweak part of the gauge fields Lagrangian into

$$\mathcal{L}_{\mathrm{EW}} = -\frac{1}{4} F_{\mu\nu}(x) F^{\mu\nu}(x) - \frac{1}{2} F^{\dagger}_{W\mu\nu}(x) F^{\mu\nu}_{W}(x) - \frac{1}{4} F_{Z\mu\nu}(x) F^{\mu\nu}_{Z}(x), \tag{2.18}$$

where $F_{\mu\nu}$ is the electromagnetic tensor for the photon field A_μ, $F_{W_{\mu\nu}}$ is the electroweak tensor for the $W^{\pm}$ bosons field and $F_{Z_{\mu\nu}}$ is the electroweak tensor for the Z^0 bosons field.

The $W^{\pm}$ and Z^0 bosons can acquire masses by adding the following terms to the electroweak part of the gauge fields Lagrangian:

$$m^2_W W^{\dagger}_\mu(x) W^{\mu}(x) + \frac{1}{2} m^2_Z Z_\mu(x) Z^{\mu}(x). \tag{2.19}$$

This addition of mass terms for the gauge bosons, and for the fermions, corresponds to a breaking of the local $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$ gauge invariance. So, a scalar potential is added to the Lagrangian in order to generate the masses without explicit breaking of the $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$ gauge symmetry. This is introduced by the BEHGHK mechanism.

As can be seen from Figure 2.5, weak neutral interactions resemble electromagnetic interactions. The Z-boson contributes in every process the photon does, but due to its relatively large mass, the contribution is often considered negligible. The weak and electromagnetic interactions derive from a single interaction called the electroweak interaction. Theoretically at earlier times of the universe these interactions were more unified, and as the universe cooled, electroweak symmetry breaking split the interactions – a time marked by the appearance of the Higgs field.

Coarsely conceptualised, as a particle such as an electron moves through space, it interacts constantly with Higgs bosons – excitations of the Higgs field. When a left-handed electron interacts with a Higgs boson, the electron might change direction and become right-handed, then interact with another Higgs boson and become left-handed again, and so on and so on. In a sense, these interactions slow down the electron and result in behaviours referred to as mass and inertia. Generally the more a particle interacts with the Higgs boson, the more mass it has.

Spontaneous symmetry breaking occurs when the Lagrangian of a system is invariant under the symmetry group, but the ground state breaks the symmetry. In order to achieve this, a complex scalar field is introduced. This field is required to couple to the gauge fields, so has a non-zero weak hypercharge and isospin. The electric charge is zero in order to preserve $\mathrm{U}(1)_{\mathrm{Y}}$. Further, there must be sufficient degrees of freedom to give mass naturally to the fermions and the vector gauge bosons. A simple representation of this Higgs field is a complex scalar doublet

$$\phi = \begin{pmatrix} \phi^{+} \\ \phi^{0} \end{pmatrix} \tag{2.20}$$

and the Higgs Lagrangian is

$$\mathcal{L}_{\mathrm{Higgs}} = \left(D^{\mu}\phi\right)^{\dagger}\left(D_{\mu}\phi\right) - V\left(\phi^{\dagger}\phi\right), \tag{2.21}$$

where the complex scalar potential $V\left(\phi^{\dagger}\phi\right)$ is of the form

$$V\left(\phi^{\dagger}\phi\right) = \mu^{2}\phi^{\dagger}\phi + \lambda\left(\phi^{\dagger}\phi\right)^{2}. \tag{2.22}$$

For $\mu^{2} > 0$ the minimum is at $\phi = 0$, so the vacuum is empty space and the $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$ gauge symmetry is unbroken at this minimum. For $\mu^{2} < 0$ and $\lambda > 0$ a Sombrero potential is formed, as shown in Figure 2.8. This potential $V(\phi)$ features two critical points, a local maximum at $\phi = 0$ and the non-zero minimum at $\phi_{0} = \sqrt{\frac{-\mu^{2}}{2\lambda}}$, which breaks the $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$ symmetry invariance. That is, the Higgs potential has minima at values of the field strength that are non-zero (the valleys surrounding the central peak) and, conversely, when the field strength is zero (at the centre), the Higgs potential is not at its lowest possible value. It is not known why the Higgs potential has this shape (as opposed to having a minimum at the origin); it is an input to the Standard Model; perhaps a more complete theory might give some explanation for it. Given the potential as described, the Higgs vacuum (with the field strength being zero) is unstable. It decays into a lower energy state at which the field strength is non-zero. This

is the spontaneous symmetry breaking mechanism of the Higgs field, which necessarily arises because of the quartic self-interaction potential of the field.

The Higgs field interacts with other fields, so this symmetry breaking has implications. If the minimum energy state were at the centre, other fields would have nothing with which to interact, as the Higgs field strength would be zero, but this is not the case. Since the minimum energy state has the Higgs field strength as non-zero, other fields almost always have something with which to interact. This interaction with the minimum energy Higgs field is the Higgs field vacuum expectation value, and it means that other interacting fields behave as though they have mass. This is why the Higgs field is said to 'give' mass to charged fermions.

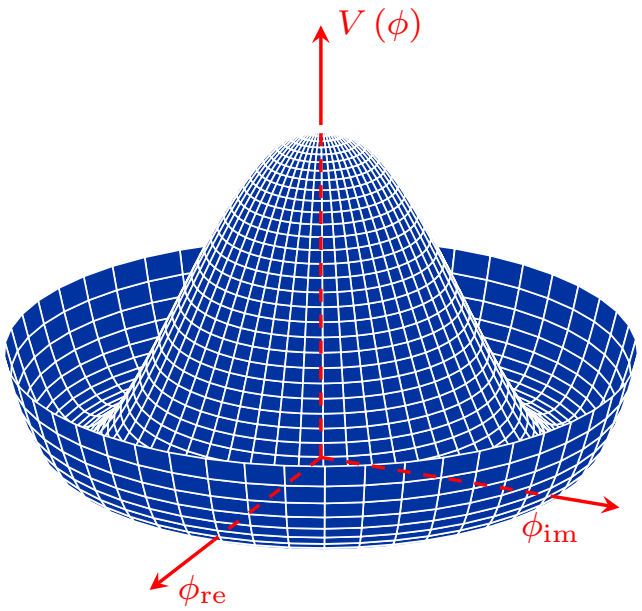


Figure 2.8: Higgs potential.

The Higgs field can be expanded around the minimum to produce the expression

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{2.23}$$

where $\nu = \sqrt{\frac{-\mu^2}{\lambda}}$ is the vacuum expectation value and the field $h(x)$ describes small perturbations about the ground state. It is the lowest energy excitation of the Higgs field above its ground state that is referred to as the Higgs boson[2].

Now, the Higgs Lagrangian can be expressed in terms of the Higgs field:

$$\mathcal{L}_{\text{Higgs}} = \left| \left(\partial_\mu - ig_2 I_i W_\mu^i - ig_1 \frac{Y_q}{2} B_\mu \right) \frac{(v+h)}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right|^2 - \mu^2 \frac{(v+h)^2}{2} - \lambda \frac{(v+h)^4}{4}, \tag{2.24}$$

where the first term, corresponding to the vector bosons, can be expanded to

$$\frac{1}{2}\left(\partial_\mu h\right)^2 + \frac{1}{8} g_2^2 (v+h)^2 \left| W_\mu^1 + iW_\mu^2 \right|^2 + \frac{1}{8}(v+h)^2 \left| g_2 W_\mu^3 - g_1 B_\mu \right|^2. \tag{2.25}$$

This can be expressed in terms of the photon field A_μ and the vector boson fields $W^\pm$ and Z^0 field:

$$\frac{1}{2}\partial_\mu h \partial^\mu h + \frac{g_2^2}{4}(v+h)^2 \left(W_\mu^+ W^{-\mu} + \frac{g_2^2 + g_1^2}{2 g_2^2} Z_\mu Z^\mu \right). \tag{2.26}$$

The bilinear terms in the fields are

$$m_W^2 W_\mu^\dagger W^\mu + \frac{1}{2} m_Z^2 Z_\mu Z^\mu + \frac{1}{2} m_A^2 A_\mu A^\mu. \tag{2.27}$$

The mass of the photon is kept as zero ($m_A = 0$), while the masses of the vector bosons are

$$m_W = \frac{1}{2}\nu g_2 \text{ and } m_Z = \frac{1}{2}\nu\sqrt{g_2^2 + g_1^2}. \tag{2.28}$$

Expressing the Higgs field in terms of its ground state via the addition of the real scalar field h induces effective masses for particles propagating through it.

The second part of the Higgs Lagrangian (see Equation 2.24) gives rise to terms involving the scalar field h:

$$-\frac{1}{2}\left(-2\mu^2\right) h^2 + \frac{1}{4}\mu 2v^2 \left(\frac{4}{v^3} h^3 + \frac{1}{v^4} h^4 - 1 \right), \tag{2.29}$$

and, with the vacuum expectation value $\nu = \sqrt{\frac{-\mu^2}{\lambda}}$, the Higgs boson mass term is

[2]The *higgson* is perhaps a more convenient name [46].

$$m_h = \sqrt{-2\mu^2} = \sqrt{2\lambda\nu}. \tag{2.30}$$

The relation $\nu = 1/\sqrt{\sqrt{2}G_F^0} \simeq 246.22\,\text{GeV}$ can be used to determine the vacuum expectation value of the Higgs field ν, where G_F^0 is the reduced Fermi constant. However, the parameter λ is associated purely with the scalar field, and so the Higgs mass cannot be predicted from the current theory (without knowledge of the scalar field itself).

Fermions are given their masses by interacting with the scalar field ϕ. The $\text{SU}(2)_\text{L} \times \text{U}(1)_\text{Y}$ Yukawa Lagrangian is

$$\mathcal{L}_\text{Yukawa} = g_f \bar{f} f \phi. \tag{2.31}$$

Replacing the Higgs field with its ground state value $\frac{\nu}{\sqrt{2}}$ changes the Lagrangian to

$$\mathcal{L}_\text{Yukawa} = g_f \frac{\nu}{\sqrt{2}} \bar{f} f, \tag{2.32}$$

where the first constant is defined as the fermion mass:

$$g_f \frac{\nu}{\sqrt{2}}. \tag{2.33}$$

The fermion masses can be introduced in a consistent way using the Higgs mechanism, however their values are not predicted by the Standard Model.

The t–H Yukawa coupling, g_t is notable for being of the order of 1:

$$g_t = m_t \frac{\sqrt{2}}{\nu} \simeq 1. \tag{2.34}$$

Higgs couplings are proportional to the coupled particle mass. Thus, the Higgs boson is produced in association with very massive particles and decays to the most massive particles kinematically possible. Higgs coupling strengths to fermions, vector gauge bosons and to itself are shown in Figure 2.2.

Coupling	Coupling strength
$Hf\bar{f}$	$\frac{m_f}{\nu}$
HVV	$\frac{2m_V^2}{\nu}$
$HHVV$	$\frac{2m_V^2}{\nu^2}$
HHH	$\frac{3m_H^2}{\nu}$
$HHHH$	$\frac{3m_H^2}{\nu^2}$

Table 2.2: Higgs boson couplings to fermions f, vector gauge bosons V and to itself H.

2.3 The *t*-quark

The t-quark is the most massive particle of the Standard Model, with a mass of $173.0 \pm 0.4\,\text{GeV}$ [43]. Its estimated lifetime of 10^{-25}s is two orders of magnitude shorter than the timescale of hadronisation, so it is unusual in that it decays before it can form any hadrons. It can decay directly via $t \to W^+ + q$, where $q = d, s, b$, according to flavour-changing weak decays. The couplings of the t-quark to the d and s-quarks are negligible, while in the CKM matrix, $|V_{tb}| \simeq 1$ indicates that the t-quark decays dominantly to a b-quark and a $W^{\pm}$-boson. Thus, examinations of t-quark decays can elucidate properties of a quark, such as its spin. Due to these apparently unique properties, the t-quark has a very recognisable decay signature and provides a way to examine the properties of quarks before hadronisation.

Typically, the decay channel of the $W^{\pm}$-boson is used to categorise t-quarks in experimental analyses. For a decay to a quark pair, the t-quark is considered hadronic and for a decay to an electron or muon neutrino, the t-quark is considered leptonic. For a decay to a tauon neutrino pair, if the tauon decays leptonically then it is considered leptonic and if the tauon decays hadronically, then the t-quark is considered hadronic, with the tauons treated as jets or treated separately with the tauons reconstructed experimentally. These decay channels are used further to classify $t\bar{t}$ and $t\bar{t} + X$ events using the decays of the two t-quarks. *All-hadronic* events feature both t-quarks decaying hadronically, *semileptonic* events feature one t-quark decaying leptonically and the other hadronically (also referred to as single lepton and lepton + jets) and *dileptonic* events feature both t-quarks decaying leptonically. Thus, excluding $t\bar{t}$ events with tauons (approximately 20 % of the $t\bar{t}$ events), the all-hadronic channel is approximately 46 % of the $t\bar{t}$ events, the semileptonic channel is approximately 30 % of the $t\bar{t}$ events and the dileptonic channel is approximately 4 % of events.

Quark and lepton masses are shown in Figure 2.9. The t-quark mass is approximately the same as the average energy of the Higgs field. There are theoretical conceptions that suggest that only the t-quark propagates through the Higgs field in a standard way, while other particles experience the Higgs field

indirectly [69]. For example, virtual t-quarks could expose a muon to the Higgs field by means of a mutual interaction with some new theoretical particle, resulting in the muon having some mass, but the muon remaining much less massive than the t-quark. A similar approach can be taken for the case of the electron. This is one approach under consideration for deriving the rough spacing of the masses of the generations of particles, with the masses of the particles becoming approximately exponentially less massive going down the generations.

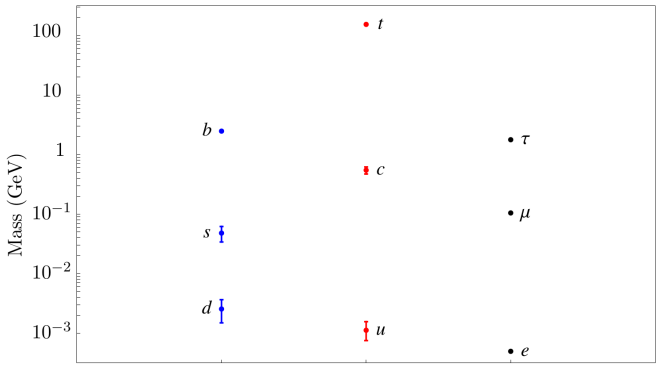


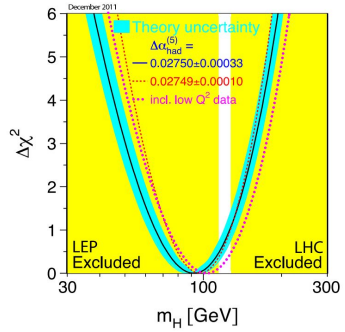
Figure 2.9: Quark and lepton masses at the 1 TeV scale [69].

Weinberg was involved in research on this topic and considered a variety of ways by which this scheme could function, where broadly mass propagates down from third generation particles (the t-quark, the b-quark and the tauon) to the second and first generation particles via interactions with theoretical exotic particles [70]. So, essentially, electroweak symmetry breaking is considered as resulting in mass only for the third generation quarks and leptons to tree-level approximation, and mass for the second and first generation quarks and leptons is considered as arising from the emission and reabsorption of some sort of gauge bosons. There remain shortcomings with these proposals.

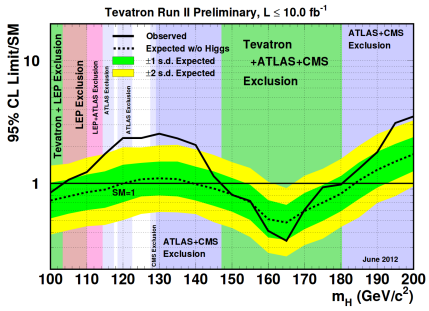
2.4 The Higgs boson

2.4.1 Past searches for the Higgs boson: from stellar emissions to the LEP to the Tevatron

Searches for the Higgs boson go back to at least 1974, with early investigations examining everything from neutron-electron scattering to stellar emissions [71, 72, 73, 74]. However, the 1976 CERN Yellow Report *Physics with Very High-Energy* e^+e^- *Colliding beams* offered detailed plans for searches for the Higgs boson using the Large Electron-Positron Collider (LEP) [10]. The LEP collider enabled investigation of the associated production $e^+e^- \to ZH$ and $Z \to H\left(\mu^+\mu^-\right)$ Higgs boson production processes in order to constrain the Higgs boson mass [9, 75, 76]. In time, electroweak radiative corrections measured by the LEP collider were used to estimate the Higgs boson mass [77]. The Tevatron collider enabled investigation of the gluon-gluon fusion ($gg \to H$) and vector boson ($W^{\pm}H$ and ZH) Higgs boson production processes. The Tevatron Higgs boson exclusion is shown in Figure 2.10b. Discovery of the t-quark by the CDF and DØ collaborations at the Tevatron in 1995 enabled reduction of uncertainties pertinent to Higgs boson searches significantly. The LEP Higgs boson exclusion is shown in Figure 2.10a. While the LEP searches did not show any conclusive evidence for the production of a Standard Model Higgs boson, the Tevatron ultimately found evidence, with a Higgs boson mass of 120 GeV corresponding to a local significance of 3σ [78, 79, 80]. The Tevatron combined CDF and DØ observed and expected upper limits on the ratio to the Standard Model cross-section for various Higgs boson masses is shown in the exclusion plot in Figure 2.10b.



(a) LEP Higgs boson exclusion (with LHC 2011 exclusion) [78, 79].



(b) June 2012 Tevatron Higgs boson exclusion (with LHC exclusion) [80].

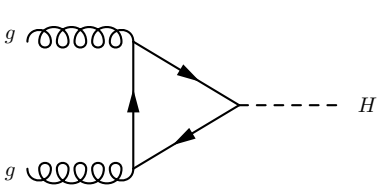
Figure 2.10: Past search exclusions for the Higgs boson.

For a more detailed history of past historical searches for Higgs bosons, please refer to the author's M.Sc. thesis [81].

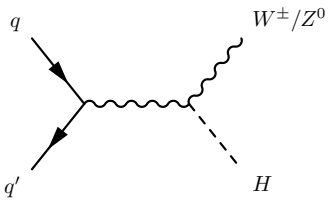
2.4.2 Searches for the Higgs boson at the Large Hadron Collider

In the Standard Model, the Higgs boson couples preferentially to the more massive particles, such as the $W^{\pm}$ and Z^0 bosons, and the t and b quarks. So, Higgs bosons are produced predominantly involving these more massive particles. The production of the Standard Model Higgs boson proceeds via gluon-gluon fusion which produces the Higgs via a quark loop, or $W^{\pm}$ or Z^0 fusion to a Higgs boson, as well as production of the Higgs in association with a vector boson or a t-quark. The Standard Model Higgs boson with a mass of 125 GeV decays predominantly to b-quark quark-anti-quark pairs. The Large Hadron Collider (LHC) enabled investigation of the various Higgs boson production processes shown in

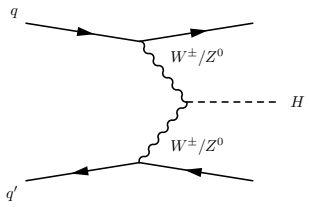
Table 2.11. Feynman diagrams of Standard Model Higgs boson LHC production processes are shown in Figure 2.11.



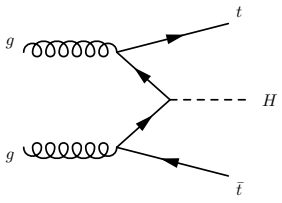
(a) Gluon-gluon fusion (ggF).



(b) Associated vector boson.



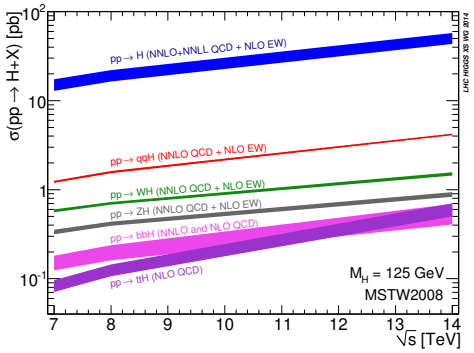
(c) Vector boson fusion (VBF).



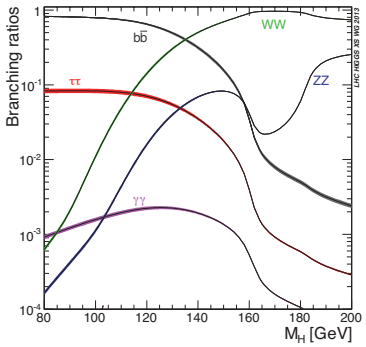
(d) Associated t-quark pair.

Figure 2.11: Higgs production processes.

Cross sections of various Higgs production processes for various collision energies are shown in Figure 2.12a while branching ratios of various Higgs decay processes for various Higgs mass assumptions are shown in Figure 2.12b.



(a) Cross sections of Higgs processes.



(b) Branching ratios of Higgs decays.

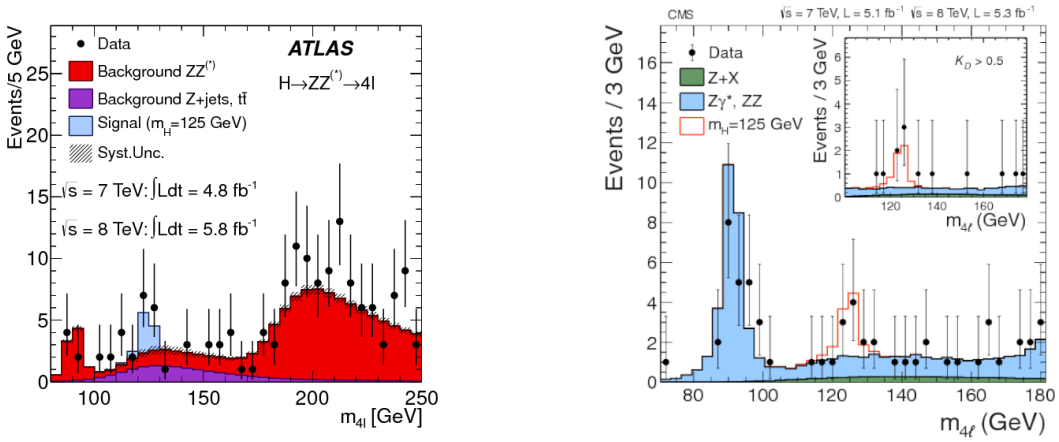
Figure 2.12: Higgs production processes for various collision energies and Higgs decays for various Higgs mass assumptions [82].

Higgs boson production cross-sections at the LHC for various production processes and an assumed Higgs mass of 125 GeV are shown in Table 2.3.

Production process	Cross section (pb)
ggF	43.92
VBF	3.748
WH	1.380
ZH	0.8696
$t\bar{t}H$	0.5085
tH	0.074

Table 2.3: Higgs boson production cross-sections at at $\sqrt{s} = 13$–14 TeV for various production processes and an assumed Higgs boson mass of 125 GeV [83, 84].

The discovery of the Higgs boson was announced in July 2012 by the ATLAS and CMS collaborations [11, 12]. The observed and expected mass distributions for events following the full selections are shown for ATLAS and CMS in Figure 2.13.



(a) ATLAS [11].

(b) CMS [12].

Figure 2.13: Distributions of invariant mass for selected events with two pairs of electrons or muons (each pair with opposite charges) presented by the ATLAS and CMS collaborations for the discovery of the Higgs boson in 2012. For the CMS case, the inset plot shows the result of applying a tight selection to a kinematic discriminant designed to separate signal and background (based on the probability ratio of the signal and background hypotheses).

Since the discovery of the Higgs boson, there has been intense research into its characteristics, particularly its couplings, which serve as a test of the precision of Standard Model predictions [11, 12, 13, 14]. Since 2012, ATLAS measurements have found the overall production rate of the Higgs boson to be in good agreement with the Standard Model, within a measurement precision of 5 % and about 4 % uncertainty in the Standard Model prediction [85].

Assuming the Standard Model Higgs boson, the ATLAS measurement of the t–H Yukawa coupling was 0.87 ± 0.15 times the Standard Model prediction, by the combination of their respective Higgs boson measurements from LHC Run-1 data collected at centre-of-mass energies of 7 TeV and 8 TeV [14]. This measurement relied largely on the production of Higgs bosons via gluon-gluon fusion, as shown in Figure 2.11a, and on the decay mode to photons. The mass of the Higgs boson was measured to be

$$m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.})\,\text{GeV},$$

with the total uncertainty dominated by the statistical component [14].

The 2020 measurement of the mass of the Higgs boson is

$$m_H = 125.10 \pm 0.14\,\text{GeV}\ [64].$$

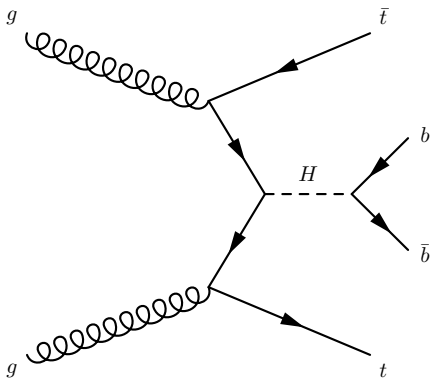
2.4.3 Searching for the Higgs boson in $t\bar{t}H$

Higgs bosons arise theoretically in multiple production processes, as described in Section 2.4.2, one of which is the $t\bar{t}H$ process. As described in Section 2.3, the t-quark, which decays almost exclusively to a $W^{\pm}$ bosons and a b-quark, is the most massive of all of the observed elementary particles, and the t–H Yukawa coupling Y_t is ~ 1, with an experimental measurement of $y_t = 1.16^{+0.24}_{-0.35}$, which is in good agreement with the Standard Model prediction [86, 87]. This is almost two orders of magnitude greater than the next largest Yukawa coupling, to the b-quark. The Higgs production mode associated with a t-quark pair, $t\bar{t}H$, is a way to measure relatively directly the t–H Yukawa coupling, as shown in Figure 2.11d [88, 89, 90, 91]. The cross-section $\sigma\,(pp \to t\bar{t}H)$ is directly proportional to the square of this coupling. The process is interesting and useful because its discovery and measurement is powerful evidence for the Standard Model and deviations could indicate physics beyond the Standard Model (BSM).

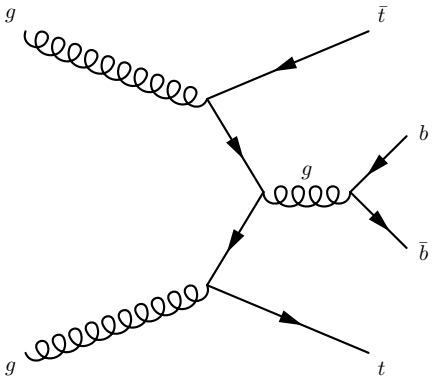
The $t\bar{t}H$ production mode contributes only about 1 % of the total Higgs boson production cross-section, as shown in Figure 2.12a [92]. One of the various Higgs boson decay modes from $t\bar{t}H$ is the Higgs boson decay to a b-quark pair, $H \to b\bar{b}$. For a Higgs mass of about 125 GeV, $H \to b\bar{b}$ is theoretically the dominant decay mode, with a branching fraction of about 58 % [92]. This decay mode is sensitive to the b-quark Yukawa coupling Y_b, the second largest in the Standard Model. While a direct search from gluon fusion is precluded by overwhelming multi-jet backgrounds, a search in association with a gauge vector

boson or a t-quark pair significantly improves the signal-to-background ratio. For $t\bar{t}H$, the cross-section is low and the process is quite suppressed, but the branching fraction to two b-quarks is very high, so there are a relatively large numbers of Higgs bosons arising from the $t\bar{t}\left(H \to b\bar{b}\right)$ process. Hence, a search for Higgs bosons in the $t\bar{t}H$ production mode that is designed to be sensitive primarily to the $H \to b\bar{b}$ decay is worthwhile; the search benefits from the large $H \to b\bar{b}$ branching fraction and leptonic decays, which contrast with a large irreducible background from $t\bar{t}+$ heavy flavour production. The work of this thesis contributed to this analysis. For $t\bar{t}H$, the cross-section and the process is heavily suppressed, as shown in Figure 2.12a, but the branching ratio to a b-quark pair is very high, as shown in Figure 2.12b), so theoretically there is a relatively large number of Higgs bosons from the $t\bar{t}\left(H \to b\bar{b}\right)$ process, though the process is rare.

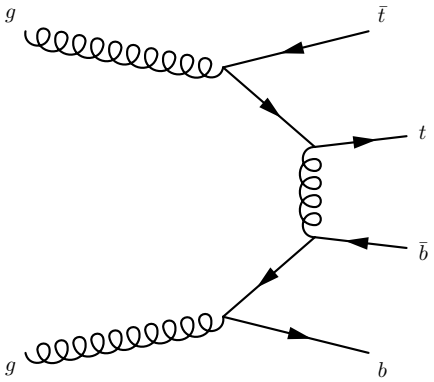
The main background for signal $t\bar{t}H\left(b\bar{b}\right)$ is t-quark pairs produced in association with additional jets. The dominant source is $t\bar{t}+b\bar{b}$ production, which results in the same final state particles as the signal. Examples of signal and background processes are shown in Figure 2.14. Other backgrounds include $t\bar{t}$ production in association with light quarks (u, d and s) and gluon jets (called $t\bar{t}$ + light) and $t\bar{t}$ + $c\bar{c}$. A more detailed Feynman diagram featuring quarks and leptons in the final state for signal is shown in Figure 2.15.



(a) $t\bar{t}H\left(b\bar{b}\right)$.

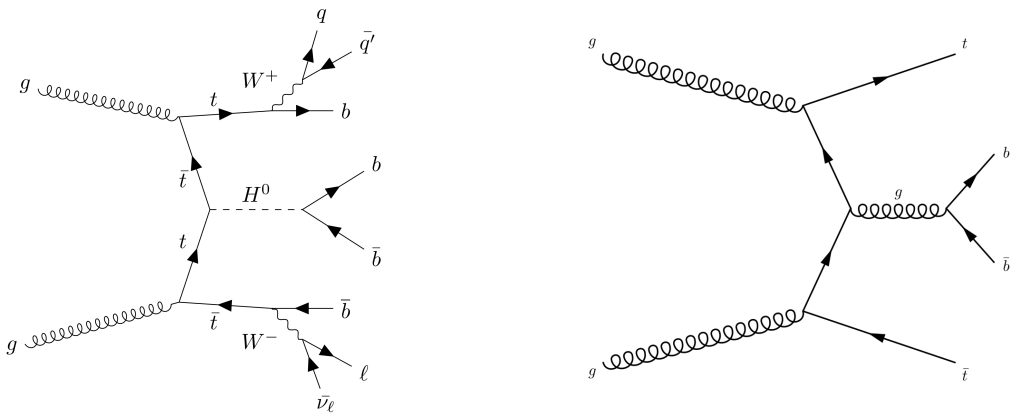


(b) $t\bar{t}b\bar{b}$.



(c) $t\bar{t}b\bar{b}$.

Figure 2.14: Signal and dominant backgrounds have similar final states. Distinguishing the processes is no simple matter.



(a) $t\bar{t}H\left(b\bar{b}\right)$. (b) $t\bar{t}b\bar{b}$.

Figure 2.15: Gaurishankar and Mount Everest: leading-order Feynman diagrams of the semileptonic signal process $t\bar{t}H\left(b\bar{b}\right)$ with the main background $t\bar{t}b\bar{b}$.

The search for the associated production of the Higgs boson with a t-quark pair and decaying into a b-quark pair at the LHC began with the Run-1 $t\bar{t}H\left(b\bar{b}\right)$ analysis [93]. Evidence for the existence of the associated production of the Higgs boson with a t-quark pair, the $t\bar{t}H$ process, was established and then declared in 2018 by the ATLAS and CMS collaborations [1, 32, 34]. The main analysis of this thesis contributed to the ATLAS version of this finding. The measurement of the process was challenging as only about 1 % of Higgs bosons are produced in association with two t-quarks, and the Higgs bosons and t-quarks decay into other particles in many modes.

Following this finding, work continued using data corresponding to an integrated luminosity of $79.8\,\text{fb}^{-1}$ and a combination of results from Higgs boson decays to $b\bar{b}$, WW^*, $\tau^+\tau^-$, $\gamma\gamma$ and ZZ^* with an observed significance of 5.8σ, compared to an expected significance of 4.9σ [2]. Combined with $t\bar{t}H$ search results using data recorded at centre-of-mass energies of $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$, the observed significance was 6.3σ, compared to an expected significance of 5.1σ. Assuming Standard Model branching fractions, the $t\bar{t}H$ production cross-section at $\sqrt{s} = 13\,\text{TeV}$ was measured as $670 \pm 90\,(\text{stat.})^{+110}_{-100}\,(\text{syst.})\,\text{fb}$, which is in agreement with the Standard Model prediction of $507^{+35}_{-50}\,\text{fb}$.

The observation of Higgs boson production in association with a t-quark pair was announced in 2018 by the ATLAS and CMS collaborations [2, 3].

2.5 Summary

The Standard Model of particle physics is a theory which provides a successful description of the electromagnetic, weak and strong interactions between known fundamental particles. Observation of the Standard Model Higgs boson, an important particle of the Standard Model, was announced by the ATLAS and CMS collaborations in July 2012 [11, 12].

Precision measurements of the Higgs boson properties, such as its mass, spin, CP and couplings, are important in searching for possible deviations from the Standard Model. Due to its large mass, the t-quark coupling to the Higgs boson is the largest of the Standard Model fermions. Indirect constrains of the t–H Yukawa coupling were published in Run-1 using Higgs gluon-gluon fusion production and the $H \to \gamma\gamma$ decay [94]. The associated production of a Higgs boson with a t-quark pair – the $t\bar{t}H$ process – is a way to measure the t–H Yukawa coupling at the LHC in a relatively direct way. Thus, the observation of the $t\bar{t}H$ process, which was announced in 2018, was a major highlight of Run-2 and advancement of the field of particle physics [2, 3]. The search for this process is one of the main topics of this thesis.

Chapter 3

The Large Hadron Collider and A Toroidal LHC ApparatuS

> I've become aware of the great number of countries which have contributed to the components of that machine: you go 'round and you look at the origins of components in the tunnel and you find this one came from the United States, the next one came from the Soviet Union and so on.
>
> Peter Higgs (2012)

This chapter provides a brief description of the experimental apparatus used in the investigation described by this thesis. The author identifies with the sentiments of Lisa Randall; specifically, the description of the LHC as an apparatus feels like a tremendous understatement. CERN, an international organisation born after terrible war and atrocity has advanced us in scientific collaboration to a point at which the experimental apparatus, the LHC and its various experiments, is not just to be described as large – indeed, city-scale – but as the largest machine ever built. Its interior is not just cold, but is the coldest place known. Its vacuum is the most vacuous over the largest region yet produced. The energy density of the collisions are the highest ever produced on Earth. This is the experimental apparatus.

3.1 The Large Hadron Collider

A particle collider is a machine that accelerates one or more beams of particles such that they collide. Early versions had a single beam colliding with a fixed target. While this is more easily built, much of the energy in the initial state is lost as kinetic energy in the final state. For the past few decades colliders

have mostly consisted of two beams of particles that collide in the centre-of-mass frame. Electric and magnetic fields accelerate the particles and focus the beams, and so the particles accelerated must be electromagnetically charged.

The year 2021 marked the 50^{th} anniversary of hadron colliders. In 1971, the Intersecting Storage Rings (ISR) at the European Organisation for Nuclear Research (CERN) were the first accelerator in human history to produce collisions between two counter-rotating beams of hadrons. Since that time, decades of learning and effort have resulted in the Large Hadron Collider (LHC), which is today the largest and most powerful particle collider in the world.

The LHC is an example of a synchrotron. A synchrotron is a cyclic particle accelerator in which the accelerating particle bean travels a fixed closed-loop path. The magnetic fields which bend the beam into its path increase in strength in synchrony with the kinetic energy of the particles in the beam. Charged particles of energy E and mass m radiate synchotron radiation at a rate proportional to $\left(\frac{E}{m}\right)^4 r^{-2}$, imposing a practical limit on acceleration [95]. This is a reason why the LHC uses relatively massive particles (protons as opposed to electrons) and has a relatively large circumference.

The LHC was built between 1998 and 2008 at CERN in the 27 km Large Electron-Positron Collider (LEP) circular tunnel, which is as deep as 175 m below the surface, with the aim of probing the TeV energy scale[1] [96, 97, 98]. It produces two counter-rotating proton beams which are focused such that they collide at four collision points around its circumference. Each collision of beam particles is referred to as a scattering event, and an aim of experiments associated is to record and analyse interesting-looking events. Collision points are surrounded by detectors designed generally to record as many of the particles emerging from a collision point as possible, measuring their momenta and charges. Some particles pass through the detector and are regarded as missing 4-momentum (a breaking of momentum conservation) in a scattering event.

Some important scientific goals of the LHC are the elucidation of the electroweak symmetry breaking mechanism and the search for the Higgs boson postulated in the Standard Model of particle physics. The LHC is designed to accelerate and collide protons at a centre-of-mass energy of 13.6 TeV, and to achieve an instantaneous luminosity exceeding $10^{34}\,\text{cm}^{-2}\text{s}^{-1}$. Protons are injected into the LHC in counter-rotating bunches. The LHC is designed to run with 2808 bunches per beam, each containing roughly 100,000 million protons with the bunches separated by 25 ns, implying a collision (bunch crossing) rate of about 600 million per second, or 40 MHz.

Stable operation of the LHC began in 2010 at a centre-of-mass energy of 7 TeV and has since increased to 14 TeV. There was a 5 TeV run in November of 2017 during Run-2 which took place over about a

[1] Once called Juratron energies [89].

week [99]. Some LHC design and delivered beam properties are shown in table 3.1 and a landscape picture of the LHC and its detectors is shown in Figure 3.1.

Parameter	Design value	2012	2015	2016
Beam energy (TeV)	7	4	6.5	6.5
Beta function β^* (m)	0.55	0.6	0.8	0.4
Max. bunches per beam	2808	1380	2232	2064
Max. protons per bunch	3.45×10^{-4}	1	1	1
Bunch spacing (ns)	25	50	25	25
Peak luminosity ($cm^{-2}s^{-1}$)	1.0×10^{-34}	7.6×10^{-33}	5.0×10^{-33}	1.0×10^{-34}
Max. average in-time pile-up	19	36.2	28.1	34.4

Table 3.1: Design and delivered beam properties of the LHC, from Run-1 to Run-2, to July 2016 (with years before 2012, called Run-1, omitted because this was to a large extent for LHC commissioning).

Figure 3.1: The LHC and its detectors, an illustration by Max Degtyarev [100].

Beyond Higgs bosons, LHC experiments have observed more than 50 new hadrons. In the 1960s, there were many hadrons observed and the quark model accounted for the composition of these baryons and

mesons. Even in Gell-Mann's earlier papers on the quark model, the notion of particles containing more than three quarks was acknowledged as a possibility. Today such particles have been observed, but it took decades to observe the first four-quark and five-quark hadrons, tetraquarks and pentaquarks.

59 new hadrons observed at LHC experiments are shown in Figure 3.2. These particles are comprised of pentaquarks, tetraquarks and new excited states of baryons and mesons. Observation and study of these particles tests the limits of the quark model and empowers a better understanding of the strong interaction.

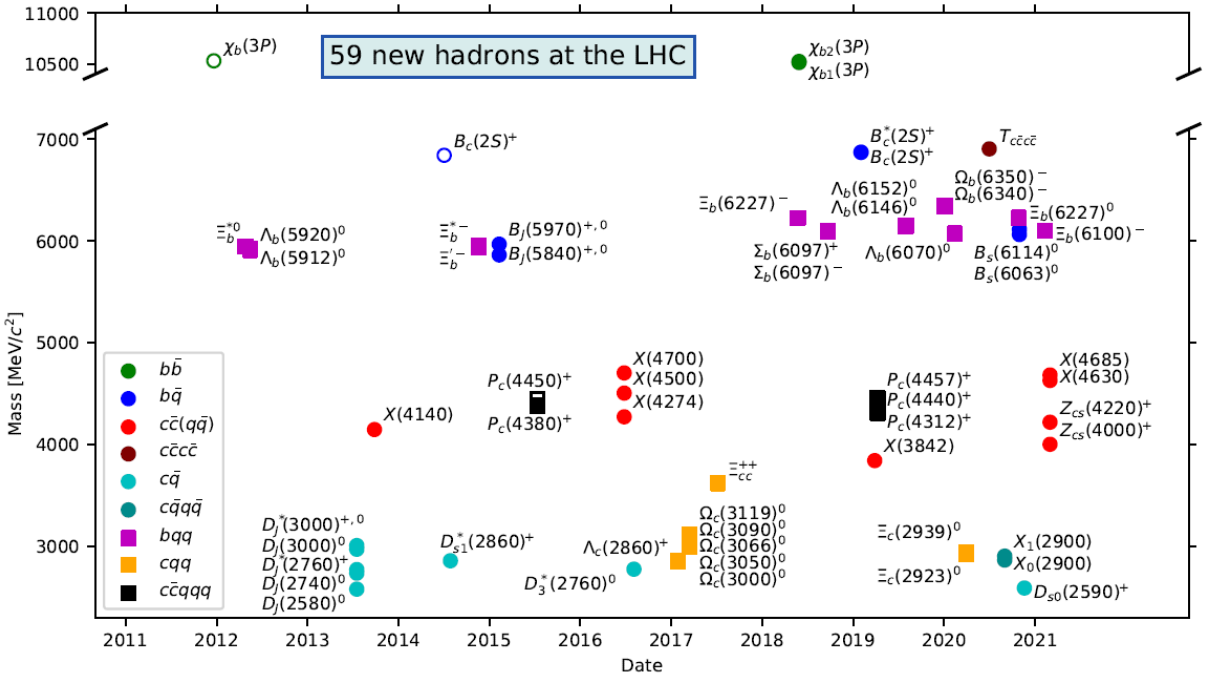


Figure 3.2: New hadrons observed by LHC experiments [101].

3.1.1 General description of LHC operation

The high energies of the LHC are achieved through the use of a series of accelerators connected in series, culminating with the LHC, as depicted in the CERN accelerator complex shown in Figure 3.3.

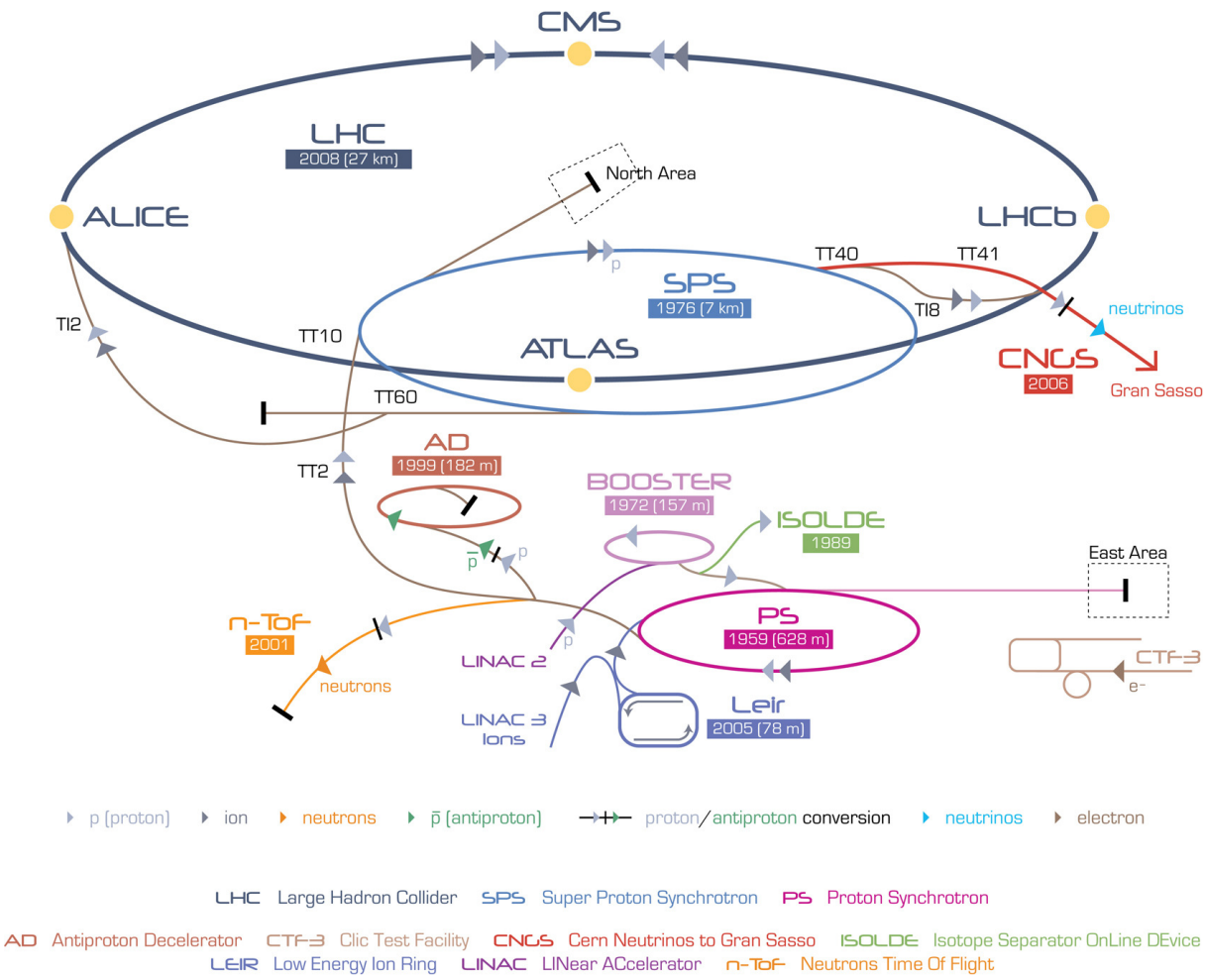


Figure 3.3: CERN accelerator complex [102].

During Run-2 of the LHC, protons from ionised hydrogen gas from "the bottle" are accelerated in the 33.3 m long accelerating radiofrequency cavities of the LINAC 2, which sends the protons in bunches to the Proton Synchrotron Booster (PSB) at an energy of 50 MeV. The PSB separates the protons into four stacked circular synchrotron rings, each 157 m in circumference, and sends them to the Proton Synchrotron (PS) at an energy of 1.4 GeV. The PS, a 628 m circumference synchrotron built from 277 magnets, forms the protons into bunch trains and sends them to the Super Proton Synchrotron (SPS) at an energy of 25 GeV. The SPS, a larger 7 km circumference synchrotron built from 1,317 magnets, accelerates the protons up to an energy of 450 GeV, which is the LHC injection energy. The SPS injects the proton beams in opposing directions in the LHC.

The LHC accelerates the proton beam around its 27 km circumference with its 1,232 superconducting cryodipole magnets and focuses the beam with 392 quadrupole magnets (and many other magnets for precise control of the beam). A cross-section of the LHC beam pipe is shown in Figure 3.4, while a three-dimensional representation of a cryodipole is shown in Figure 3.5. The cryodipole magnets keep particles on an approximately circular orbit, the radiofrequency cavities boost particles at each turn and the quadrupole magnets limit the divergence of the beam and focus beams on interaction points to obtain high luminosities. The LHC beam pipe is contained in a cryostat at a temperature of 1.9 K, facilitating the functioning of the superconducting magnets. Today, the LHC can accelerate each counter-rotating beam to 6.8 TeV. Colliding these beams results in a centre-of-mass energy of 13.6 TeV. The data with which the search presented in this thesis is concerned is Run-2 data up to a centre-of-mass energy of 13 TeV.

For a video which illustrates these accelerations clearly and with exciting music, see Reference [103].

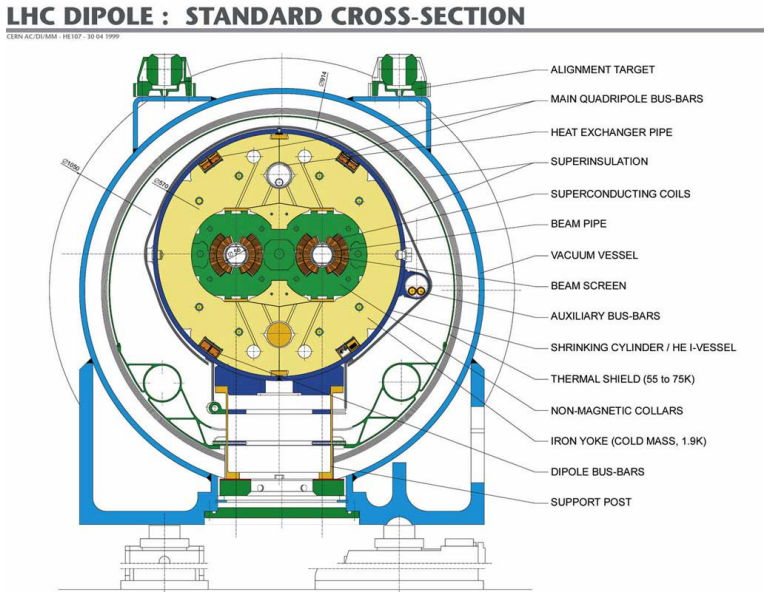


Figure 3.4: Cross section of LHC superconducting cryodipole [104].

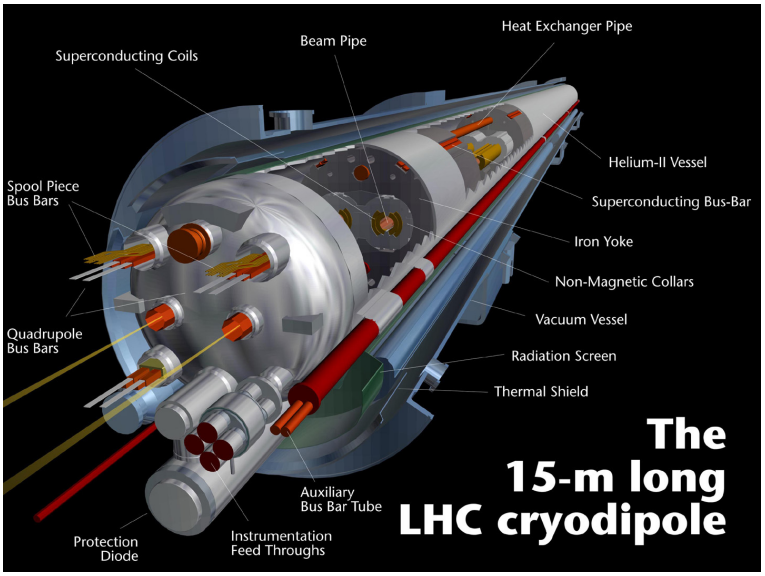


Figure 3.5: Three-dimensional representation of LHC superconducting cryodipole [105].

The beam areas can be represented as ellipses. The beam size is expressed in terms of transverse emittance ϵ and amplitude function β, with the beam area being $\sigma = \sqrt{\epsilon\beta}$. The emittance is an area containing a certain fraction of beam particles, so it is an expression of the beam confinement. The amplitude function is effectively the aspect ratio of the beam area ellipses, with a lower β corresponding to a more narrow beam and a higher β corresponding to a wider one. In order to maximise the rate of interactions, the beam size is minimised ($\beta^\star$) at the interaction point.

One injection and depletion of the LHC beam is called a fill. It involves broadly the following modes [106]:

1. Setup: this is a pre-injection setup.
2. Cycling: this is a pre-cycle before injection. The goal of this mode is to reset the magnetic history of the LHC and to prepare for a new cycle.
3. Injection: counter-rotating beams are injected ($\beta^\star = 30$ cm).
4. Prepare ramp: the injection is complete and preparations are made to accelerate the beams.
5. Ramp: the beams are gradually accelerated.
6. Flat top: the beams have reached the planned energy.
7. Squeeze: the beams are gradually made narrower ($\beta^\star = 25$ cm).

8. Stable beams: the beams are stable in the planned configuration. This is the condition under which data-taking for physics analyses usually happens. While in practice there can be occurrences that disrupt this stable condition, it can be maintained usually for a few hours before the beams are exhausted.

9. Dump: the beams are deflected out of the LHC into absorption material by the LHC Beam Dumping System. LHC beams store a lot of energy so this system is critical for safety and is used in controlled conditions and in unstable conditions such as some of the beam heating part of a superconducting magnet (a change called a quench).

10. Ramp down: there is a ramp down and there is cycling after the beam dump.

In the ATLAS Control Room, these various modes are alerted using sound effects ranging from the Starship Enterprise engaging warp speed to a toilet flushing. These various modes can be seen during the start of Run-2 in 2015 in the video of Reference [107]. A snapshot of an injection as depicted on OP Vistars LHC Page 1 during the start of Run-2 is shown in Figure 3.6.

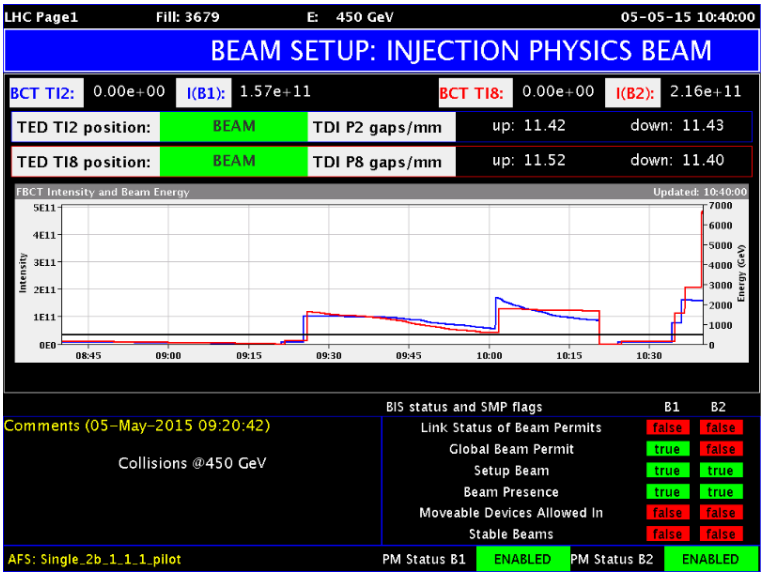


Figure 3.6: Run-2 start, as depicted on OP Vistars [107].

3.1.2 Collisions

One collision is referred to as an event (a scattering event, as discussed in Section 3.1). The outcome of a specific collision cannot be predicted with certainty, and an exhaustive list of possible interaction final states is not calculable. So, the relative probabilities of a variety of different final states, or the rates of

different types of events, are determined.

The cross-section is a measurement of the rate at which a particular process occurs; specifically the probability that an interaction will occur between a projectile particle and a target particle. It is measured by considering an effective area available for that process. In fixed-target experiments, such as the famous Rutherford α-particle scattering experiment, a beam of particles hits a target, and the cross-section is the effective area that the target presents to an incident particle of the beam. A large target with component target particles covering a large fraction of the target corresponds to a high probability of interaction. The cross-section is measured in terms of the unit of area called the barn, which is 10^{-28} m^2. At Los Alamos, Holloway and Baker defined the term barn, with a cross-section of 10^{-24} m^2 for nuclear processes felt to be "as big as a barn" [108]. Given that collider physics research has been ongoing for some time now, the cross-sections for interesting processes tend to be very small.

The Lorentz contraction of the high speed protons results in effectively flat, oval proton bunches colliding in the centre-of-mass reference frame. The bunches of particles in the LHC beam arrive at the interaction point with a certain frequency, so the collision rate depends on the numbers of particles in the bunches, the collision frequency and the size of the bunches. The smaller the bunch, the greater the packing of the constituent particles and, thus, the higher the rate of collision.

As described in Chapter 2, proton-proton collisions are effectively the interaction of partons (quarks and gluons) from each proton. Two protons can interact elastically or inelastically, analogous to collisions described by classical mechanics. For the case of elastic collisions, which occur more frequently, protons do not dissociate and there are no observable new particles created and there is no energy lost. For the case of inelastic collisions, one or both protons change energy and direction and dissociate, with the creation of new particles. Contemporary high-energy physics is concerned primarily with inelastic collisions. There are broadly two categories of inelastic collisions, non-diffractive, wherein partons from two protons interact directly, leading to the creation of new particles, and diffractive, wherein protons dissociate without direct interactions between them. Non-diffractive interactions typically result in outgoing particles spread over the whole volume of the detector while diffractive interactions typically result in outgoing particles close to the parent proton (with diffractive interaction sub-categories of single-diffractive, double-diffractive and central-diffractive possible).

Partons do not exist in isolation; they split into other partons in a process called fragmentation such that they can combine to form colour-neutral hadrons, in a process called hadronisation. Generally, all interactions except for hard scattering are referred to as an underlying event, so it is all the activity of a single proton-proton collision that is not arising from the main parton-parton interaction [109]. In contemporary high-energy physics, the search for unobserved particles takes place mainly in the hard-scattering part of collisions.

Collisions are broadly hard, with large momentum transfer, or soft, with less momentum transfer. Hard and soft collisions can be interpreted in terms of impact parameter (centrality), which is discussed in Chapter 4, and which is a measure of overlap of interacting protons (essentially the fractional areas or volumes of protons that participate in the interactions). Hard interactions tend to be central.

3.1.3 Luminosity

The instantaneous luminosity is a measure of collision rate (strictly, the number of particles put in a position to collide) and it can roughly be thought of as the 'brightness' of the source of the colliding particles. The amount of data collected by an experiment like ATLAS may be quantified by the integrated luminosity, which is the instantaneous luminosity integrated over time. This is a number which, when multiplied by the cross-section of a particular process, yields the expected number of interactions of that type. It is an important number because it is typically a dominant systematic uncertainty.

The total luminosity of collisions recorded by a detector, L, can be considered in terms of the number of events recorded, N, and the total inelastic cross-section, σ, which is the proton-proton inelastic cross-section in for the LHC (or the process under observation – the effective area which governs the probability of some scattering or absorption event,

$$N_{\text{events}} = \sigma_{\text{event}} L = \sigma_{\text{event}} \int \mathcal{L}.\text{dt}, \tag{3.1}$$

where $\mathcal{L}$ is the instantaneous luminosity, itself defined by a number of beam-related parameters:

$$\mathcal{L} = f_r n_b \frac{N_1 N_2}{2\pi \Sigma_x \Sigma_y}, \tag{3.2}$$

where f_r is the beam revolution frequency, n_b is the number of bunches, N_1 and N_2 are the respective numbers of particles in bunches 1 and 2, and Σ_x and Σ_y are the mean beam width in the respective x or y directions.

The integrated luminosity of the collision data collected by ATLAS over the years of LHC operation is shown in Figure 3.7.

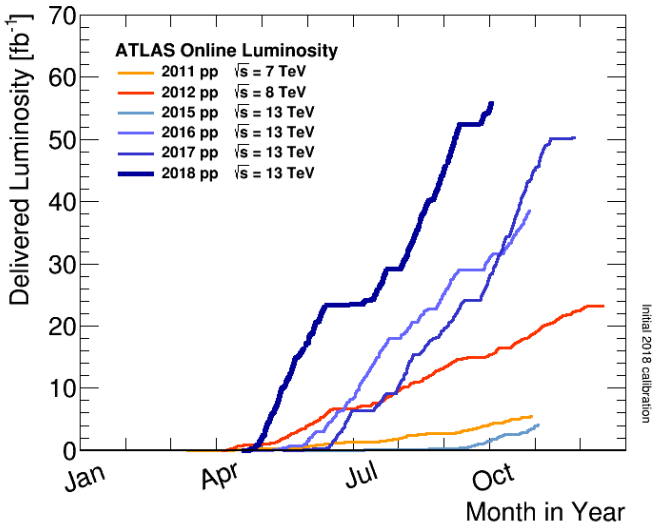


Figure 3.7: Proton-proton collision data collected by ATLAS over the years of LHC operation [110].

The numbers of interactions per crossing for $\sqrt{s} = 13\,\text{TeV}$ from 2015 to 2018 are shown in Figure 3.8. The mean number of interactions per bunch crossing for 2015 and 2016, the years of the data with which this thesis is concerned, is estimated as 24, with a distribution ranging from approximately 6 to 48.

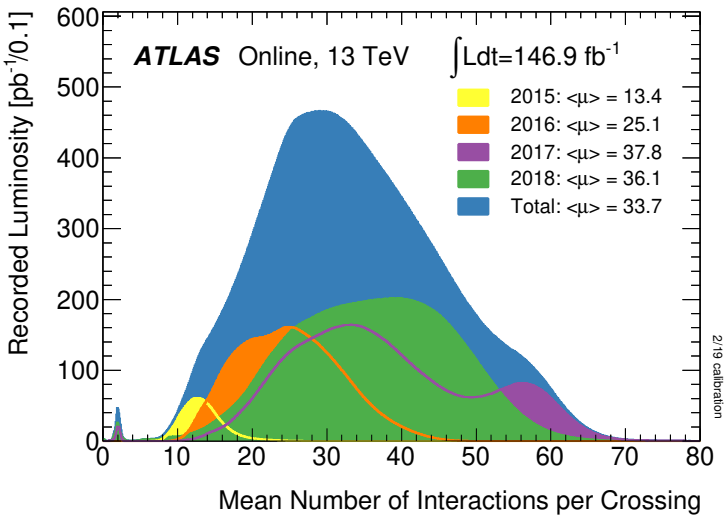


Figure 3.8: Numbers of interactions per crossing for $\sqrt{s} = 13\,\text{TeV}$ from 2015 to 2018 [111].

For the Run-2 2015–2016 datasets of the search presented in this thesis, a total of $36.1 \pm 1.2\,\text{fb}^{-1}$ was recorded.

Integrated luminosity is an important systematic uncertainty. Precise, absolute and relative calibrations of several luminosity-sensitive detectors are performed in order to minimise it [112].

First, a van der Meer (vdM) beam-separation scan is performed for an absolute calibration. In this process, the proton beams are displaced horizontally and vertically, scanning them through each other, for a measurement of the combined size of the proton bunches. Beam current measurements then enable derivation of the total number of protons in each colliding proton bunch. As described in Section 3.2.6.1, the LUCID-2 detector pair is used to infer the mean number of interactions from Cherenkov light and the total number of protons measured from beam current measurements can be related to this in order to set the scale for the absolute luminosity measurement.

Second, this calibration is extrapolated to the number of interactions per bunch crossing (about 20 to 50), and the patterns of bunches change to trains of consecutive bunches with 25 ns spacing, which is measurable again using the LUCID-2 detector pair. A correction is then applied using a track counting algorithm which relates the number of interactions to the number of tracks reconstructed in the Inner Detector of ATLAS, described in Section 3.2.4.

Third, the stability of the LUCID-2 calibration is calculated by comparing the luminosity estimate of LUCID-2 to track-counting measurements using the Inner Detector and various calorimeters over time, as shown in Figure 3.9. A preliminary uncertainty on the Run-2 luminosity estimate is 1.7 %. The full 13 TeV Run-2 combined data corresponds to an integrated luminosity of $139\,\text{fb}^{-1}$, or about 1.1×10^{16} proton-proton collisions.

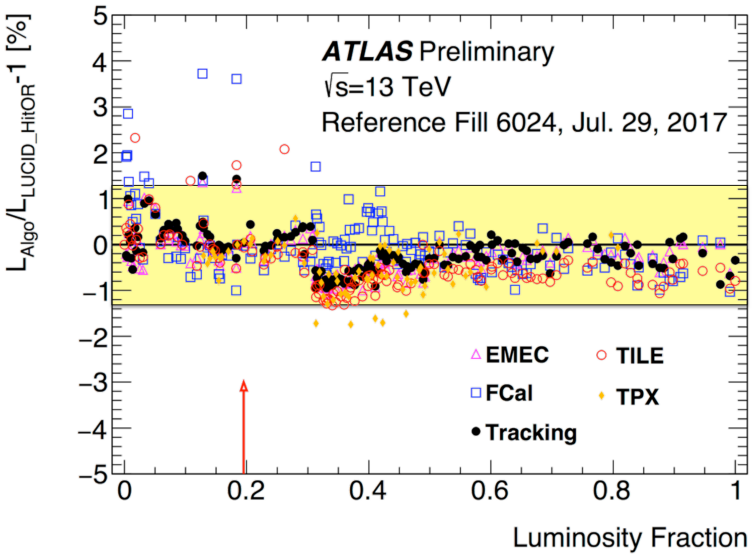


Figure 3.9: Fractional difference in run-integrated luminosity estimates of LUCID-2 compared with track counting and various detector measurements used to quantify the long-term stability of the calibration for uncertainty in the luminosity estimate [112]. The luminosity measurements from the other detectors have been normalised to that of the LUCID-2 measurements in the reference run shown by the red arrow. The assigned ±0.8 % long-term stability is shown by the yellow band.

3.1.4 LHC experiments

There are four main experimental detectors located around the circumference of the LHC. A Large Ion Collider Experiment (ALICE) is a detector designed to observe heavy ion collisions (from lead ion collisions) for the main purpose of study of the quark-gluon plasma[2] [114, 115]. Large Hadron Collider Beauty (LHCb) is a detector designed for the study of CP violation and rare decays in the b and c sectors [116, 117]. The Compact Muon Solenoid (CMS) is a general purpose detector designed to explore physics at the TeV scale, to search for the Higgs boson, and to search for physics beyond the Standard Model, such as supersymmetry (SUSY) and extra dimensions [118, 119]. As a general purpose detector, CMS is a cross-check on ATLAS, which is described in detail in Section 3.2. A variety of experiments were proposed for the LHC initially. Some were not implemented while others merged and advanced. L3P would have been an evolution from the L3 experiment of the LEP for precision lepton and photon measurements far from the interaction point, COBEX would have been an open-geometry forward detector and GAJET would have involved a gas jet target in an LHC beam. ATLAS was formed

[2] Alternatively referred to as *ylem* by Gamow *et al.* [113].

from the merging of the ASCOT (Apparatus with Super COnducting Toroids) and EAGLE (Experiment for Accurate Gamma, Lepton and Energy measurements) detector concepts, both of which involved a toroidal magnet configuration for the Muon Spectrometer, presented at the 1992 Évian-les-Bains meeting *Towards the LHC Experimental Programme* [120, 121, 122].

3.2 ATLAS

3.2.1 General description

ATLAS is a multipurpose detector with a forward-backwards symmetric cylindrical geometry, and is designed to exploit the full potential of the LHC proton-proton collision programme, with a large discovery potential, and for precision measurements [123, 124, 125]. With a length of 46 m and a diameter of 25 m, and a weight of 7,000,000 kg, it is the largest general purpose particle detector. ATLAS is capable of accurate measurements of charged leptons, largely hidden particles (through missing energy measurements), hadronic jets, b-quarks and photons, and is capable of reconstruction of complete final states. ATLAS features three main detector systems: a tracker, a calorimeter and a muon spectrometer, as shown in Figure 3.10.

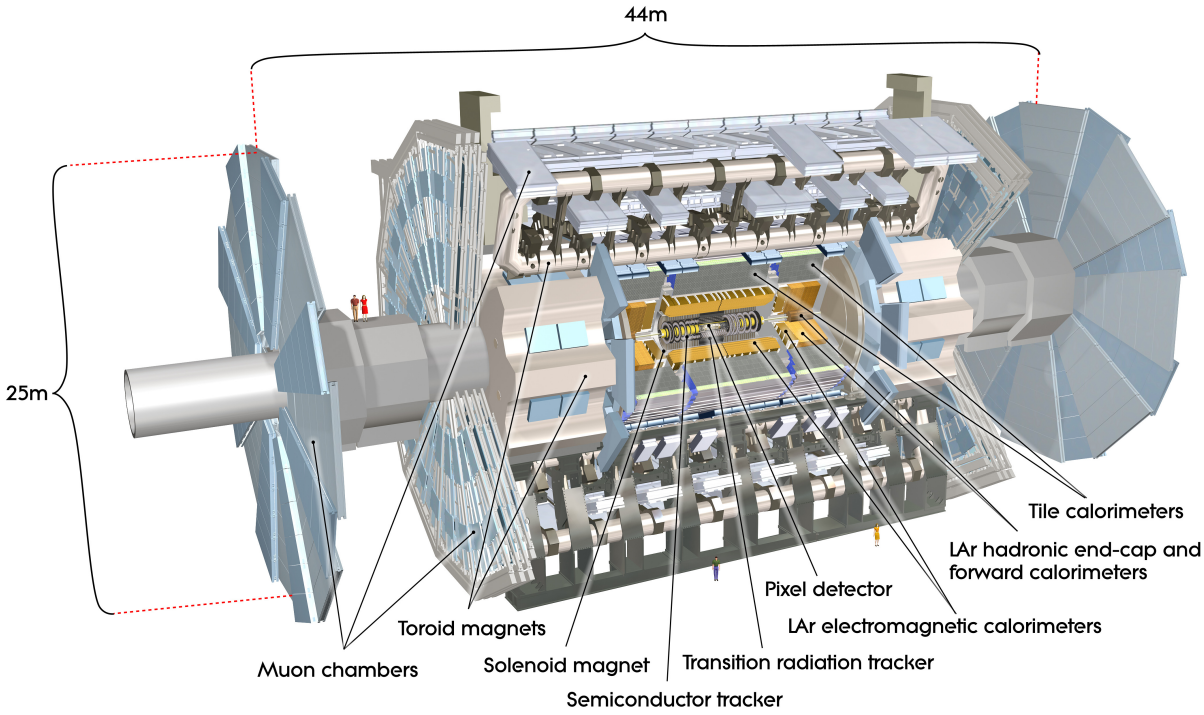


Figure 3.10: ATLAS detector [126].

3.2.2 ATLAS coordinate system

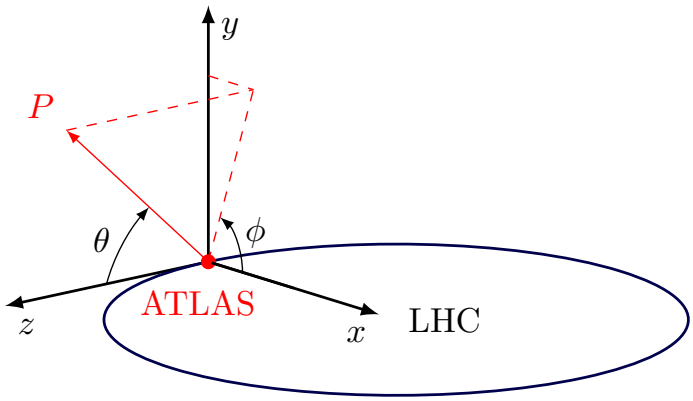


Figure 3.11: ATLAS coordinate system.

The coordinate system used for ATLAS is shown in Figure 3.11. It defines the origin as the centre of the detector at the nominal interaction point. The z axis is parallel to the beam line with the positive sense in the anticlockwise direction. The positive sense of the x axis points from the origin to the centre of the LHC. The positive sense of the y axis points upwards (perpendicular to the x and z axes).

The coordinate system is shown overlaid on an isometric projection of ATLAS in its underground chamber in Figure 3.12.

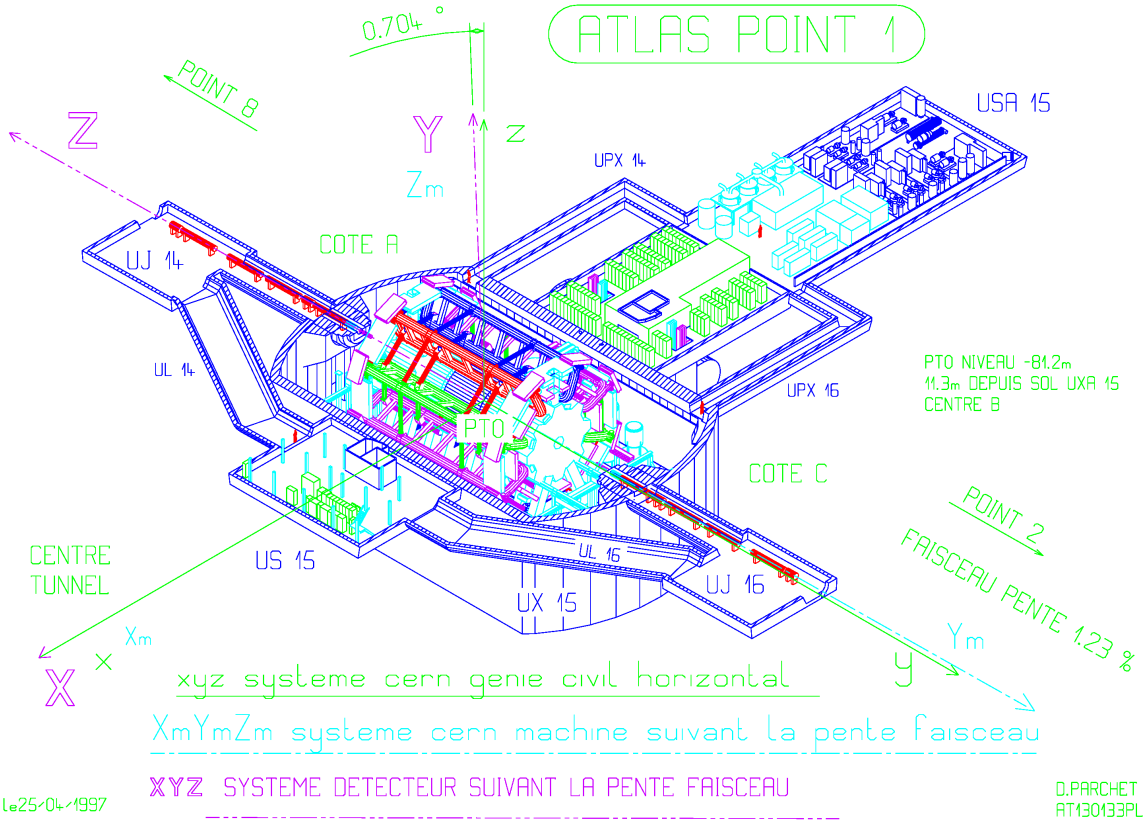


Figure 3.12: ATLAS coordinate system with ATLAS in its underground chamber [127].

ATLAS has an approximate cylindrical geometry, which is described using the (R, ϕ, z) coordinates. The polar angle θ is defined as the angle measured from the beam axis in the xz plane. The azimuthal angle ϕ is defined as the angle measured in the xy plane around the z axis (clockwise when viewing in the positive sense of the z axis). The Lorentz invariant measure of position in the detector known as pseudorapidity η is defined for relativistic particles in terms of the polar angle θ:

$$\eta = \ln \tan \left(\frac{\theta}{2} \right). \tag{3.3}$$

Transverse energy E_{T}, momentum p_{T} and missing transverse energy $\not{E}_{\mathrm{T}}$ are defined in the xy plane. Distances between objects in the detector can be considered in the $\eta\phi$ plane:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}, \tag{3.4}$$

where $\Delta\eta$ is the difference in pseudorapidity and $\Delta\phi$ is the difference in azimuthal angle of two objects under consideration.

3.2.3 Magnet system

The ATLAS magnet system is used to measure the momentum of charged particles because the magnetic fields cause the paths of the charged objects to bend with the curvature of the path correlated to the momentum [125]. So, measuring the degree of curvature of the path made by charged particles gives the momentum of the particle for a known field strength. It is composed of four superconducting magnets: the solenoid, the barrel toroid and two end-cap toroids. When powered they have a stored energy of 1.6 GJ.

The solenoid generates a uniform 2 T axial magnetic field aligned along the beam axis surrounding the Inner Detector. The barrel toroid generates a toroidal magnetic field of 0.5 T for the central muon detectors using eight separate coils. The end-cap toroids produce a magnetic field of 1 T for the end-cap regions of the muon spectrometer using eight flat, square coils.

The configuration of a toroid muon magnet system complimented by the inner superconducting solenoid was chosen because it provided few constraints on calorimetry and the Inner Detector, allowing non-compromised technological solutions, and a large acceptance and robust standalone muon detection system. Precise measurement of the magnetic field produced by the ATLAS solenoid was achieved using a special field mapping machine [128].

3.2.4 Inner Detector

The ATLAS Inner Detector (ID) is used to measure the position of charged particles in the centre of ATLAS in order to determine their tracks [125]. It is composed of three sub-detectors: the Pixel Detector (PD), the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT), as shown in Figure 3.15.

As shown in Figure 3.13, the Inner Detector starts a few centimetres from the beam and extends to 1.15 m, providing full tracking over $|\eta| < 2.5$. It is 7 m long and is composed of a barrel section that covers ± 80 cm along the beam and two end-caps. In the barrel region, the high precision detectors are mounted on concentric layers around the beam. The structure of the Inner Detector can be seen by plotting secondary-vertex positions following some selection criteria, as shown in Figure 3.14.

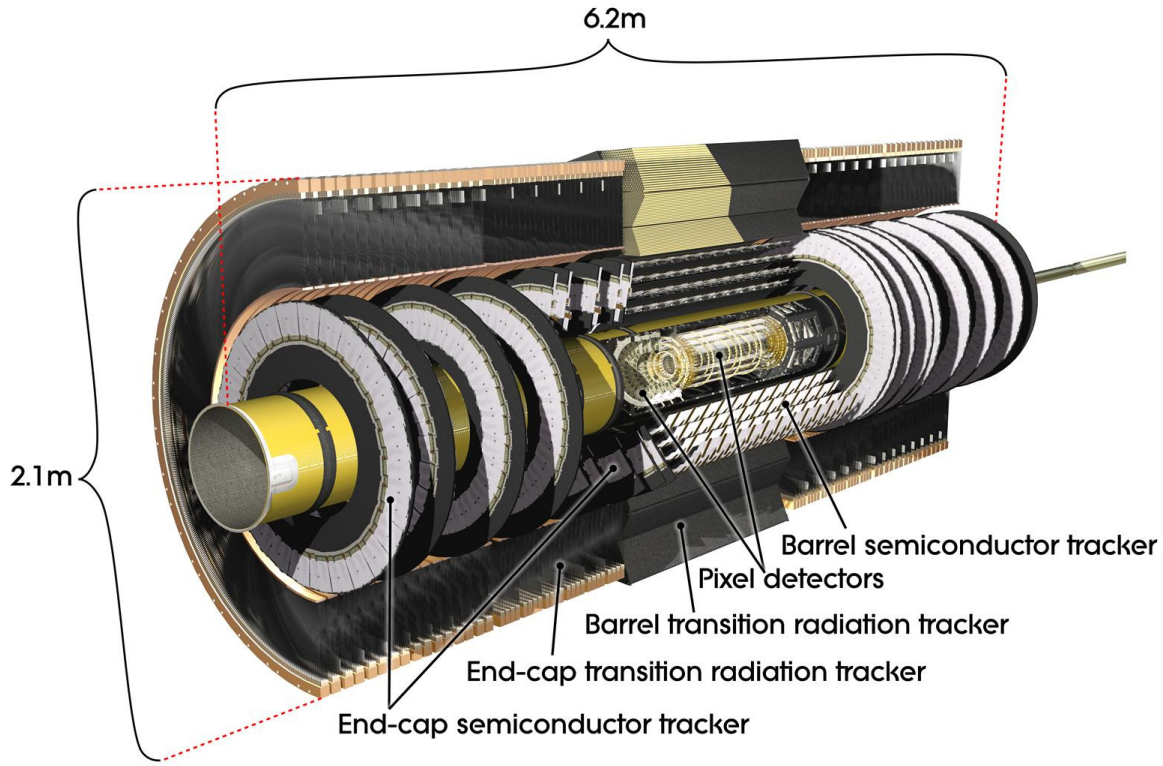


Figure 3.13: ATLAS Inner Detector [129].

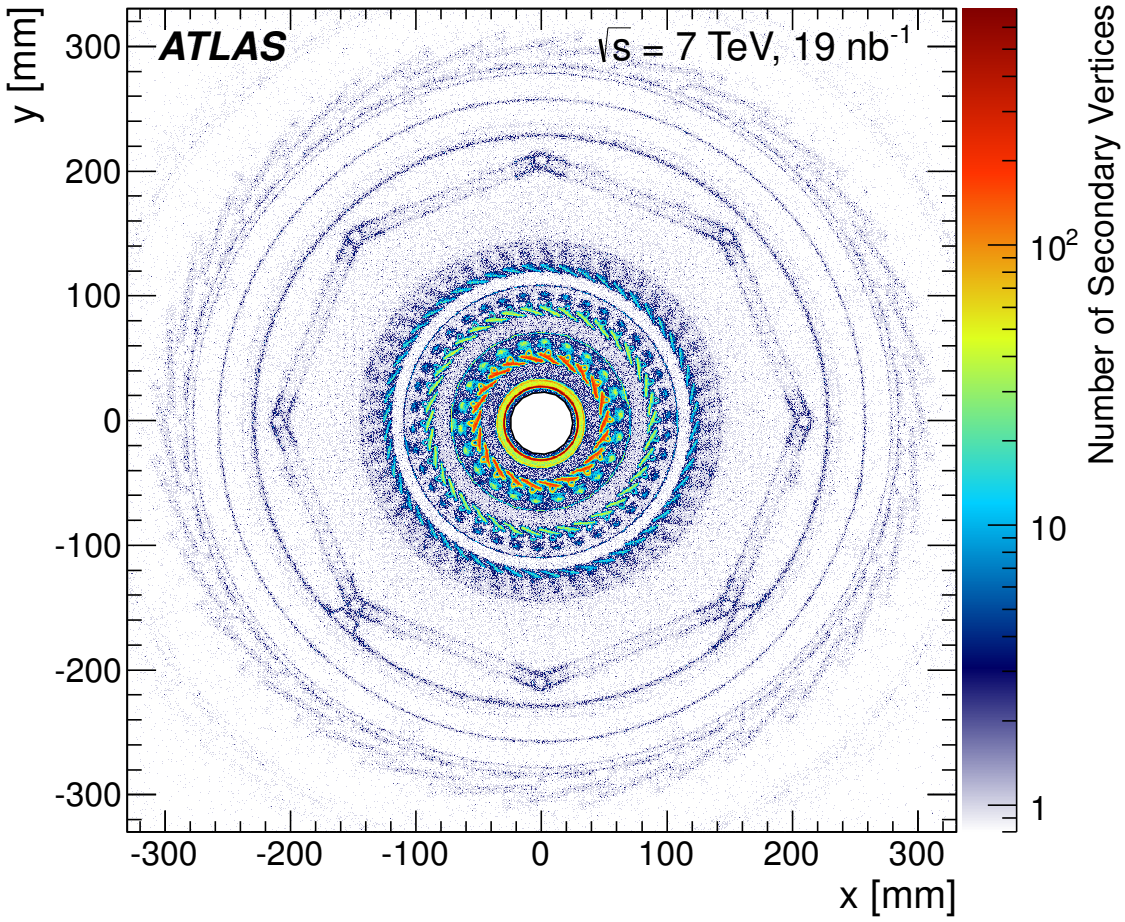


Figure 3.14: Secondary-vertex positions in the ATLAS Inner Detector [130].

In performing tracking, precision measurements of particle momentum above 0.5 GeV are made for charged tracks between the beam pipe and the electromagnetic calorimeter. Secondary vertex identification using reconstructed tracks and impact parameters are used to indicate the presence of short-lived particles such as tauons and b-quarks.

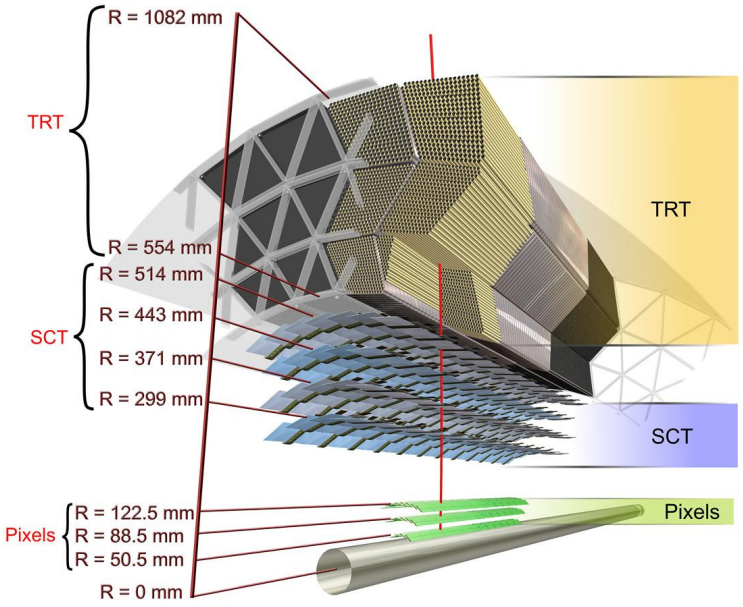


Figure 3.15: Cross section of ATLAS Inner Detector [129].

The Inner Detector can measure the position of transiting charged particles to better than 100^{th} of a millimetre. In order to achieve that precision, the detector must be aligned to equal or greater accuracy. During LHC collisions, it was observed that parts of the Inner Detector showed evidence of short-timescale movements. ATLAS components can move due to temperature fluctuations and changes in magnetic field strength and these movements are particularly pronounced in the Inner Detector. Vertical movement of the Pixel Detector from the start of an LHC fill is shown in Figure 3.16. In order to address these movements, ATLAS uses an automated dynamic alignment scheme for the Inner Detector, which corrects data recorded by ATLAS [131]. During the first hour of data-taking–when there are significant thermal changes–alignment constants are derived every 20 minutes and every 100 minutes thereafter.

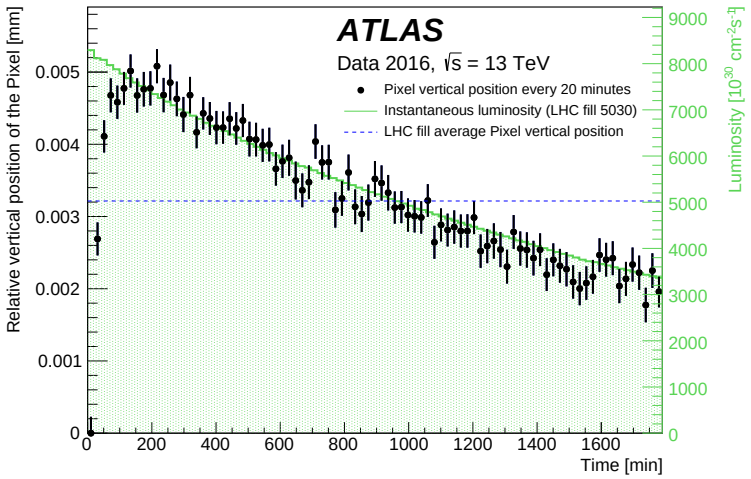


Figure 3.16: Vertical movement of the Pixel Detector as a function of time since the start of an LHC fill [131].

3.2.4.1 Pixel Detector and Semiconductor Tracker

High granularity is required close to the beam pipe as this is where particle track density is greatest. The Pixel Detector (PD) and Semiconductor Tracker (SCT) are precision tracking detectors designed for this purpose [125]. In the barrel region they are concentric cylinders around the beam, and in the end-caps they are discs perpendicular to the beam, as shown in Figure 3.13.

The Pixel Detector is the innermost part of the Inner Detector and provides precise tracking close to the nominal interaction point, with each track transiting approximately three pixel layers. This provides three points in space for reconstruction of the track. The first layer of the Pixel Detector is the vertexing layer, which serves to resolve the secondary vertices of short lived particles.

The Insertable B-Layer (IBL) is a fourth layer that was added to the Pixel detector between the beam pipe and the inner Pixel layer (B-layer) in 2014 for Run-2 [132]. This layer, shown in Figure 3.17, is suited to b-tagging and supplements the B-layer of the Pixel detector. It is beneficial because it provides robustness against radiation damage to b-tagging capabilities of ATLAS (which are essential for the analysis presented in this thesis) and it provides better tracking capabilities for ATLAS because it is close to the beam.

The IBL was integrated into the detector by shrinking the diameter of the beam pipe and inserting the layer into the gap between the Pixel Detector and the beam pipe, making it effectively the innermost layer of the Pixel Detector. The pixel size was set to 50 by 250 µm, a CO_2 cooling system was introduced and

new carbon foam structures were invented to support the modules that make up the IBL. A plot of the occupancy readout of the IBL during commissioning is shown in Figure 3.18 and a plot of the occupancy readout of the IBL during stable beams with 13 TeV proton-proton collisions is shown in Figure 3.19.

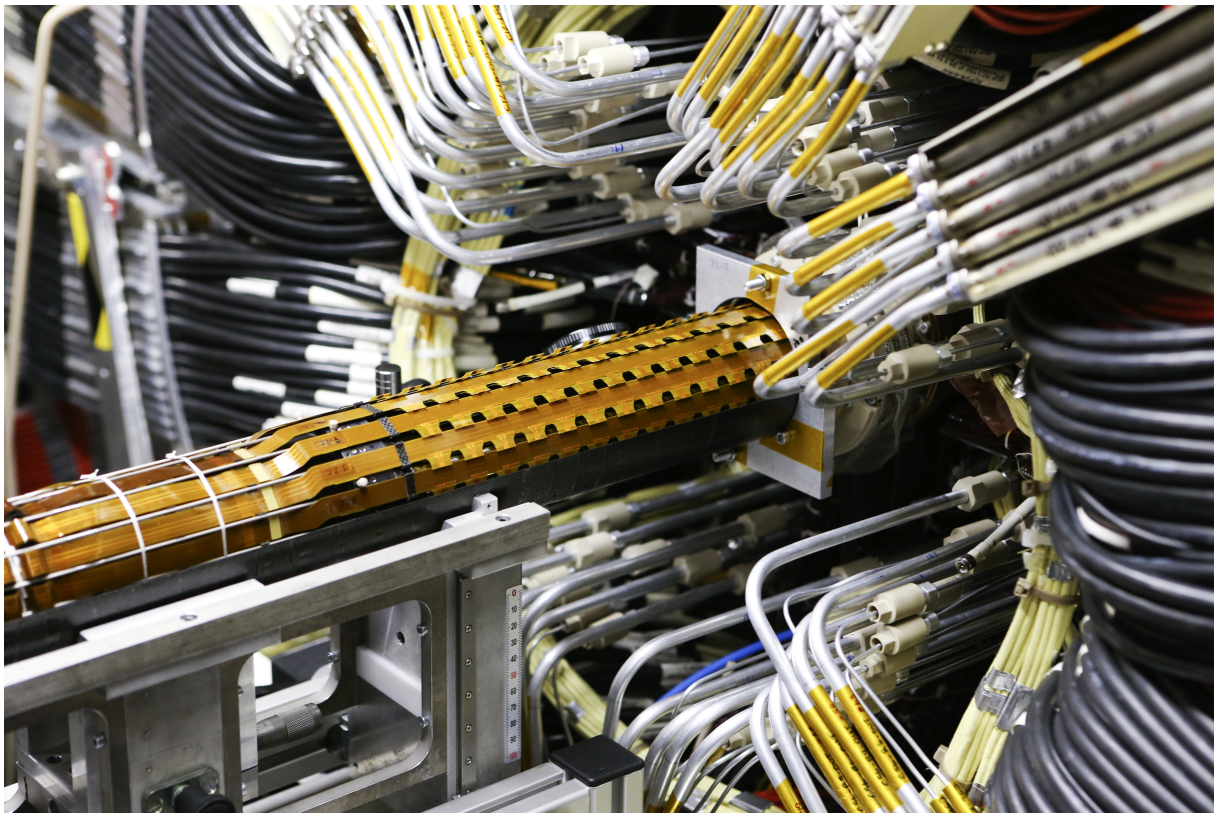


Figure 3.17: Insertable B-Layer (IBL) of ATLAS Inner Detector [133].

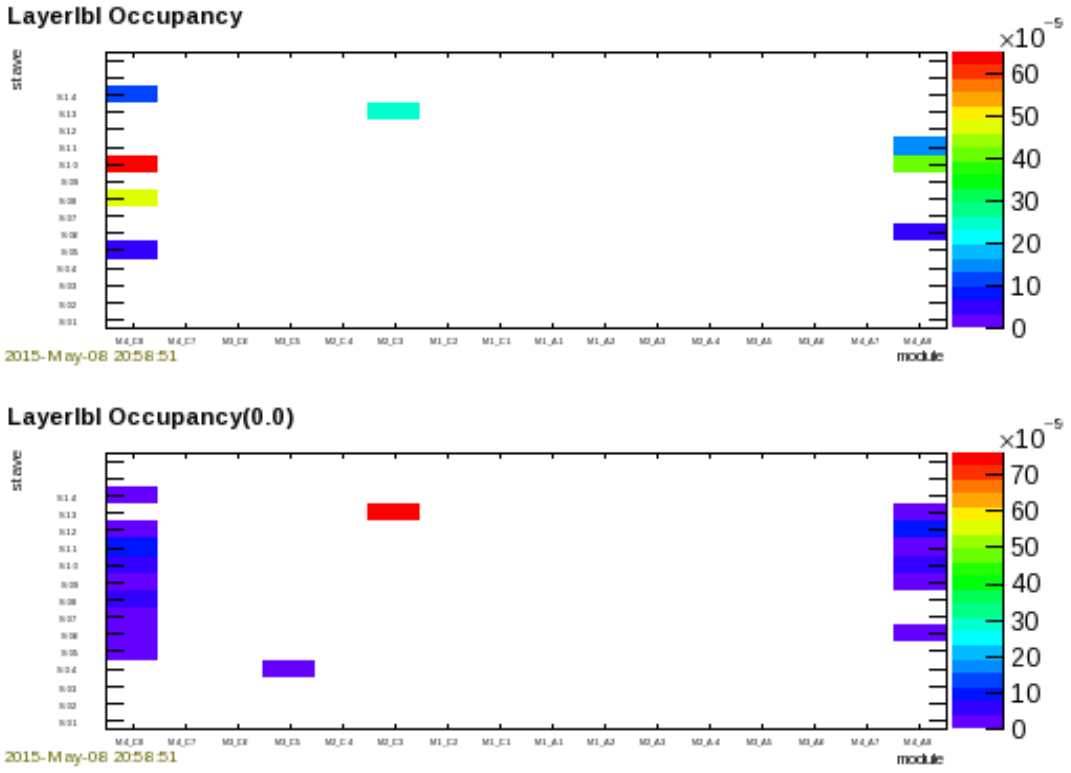


Figure 3.18: Occupancy readout from IBL during commissioning and run 264336 [134].

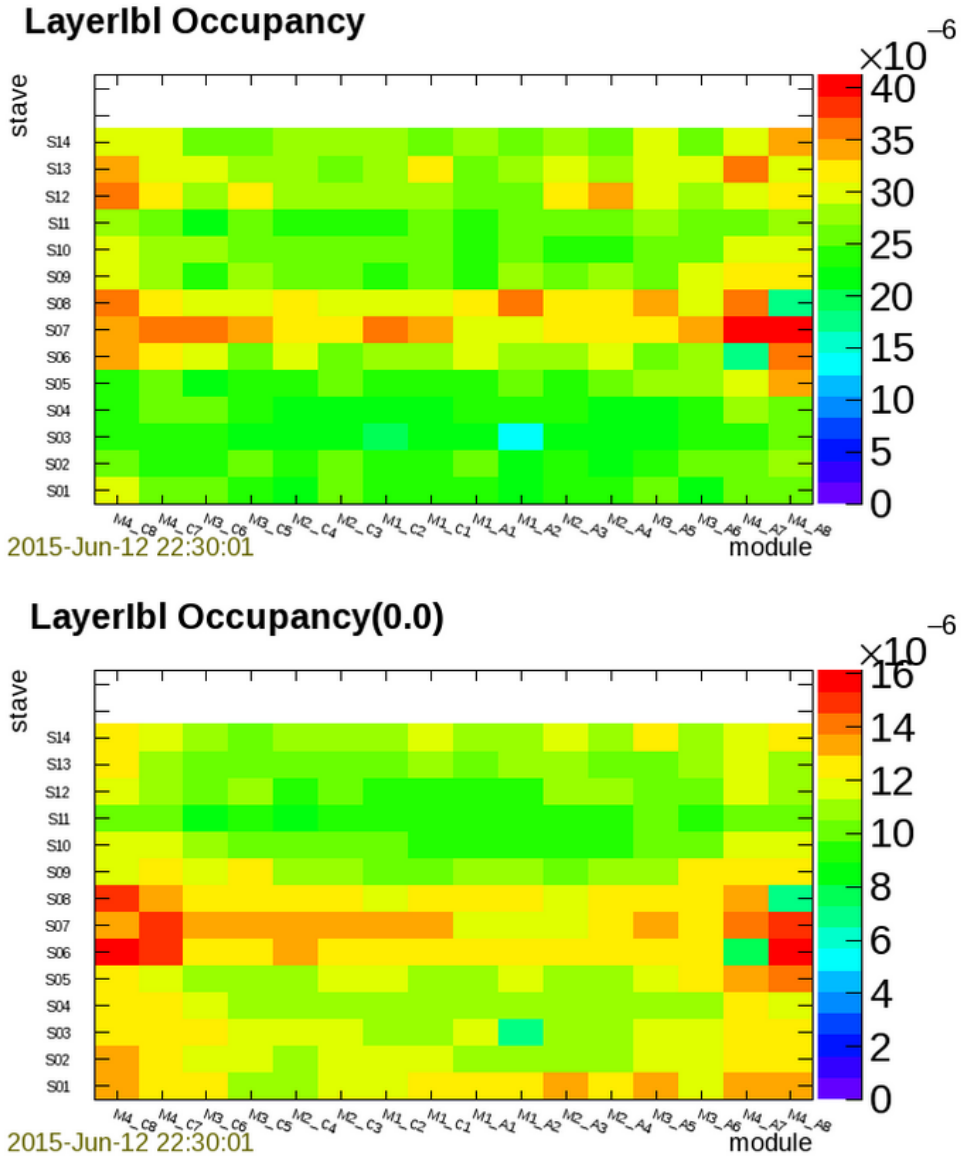
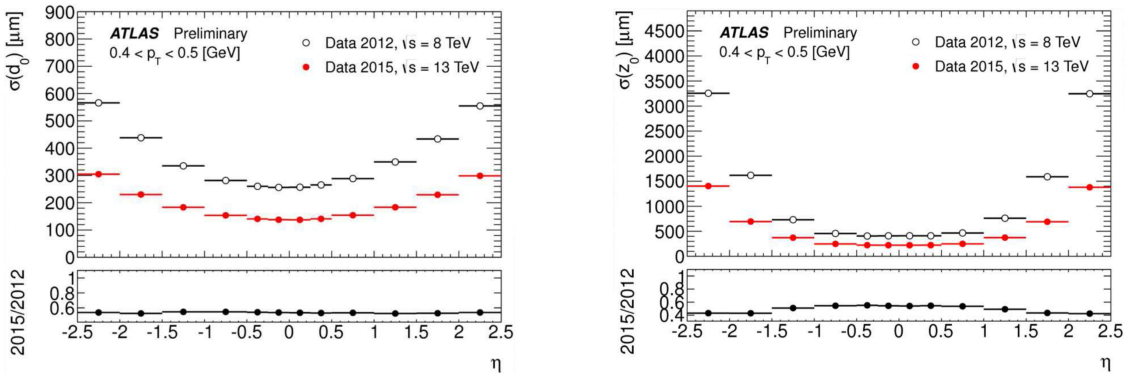


Figure 3.19: Occupancy readout from IBL during ATLAS run 267599 with stable beams and 13 TeV proton-proton collisions [135].

A variety of concerns about the IBL were monitored extensively and continuously during its commissioning. One concern arose from observations of the IBL being at a greater temperature for the same power

usage on a subsequent day, with the mean IBL thermal figure of merit (TFM) losing linearity following the prior day. Actions such as taking the IBL out of usage were intended to protect the thermomechanical integrity of the IBL modules from irreversible damage. Ultimately the temperature issue was identified as a CO_2 cooling flow being too high. Another concern arose from observations of a drift in low voltage currents in the IBL frontends, with currents exceeding safe operating parameters. The impact of running the IBL warm for one year was assessed and ultimately the recommendations for responding to the low voltage drift in the context of increased luminosity in Run-2 involved a higher temperature and a lower digital supply.

The IBL resulted in significantly improved performance. Comparisons of impact parameter resolutions of Run-2 (with the IBL) to impact parameter resolutions of Run-1 (without the IBL) are shown in Figure 3.20.



(a) Transverse impact parameter

(b) Longitudinal impact parameter

Figure 3.20: Comparisons of impact parameter resolutions of Run-2 (with the IBL) to impact parameter resolutions of Run-1 (without the IBL) [136].

Including the IBL, the Pixel Detector enables tracking starting at 33.25 mm from the nominal interaction point, allowing discrimination of primary and secondary vertices through precise measurement of the origin of tracks.

The Semiconductor Tracker is the middle component of the Inner Detector and is similar in concept and function to the Pixel Detector but is composed of silicon microstrips instead of silicon pixels. The Semiconductor Tracker provides basic tracking in the plane perpendicular to the beam. It starts at

299 mm from the nominal interaction point.

3.2.4.2 Transition Radiation Tracker

The Transition Radiation Tracker (TRT) is the outermost component of the Inner Detector and is a combination of a straw tracker and a transition radiation detector [125]. The detecting elements are drift tubes (straw tube detectors). The tubes are made of two multi-layered films attached back-to-back, designed to give good electrical and mechanical properties while maintaining a very thin cross-section, with a thin, gold-plated tungsten wire in the centre of the tubes, making capacitors, with the wire forming the anode and the tube forming the cathodes. The tubes are filled with a gas (composed mostly of xenon or argon) that is ionised when charged particles pass through it. When a charged particle transits, the ions from the gas are attracted to the wire, creating a current in the wire. As the magnetic field of the tube is known, the distance from the wire can be calculated using the time the ion takes to drift to the wire.

Transition radiation photons are emitted when a highly relativistic charged particle traverses through a boundary between two media with different dielectric constants. The quantity of transition radiation is greatest for highly relativistic particles, so, for a particle of known energy the radiation is dependant on the mass of the particle. To provide such boundaries, the Transition Radiation Tracker straws are interleaved with fibres in the barrel region and foils in the end-caps. This enables some differentiation between electrons and pions.

3.2.5 Calorimeters

Calorimeters in contemporary high-energy physics detectors often are composed of layers of lead and plastic scintillator sheets. Broadly, particles interact with lead nuclei and create a showering cascade of secondary interactions, with resulting particles generating light as they traverse the scintillator sheets. By measuring the light scintillated, the energies of the originating particles can be inferred. Calorimeters are important components in contemporary high-energy particle physics detectors for a number of reasons. They are sensitive to charged and neutral particles produced in interactions. Their performance generally improves with increasing LHC energies, in contrast to other detector components which often are less effective with increasing energy. Calorimeters are relatively fast at providing information needed for decisions on recording physics events. Specifically, they provide information on energy flow in events in a relatively direct way, and estimations of such information as missing energy, transverse energy and jet production are particularly useful in assessing events.

The calorimeters of ATLAS are located outside the solenoid, as shown in Figure 3.21. They measure the energy of particles. This is accomplished through the use of materials that increase the probability of the

particle decaying. This results in a shower of many successive particle decays leaving a large number of low energy particles. These low energy particles are then measured and the sum of the energy gives the energy of the original particle.

Information such as the width of the shower and its depth in the detector is used to determine whether the incident object is a jet or a lepton. To aid in this determination, the calorimetry in ATLAS is accomplished using two main calorimetry systems, the Electromagnetic Calorimeter (EMCAL) and the Hadronic Calorimeter (HCAL).

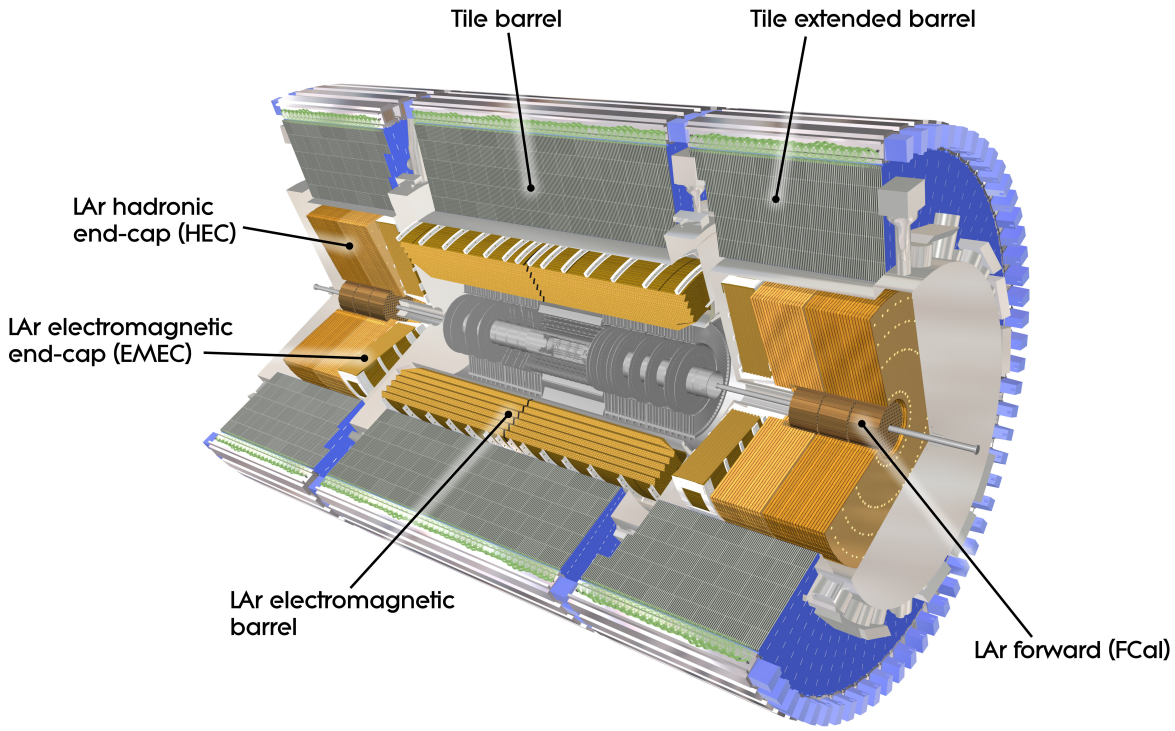


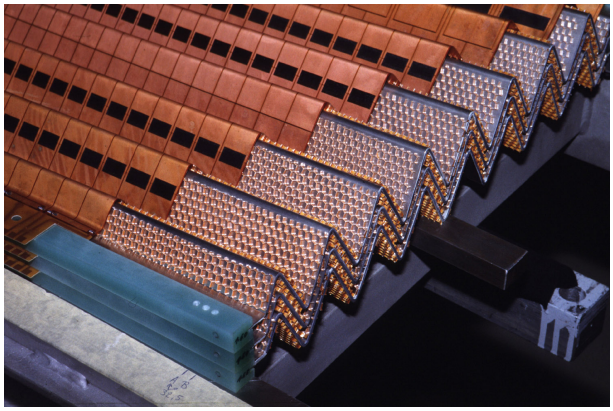
Figure 3.21: Calorimetry in ATLAS.

Together, the Electromagnetic Calorimeter and the Hadronic Calorimeter cover out to $|\eta| \leq 4.9$.

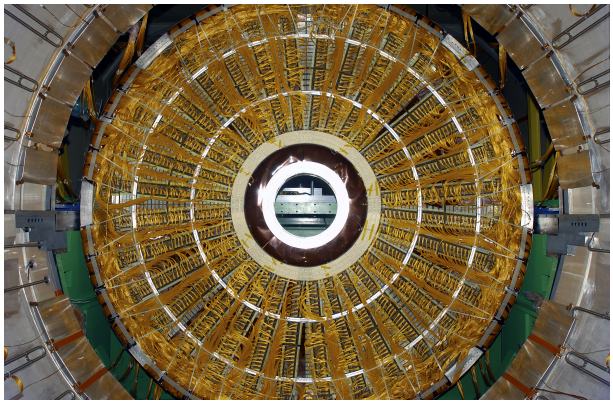
3.2.5.1 Electromagnetic Calorimeter

The Electromagnetic Calorimeter is a high granularity sampling calorimeter used to measure the energy of charged particles and photons [125, 137]. It is composed of liquid argon sampling calorimeters, where

liquid argon is the active material and lead, copper and tungsten are used as the passive absorber, depending on the location. When charged particles cross the liquid argon gap between electrodes and absorbers, they ionise it. Under the influence of the electric field, the ionised electrons drift towards the electrode, inducing a current that is converted to a digital signal. The Electromagnetic Calorimeter consists of alternating layers of lead, for shower production, and liquid argon, for particle detection, and it covers out to $|\eta| \leq 2.5$. These layers are arranged in an accordion geometry oriented in a radial direction in order to provide full azimuthal coverage, as shown in Figures 3.22a and 3.22b. Essentially, the Electromagnetic Calorimeter absorbs energy from particles that interact electromagnetically, which include charged particles and photons. The whole Electromagnetic Calorimeter is made up of four sub-detectors: the electromagnetic barrel (EMB), the electromagnetic end-cap (EMEC), the hadronic end-cap (HEC) and the forward calorimeter (FCal).



(a) Electromagnetic Calorimeter alternating layers [138].



(b) Electromagnetic Calorimeter end-cap [139].

Figure 3.22: ATLAS liquid argon Electromagnetic Calorimeter.

3.2.5.2 Hadronic Calorimeter

The barrel region of the Hadronic Calorimeter (the Tile Calorimeter) consists of steel with scintillating tiles that sample the energy deposited [125]. The scintillators are instrumented with wavelength-shifting fibres which direct light to photomultiplier tubes. The Hadronic Endcap Calorimeter (HEC) is a sampling calorimeter that consists of copper and liquid argon. Essentially, the Hadronic Calorimeter absorbs energy from particles that pass through the Electromagnetic Calorimeter and interact strongly. These particles are mostly hadrons.

3.2.6 Muon Spectrometer

The Muon Spectrometer is the outermost part of the detector, as shown in Figure 3.23. Due to the containment provided by the calorimeters, aside from weakly-interacting particles like neutrinos, few particles other than muons interact with the muon spectrometer; with no strong interactions and large mass, muons lose energy primarily in ionisation, so they pass through the calorimeters with minimal energy loss. The Muon Spectrometer is in the magnetic field of the superconducting air core toroid magnets which deflects the muons in order to provide information on their charge and momentum. The muon spectrometer can provide accurate momentum information for transverse momenta from about 3 GeV (which is limited by losses in the calorimeter) to 3 TeV, and it has a large acceptance covering a region $|\eta| < 2.7$ [125].

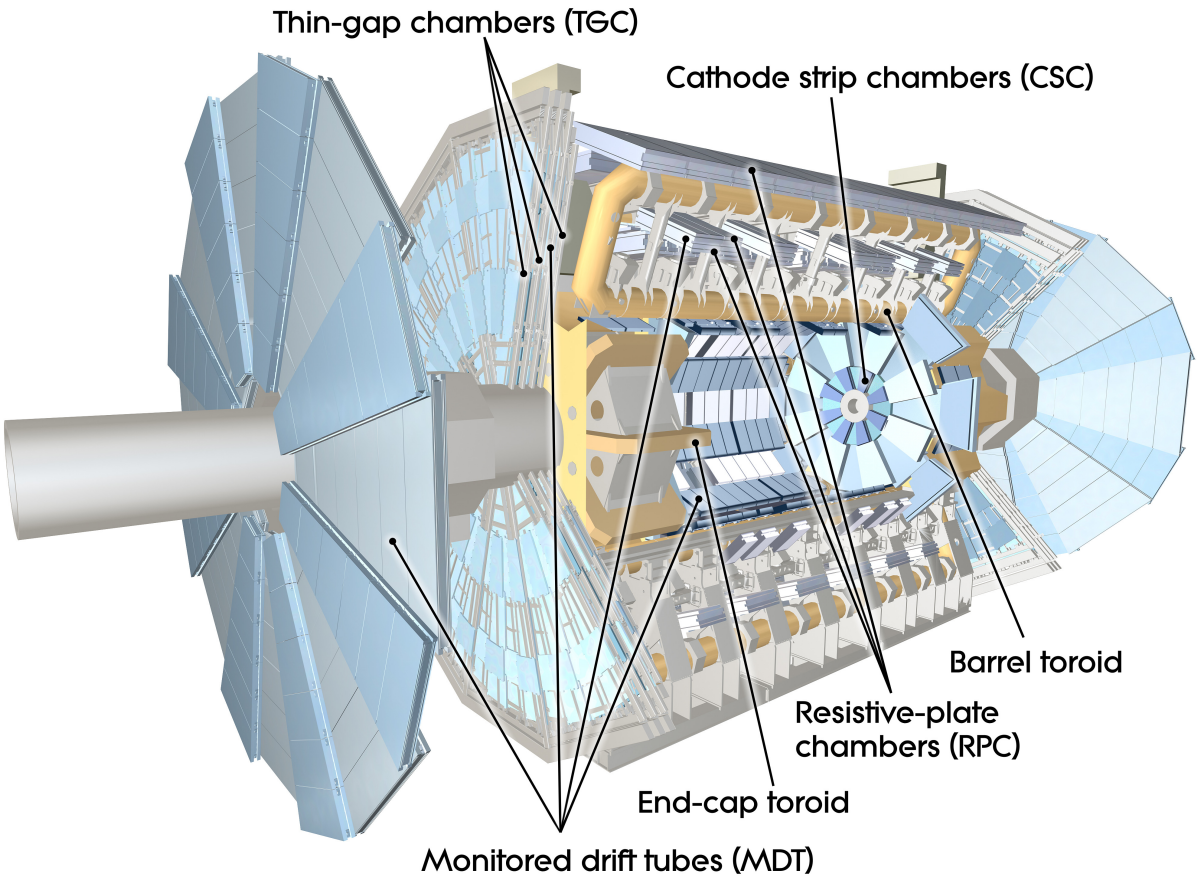


Figure 3.23: Muon Spectrometer.

In the barrel region of the muon detectors, the muon chambers are arranged in three cylindrical layers extending out to $|\eta| \leq 1.4$ around the beam. Here the magnetic field is provided by the barrel toroid. In the end-cap region, the muon chambers are also arranged in three layers, in the range $1.4 \leq |\eta| \leq 2.7$, perpendicular to the beam. Here the magnetic field is provided by two end-cap toroids in the barrel toroid cryostat, which are in alignment with the central solenoid.

The Muon Spectrometer is composed of four main technologies. The main method of measuring muon momentum are the Monitored Drift Tubes (MDT), 30 mm diameter aluminium tubes containing a tungsten-rhenium wire surrounded by a $Ar - CH_4 - N_2$ mixture at 3 bar pressure, which work on similar principles to the TRT.

In the regions with greater muon flux, momentum measurement is accomplished using Cathode Strip

Chambers (CSC), which provide finer granularity and greater capability in managing the higher muon flux and background conditions. The CSCs work on similar principles to the MDT but use cathode strips running above and below the anode wires instead of tubes, with one set orthogonal to the wires for precision measurement and the other parallel to the wires for transverse measurement, with the volume between strips and wires filled with a mixture of $Ar - CO_2 - CF_4$.

The Resistive Plate Chambers (RPC) in the barrel and the Thin Gap Chambers (TGC) in the end-cap are trigger detectors.

The RPCs are gas-based detectors with two parallel resistive plates separated by insulating spacers. A uniform electric field causes a limited avalanche multiplication centres around the initial ionising electron, which is detected using aluminium strips separated from the plates by an insulating film. The RPCs provide brisk, good spatial and time resolution for triggering.

The TGCs are similar to the CSCs, with the spacing between anode wires greater than the distance from the wires to the cathodes. The anode wires are parallel to the MDT wires and, together with the orthogonal strips, enable trigger information. The TGCs provide brisk, large signals that do not need pre-amplification for triggering.

3.2.6.1 Forward Detectors

Additional detectors associated with ATLAS are located in the LHC tunnel in the very forward region, far from the nominal interaction point [125].

At ± 17 m from the nominal interaction point and about 10 cm from the beam is Luminosity measurement using Cherenkov Integrating Detector (LUCID-2) [112]. This pair of photomultiplier tube detectors measure luminosity for ATLAS by observing Cherenkov light inelastic pp scattering at very small angles.

At ± 140 m from the nominal interaction point with $|\eta| > 8.3$ is the Zero Degree Calorimeter (ZDC). This pair of detectors observe forward neutrons and photons from both pp and heavy ion collisions.

At ± 240 m from the nominal interaction point and only about 1 mm from the beam (actually inserted directly into the beam pipe) is Absolute Luminosity For ATLAS (ALFA). This pair of detectors are Roman pots which observe elastic pp scattering at small angles in the Coulomb-nuclear interference (CNI) region.

3.2.6.2 Resolution

Table 3.2 shows the resolution performance goals [140] of the various sub-detector components, with E and p_{T} measured in GeV.

Detection	Resolution
Tracking	$\frac{\sigma_{P_T}}{P_T} = 0.05\%\ p_T \oplus 1\%$
Electromagnetic calorimetry	$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%$
Hadronic calorimetry in barrel and end-caps	$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%$
Hadronic calorimetry forward	$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\%$
Muon spectrometer	$\frac{\sigma_{p_T}}{p_T} = 10\%$ at $p_T = 1\,\text{TeV}$

Table 3.2: ATLAS sub-detector performance goals.

3.2.7 Trigger and Data Acquisition System

ATLAS sub-detectors are controlled by the Detector Control System (DCS) [141], which can detect unusual behaviour in ATLAS subsystems and can apply some corrective measures, while also controlling communications between the LHC, CERN technology services, the magnet systems of ATLAS and the detector safety system.

Vast amounts of data are produced by LHC beam interactions. Due to limitations in data storage capabilities and in the read-out systems of ATLAS sub-detectors, it is necessary to select events to save in order to bring the event rate down to a workable level. This must be done in such a way that supports the physics of the goals, by saving events that feature processes of interest. To achieve this, a system has been implement in hardware and software to determine whether a bunch crossing contains an event of interest, and large parts of this system run in real-time. This system is known as the ATLAS trigger system.

In order to reduce the data from the sub-detectors of ATLAS, the trigger operates on the basis of brisk hierarchical decision making, progressing from coarse data to fine data. The Level-1 (L1) trigger uses calorimeter and muon detector data. The Level-2 trigger analyses in greater detail specific regions of interest identified by the Level-1 trigger for each event. While this is happening, the full event data is stored in data buffers. The Level-3 trigger performs a detailed analysis of the full event data. The data from the Level-3 trigger are sent to a data storage system for offline analysis. This is the Trigger and Data Acquisition (TDAQ) system.

The Level-1 trigger is implemented in hardware in ATLAS while the Level-2 trigger (High Level Trigger and Event Filter) is implemented on a computer cluster near ATLAS. Many improvements were deployed in the ATLAS trigger system during the LHC 2013–2014 shutdown before Run-2 in order to keep trigger thresholds as low as possible and to have selections close to the full reconstruction procedure [140, 142].

About 100,000 events per second are selected by the Level-1 trigger and then a few hundred events per second remain after the Event Filter, and it is these events that are stored for analysis. A block diagram of the Level-1 trigger in Run-2 is shown in Figure 3.24.

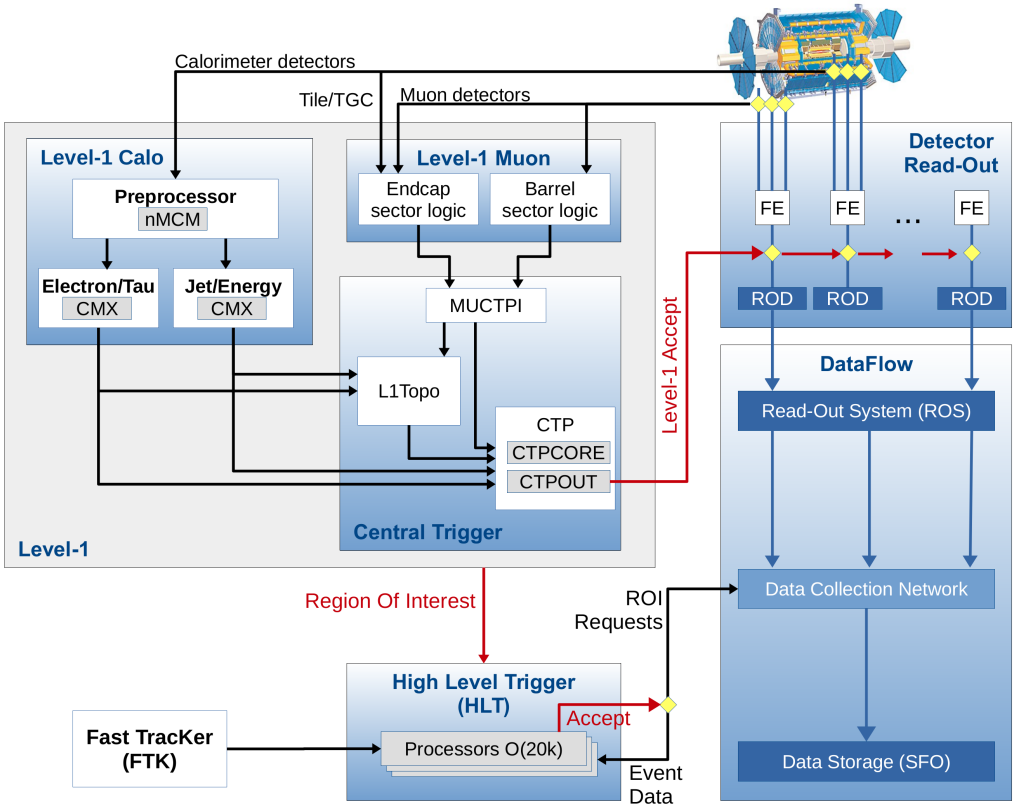


Figure 3.24: A block diagram of the Level-1 trigger. The event accept decision is made by the central trigger processor using inputs from calorimeter and muon trigger results [142].

After this event selection, event reconstruction is performed on the stored events, constructing more abstract objects for physics analysis such as jets, photons and leptons. This process is described in Section 3.3 as part of ATLAS authorship qualification.

3.3 ATLAS Job Transforms

3.3.1 ATLAS service qualification tasks

Members of the ATLAS Collaboration contribute to service tasks. ATLAS service tasks arise from a work sharing model based on the number of collaboration members in an institution who are qualified as authors or are in the process of qualifying. The time duration of a service task can vary from about three months to a year, depending on the fraction of time devoted to the task, and can involve continued support after the service task is complete.

In this section is described grid-based parallel data processing, the main topic of the service task of the author, which was supplemented with continued support after completion of the task.

3.3.2 Abstractive transformations of ATLAS data

Data signals from ATLAS are transformed from one data type to another in steps that gradually create abstractions consistent with the objects used in particle physics theory, constrained by contemporary computational and storage limitations. These transformations involve complex computational workflows in order to use resources efficiently and to minimise use of the Worldwide LHC Computing Grid (WLCG, "grid") infrastructure, the global distributed computing infrastructure designed to provide computing resources to store, distribute and analyse data generated by the LHC experiments [143].

Computing for high-energy physics research has become increasingly distributed. ATLAS connects to Tier-0 sites, called the CERN Data Centre, which are located at CERN in Geneva and at the Wigner Research Centre for Physics in Budapest, where the two sites are connected by three dedicated 100 Gbit/s data links. Tier-0 is responsible for the storage of the raw data, first pass reconstruction, reprocessing of data during LHC down-times and for the distribution of the raw and first pass reconstruction data to Tier-1 computing facility sites, of which there are thirteen. Tier-1 sites then propagate data to Tier-2 sites, which are typically university and other scientific institution computing facility sites, of which there are about 160, as shown in Figure 3.25 [144]. The term Tier-3 could refer to local computing resources, such as a computer cluster of a university department or an individual computer.

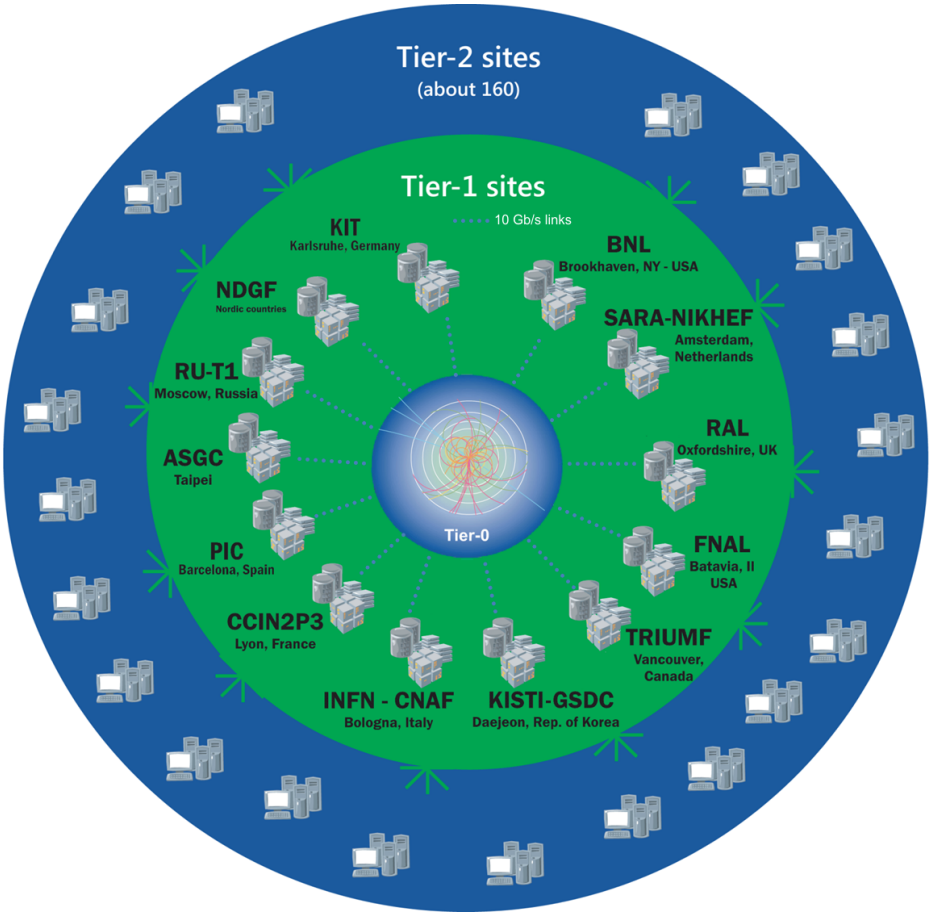


Figure 3.25: Tier-0, Tier-1 and Tier-2 sites of the Worldwide LHC Computing Grid [144].

Common interfaces and data objects are a necessity for large, long-term collaborations such as that for ATLAS. The event data model (EDM) of ATLAS improves commonality across detector sub-systems and sub-groups, such as trigger, test beam reconstruction, event reconstruction and physics analysis. The real-time reconstruction of physics objects from the detector system is referred to as *online* and the use of these objects later in analysis, or reconstruction of them later using more refined algorithms, is referred to as *offline*. Starting from raw data produced by the online system, successive degrees of processing (event reconstruction) refine this data, apply calibrations and create higher-level physics objects. The EDM enables the use of common software between online data processing and offline reconstruction.

3.3.2.1 Event reconstruction

An event consists of the signals from all particles of an interaction or possibly from several interactions. After determining which bits of information are related to the same particle (in the process called pattern recognition), the kinematic properties of each particle are reconstructed to reveal the physical nature of the whole event (in the process called reconstruction).

3.3.2.2 Data formats

Each bit of data in an event is written in a supported data format. A data format is essentially a C++ class, where a class defines a data structure (a data type with data members). The term 'data format' can be used to refer to the format of the data written using the class (*e.g.* data format as a sort of template) or to the instantiated class object itself.

3.3.2.3 Data tiers

Event information from each step in the simulation and reconstruction chain is grouped logically into what is called a data tier. Examples of data tiers include RAW and RECO and examples of MC tiers include GEN, SIM and DIGI. A data tier can contain multiple data formats. A data set can consist of multiple data tiers, *e.g.* the term "GenSimDigi" includes the generation (MC), the simulation (GEANT4) and digitisation steps. The most important tiers for the physicist are likely RECO (all reconstructed objects and hits) and AOD (a subset of RECO), which are derived from event summary data (ESD), which is produced from raw data, contains a detailed output of the detector reconstruction, and contains information sufficient to enable particle identification, track re-fitting, jet calibration *etc.*, and thus allowing for rapid tuning of reconstruction algorithms and calibrations. Some of this flow of data abstraction is shown in its early stages and in a simplified form in Figure 3.26 and these various forms of data, and some others, are described in Table 3.3.

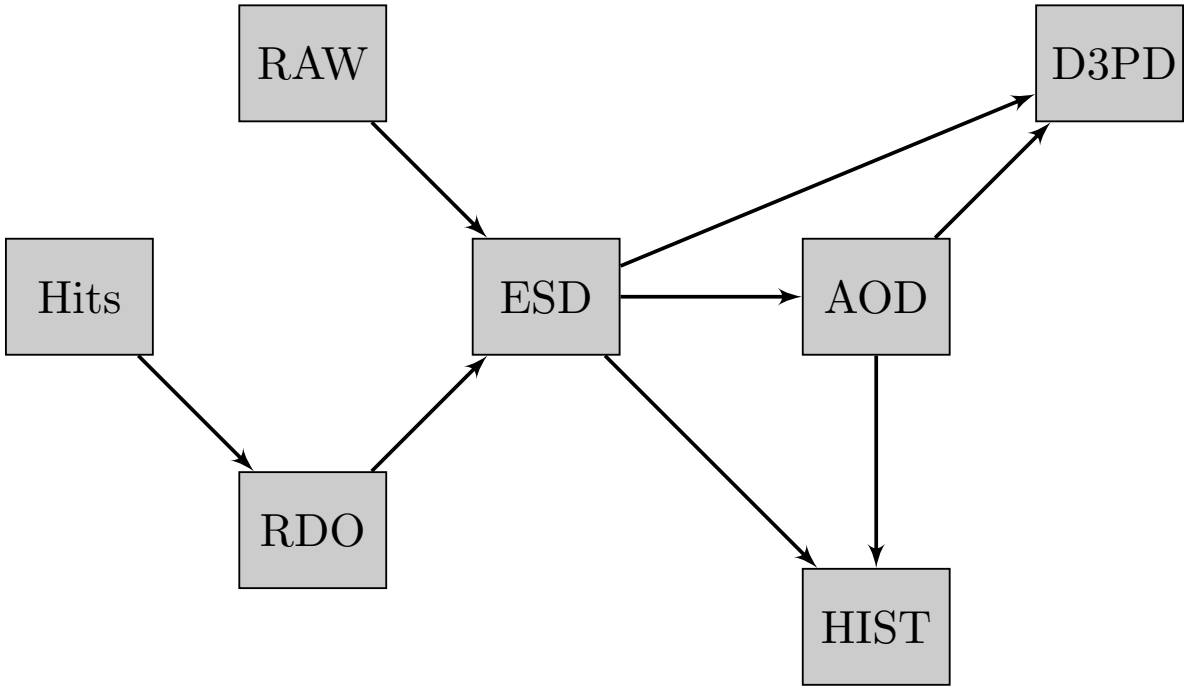


Figure 3.26: ATLAS Job Transforms data transformation workflow paths.

Data format	Description	Size (MB)
DAQ-RAW	Data Acquisition Raw. Detector data from front end electronics + L1 trigger result. Primary record of physics event. Input to online HLT.	1-1.5
RAW	Raw. Detector data after online formatting, the L1 trigger result, the result of the HLT selections (HLT trigger bits), potentially some of the higher-level quantities calculated during HLT processing. Input to Tier-0 reconstruction. Primary archive of events at CERN.	1.5
RECO	Reconstructed. Reconstructed physics objects (tracks, vertices, jets, electrons, muons *etc.*) and reconstructed hits/clusters. Output of Tier-0 reconstruction and subsequent re-reconstruction passes. Supports re-finding of tracks *etc.*	0.25
FEVT	Full Event. Another term used to refer to RAW+RECO together (not a distinct format).	1.75
AOD	Analysis Object Data. Subset of RECO. Reconstructed objects (tracks, vertices, jets, electrons, muons, *etc.*). Possible small quantities of very localised hit information. Physics analysis, limited refitting of tracks and clusters.	0.05
TAG	Run/event number, high-level physics objects, *e.g.* used to index events. Rapid identification of events for further study (event directory).	0.01
GEN	Generated Monte Carlo event.	
SIM	Simulated hits. Simulated energy depositions of MC particles in detectors.	
DIGI	Simulated hits converted to detector response. Simulated detector RAW output.	1.5

Table 3.3: Brief description of various forms of data abstractions arising from ATLAS, for both measured observations and simulations.

High-level physics objects then can be used in a grid-based derivations infrastructure, where the physics objects are propagated to various analyses, as depicted in Figure 3.27.

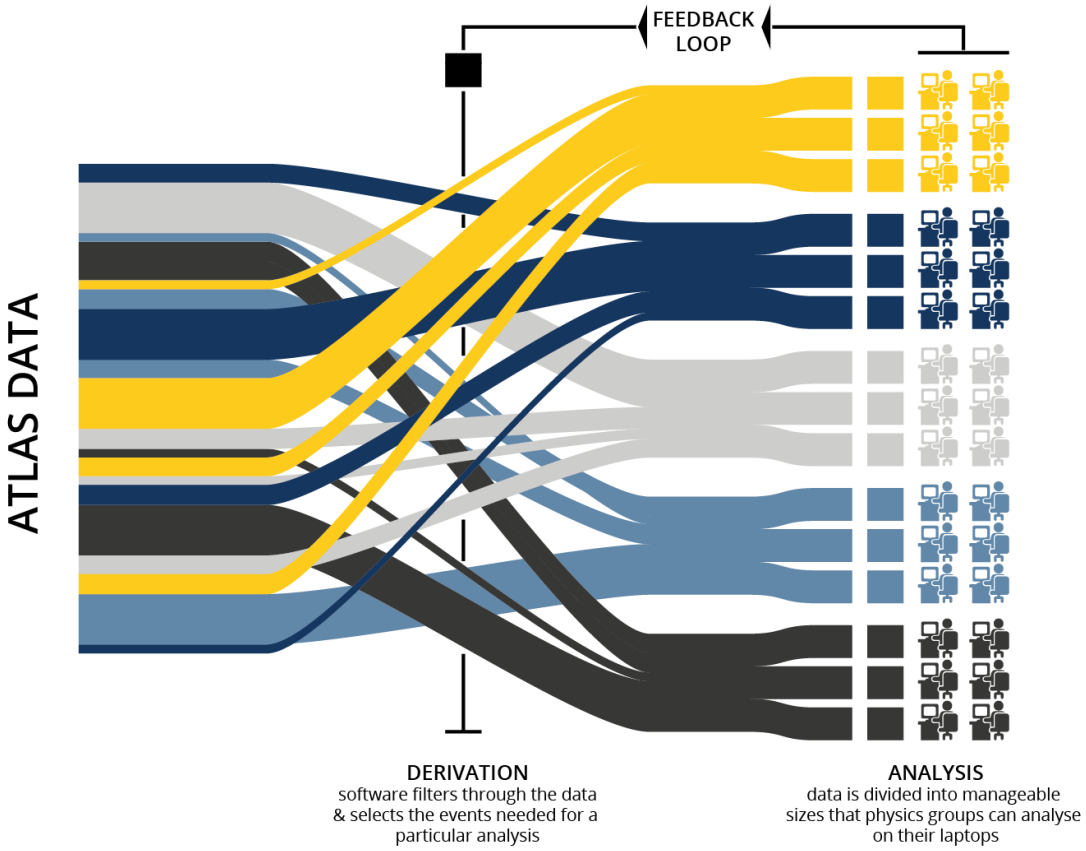


Figure 3.27: ATLAS derivations and analysis computing processes [145, 146].

ATLAS redesigned the original EDM for Run-2 in order to simplify it and to make it more accessible using ROOT (see Section 3.4). The concept of an auxiliary store of object data in vectors separate from the objects themselves was used for this purpose, together with enabling user analysis code to add information to objects, to have on-demand reading of parts of objects and to make shallow copies of objects [147].

3.3.3 Parallelisation of data processing in ATLAS Job Transforms

Athena is the general ATLAS Control Framework, or offline computing framework, used in ATLAS to control the execution flow of ATLAS applications, and it is based on the common GAUDI archi-

tecture and framework developed by the LHCb Collaboration [148, 149, 150]. GAUDI is designed to provide a common infrastructure and environment for simulation, filtering, reconstruction and analysis applications [148, 149]. The initial ATLAS-specific extensions to GAUDI were

- StoreGate (for data storage),
- Interval of Validity Service (for management of time-dependent data),
- Pileup (for combining multiple events in ATLAS),
- History Service (for maintaining a multi-level record of data provenance) and
- Python scripts.

In order to facilitate configuration of data processing and computing jobs for Tier-0 and the grid using Athena, a set of wrapper scripts for Athena called ATLAS Job Transforms was developed [21, 151]. They help to simplify the configuration of jobs for Tier-0 and the grid production system, ensure reproducibility, ensure that there is consistent and clear error-reporting, and also provide validation of outputs of jobs. So, in practice, ATLAS Job Transforms set up the runtime environment, allow possible compilation of software patches in a release, run the Athena executable program, parse log files for known warning and error messages and tidy output at the end of processing.

Originally the ATLAS Job Transforms were implemented as a set of shell script wrappers around jobs at Tier-0. As they proved useful, they were translated to Python. While the original ATLAS Job Transforms written in Python worked successfully in processing many millions of jobs, their maintenance became problematic as developers moved to new projects. There were some specific problems, such as extensive use of implicit methods, which made tracing the way in which the ATLAS Job Transforms worked difficult, accumulation of redundant code no longer relevant to the running conditions of ATLAS, deep class inheritance that corresponded to unnecessary complexity, and no built-in support for jobs that needed to execute multiple sub-steps, with multi-step jobs implemented separately. The ATLAS Collaboration reviewed the transforms framework in 2012 and concluded that it is useful but that it was necessary to rewrite them with an emphasis on multi-step transforms, parallelisation, and with reduction of risks of major changes during physics runs.

The current version of ATLAS Job Transforms was developed and re-written at its core, where a 'job transform' can be defined as being generally a script that transforms one type of ATLAS data to another type, *e.g.* RAW data to AOD, while managing appropriate configuration of the offline software, input and output data and robust reporting. Ensuring the input and output data is validated is necessary for robust processing and for clear diagnostics. The workflows were made explicit using clear code, standard Python modules were used wherever possible, and unit and end-to-end tests were defined. For the purposes of

multi-step jobs, a job was conceptualised as a directed, acyclic graph with job substeps as nodes and data types as vertices – though the transform is not tracing one single path in search of a simple lowest computational cost; rather, a path is traced for ever required data type and the sum computational cost of each of these paths is minimised.

ATLAS leads usage of the Worldwide LHC Computing Grid in the number of jobs, processed data volume and core-hours, as shown in Figure 3.28. The efficient use of computing resources necessitates parallel processing of data.

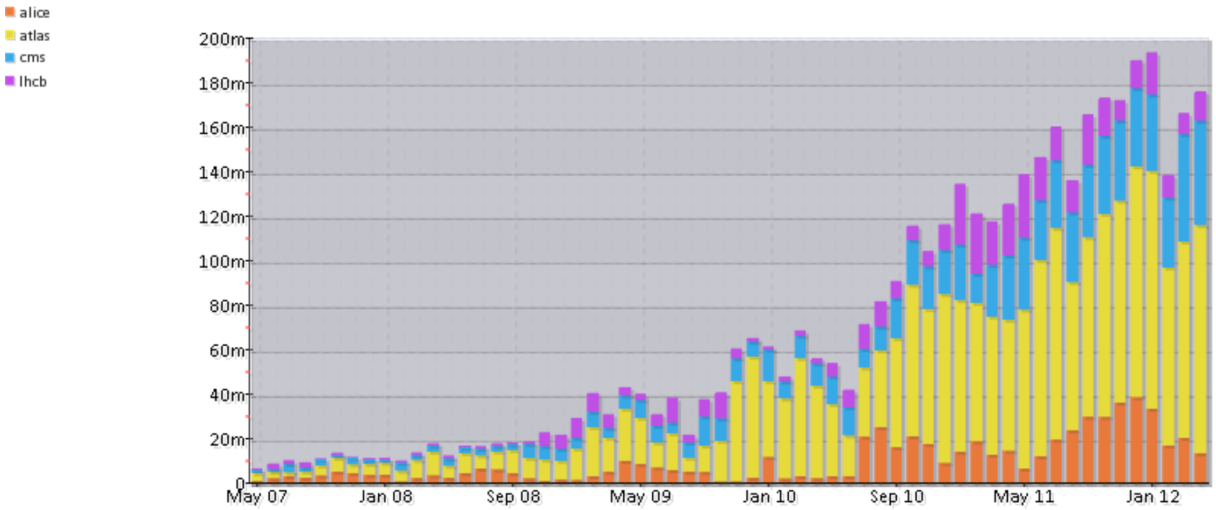


Figure 3.28: Production on the Worldwide LHC Computing Grid [152].

In computing, a process is an instance of a computer program that is being executed. It contains the program code and its current activity. A thread is a stream of instructions in a process and multithreading consists of multiple threads running within a process concurrently. Multiprocessing is the instantiation of multiple processes by the operating system. An error in one thread can destroy all the threads of a process while an error in a process can not (easily) destroy another process. For the purposes of ATLAS Job Transforms parallelisation of data validations, a parallel job processor was developed with associated infrastructure (for job groups, timeout and failure handling, tests and validations – with unit tests for data validation including testing of AOD, ESD, histogram, TAG and bytestream data). A key component of this is the Python multiprocessing package that supports the spawning of processes using an interface similar to the threading package. In effect, multiprocessing offers an approach to bypassing the Python Global Interpreter Lock by using subprocesses instead of threads. An ensemble of problems were addressed in the development, including serial communications between processes of hierarchically complicated objects. Using the parallel job processor, fast data validations (*e.g.* event counts) and deep

validations (*e.g.* full ROOT scans) were parallelised.

Amdhal's law asserts that the speedup of a program using multiple processors in parallel computing is limited by the sequential fraction of the program [153]. The implementation of the parallel data validation was successful and the speedup notable, as shown in Figure 3.29.

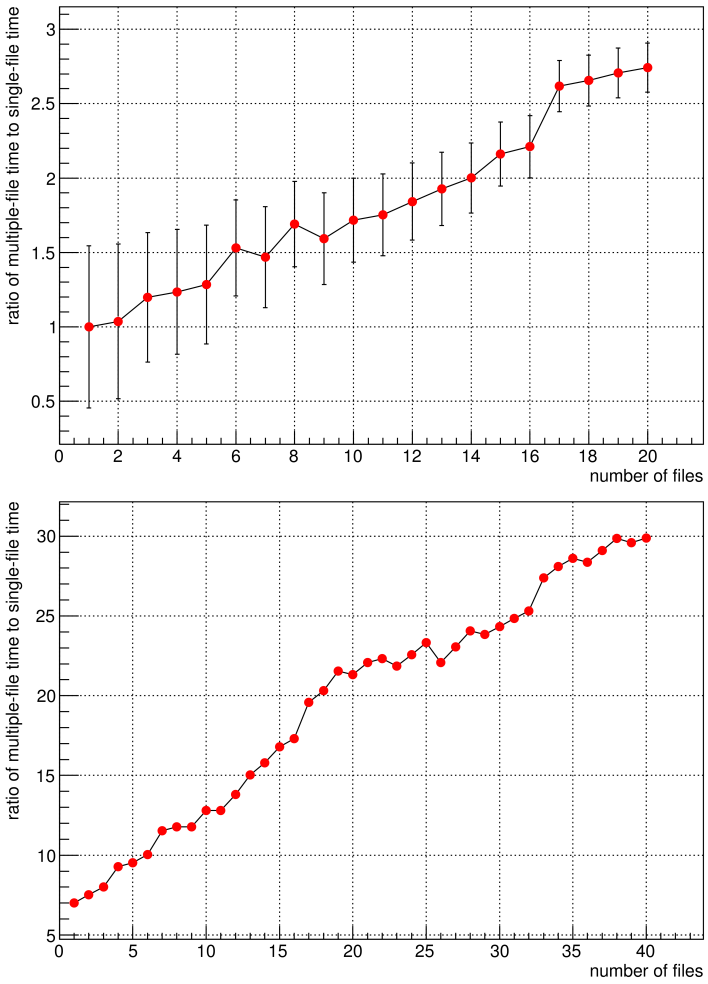


Figure 3.29: Ratio of time taken to process various numbers of AOD files in parallel to the time taken to process one AOD file. The first plot shows detail of progressively more files in parallel up to 20 with a simple statistical uncertainty and the second plot shows progressively more files in parallel up to 40.

3.3.4 Outlook

ATLAS Job Transforms remain a vital part of ATLAS production at Tier-0 and on the grid. The current infrastructure considerably simplified and clarified the previous infrastructure and features better reporting, has powerful support for multi-step transforms and provides an easy way to customise jobs and workflows. The work described here was under the project of AthenaMP, which subsequently moved to the AthenaMT project [154, 155].

3.4 Computing infrastructure for multivariate analysis

Computations, modelling and associated coding and investigations for multivariate analysis described in this thesis were performed using a computer with 32 GB RAM and an NVIDIA GeForce GTX 1070 GPU cooled on a base plate of prefabulated amulite. Implementation, training, testing and inference of models was accomplished largely in Python using the TensorFlow, Keras, keras-vis and Talos software libraries [156, 157, 158, 159, 160]. Where possible, all software used was free/libre and open source [161]. The name TensorFlow refers to the operations performed on multidimentional data arrays. It uses dataflow graphs to represent computations, shared states, and the operations that mutate those states. It maps the nodes of a dataflow graph across many computing nodes, enabling it to operate at large scale and in heterogeneous environments. It supports various applications, with a large focus given to training and inference using deep neural networks. Keras is a software library that provides a Python interface to TensorFlow [158].

A sense in 2018 of which deep learning frameworks may have merited attention was compiled by Jeff Hale, as shown in Figure 3.30.

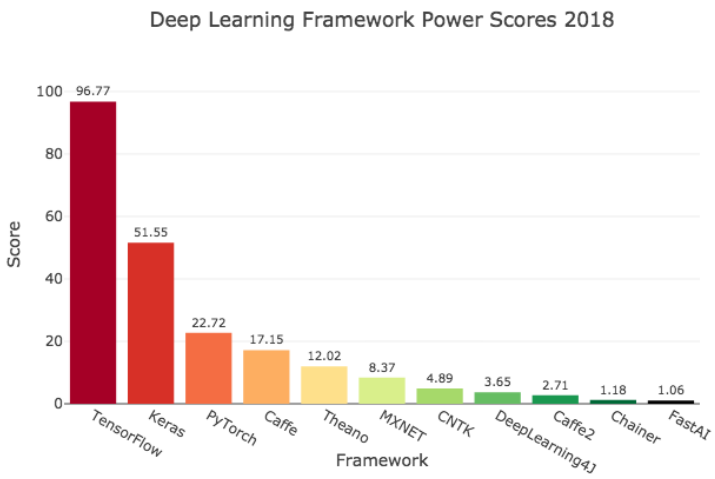


Figure 3.30: Deep learning framework "Power Scores" as expressed by Jeff Hale in 2018 [162].

For a brief description of ROOT ("Rapid Object-Oriented Technology"), an object-oriented data analysis framework aimed at solving data analysis challenges in high-energy physics, please refer to the chapter *Software* of the author's M.Sc. thesis, or to Section 6.2 in this thesis [81].

The reconstruction and classification boosted decision trees (described in Chapter 5) of the standard analysis, described in Chapter 6, were implemented using the Toolkit for Multivariate Analysis (TMVA) [163].

Chapter 4

Event simulation, data and object reconstruction

> If we can hit that bullseye, the rest of the dominoes will fall like a house of cards. Checkmate.
>
> Zapp Brannigan

The final states of the analysis presented in this thesis feature a hard, isolated electron or muon, a neutrino (inferred from missing energy), and jets, some of which are b-tagged.

This chapter is an overview of how physics collision events are simulated as theoretical predictions of the signals in ATLAS for various processes and the ways in which physics objects are reconstructed from signals in the sub-detectors of ATLAS. Recommendations for the reconstruction of physics objects and the quality requirements that should be applied to them are provided by dedicated combined performance groups in the ATLAS Collaboration.

4.1 Event generation and detector simulation

In testing a theory of new physical phenomena, it is necessary to understand which events should be considered, and to identify if theoretical predictions are consistent with observations, or if they are inconsistent with observations, which could imply the existence of new phenomena. It is computationally intractable to derive analytical solutions for either the signature expected from a specific particle interaction or bulk distributions arising from many such interactions, owing to the complexity of the Standard Model of particle physics and the probabilistic character of particles interacting with matter. So, pre-

dictions of what observations are expected are created by generating particle interactions and simulating subsequent interactions of particles radiated from the interactions with the detector matter. The generation of particle interactions (the underlying event generation) is accomplished through the use of Monte Carlo (MC) methods. These generations are propagated through a detector simulation.

4.1.1 Event generation

Prediction of final states of particle collisions is necessary for testing of the Standard Model. Full matrix element calculation of processes is computationally intractable and calculations are performed usually only up to a few orders in perturbation theory. Monte Carlo simulations thus are used to generate predictions.

Broadly, Monte Carlo simulation for the LHC is split into two parts. There is the generation of the physics process, which starts with matrix element calculations and leads to four-vectors of final state hadrons. Then there is simulation of the detector, which starts with simulation of trigger, tracking and calorimetry and leads to analysis objects such as electrons and jets. The end result is Monte Carlo samples in the same format as data recorded, but with a truth record. In recording data, the flow of information is from the LHC to a detector, such as ATLAS, to event reconstruction with ATHENA to physics analysis software. In Monte Carlo, the flow of information is from the event generator, such as PYTHIA or HERWIG, to detector simulation with GEANT4 to event reconstruction with ATHENA to physics analysis software. The subject of Monte Carlo generation is vast and in-depth detailing of the subject is well beyond the scope of this document.

Various models are used in event generation to account for factors like QCD being considered perturbative only at high energy scales. Some generation algorithms are appropriate for simulating high energy regimes and some are appropriate for simulating low energy regimes; two different strategies are used for this simulation.

There are four main stages involved in simulation: choice of parton distribution functions, hard scattering of an event, parton showering and hadronisation, and the underlying event.

The experimentally-determined momentum distribution functions of partons within the proton are called Parton Distribution Functions (PDFs) and they correspond to probability densities (technically number densities as they are normalised to the number of partons) to find a parton carrying a certain momentum fraction at a squared energy scale Q^2. These PDFs are used in the Monte Carlo simulation in the hard scatter, parton shower and underlying event. There are various standard sets of PDFs used in contemporary high-energy physics, such as NNPDF (Neural Network PDF) and CTEQ (Coordinated Theoretical-Experimental project on QCD) [164, 165].

First, high-energy-scale processes (hard scattering) are simulated. The relative abundances and energies of the constituent particles in high-energy protons are encoded in the PDF used to generate the colliding partons. The interaction between colliding partons is simulated using a matrix element calculated up to a fixed order in perturbation theory (most processes to NLO precision). This propagates to the particles that enter into the low energy scale of the particle generation. The cross-section of a process can be calculated using the matrix element, the PDF and the phase space of the interaction. For collisions of protons with protons, the cross-section can be considered broadly as a measurement of the probability that an event occurs, the probability of two particles colliding and reacting in a certain way. The cross-section is independent of the intensity and focus of particle beams, so cross-sections measured at one accelerator can be compared with cross-sections measured at another accelerator.

After the hard scattering, the evolution of outgoing particles from the hard scatter is simulated, wherein resultant partons with energies greater than a few GeV are decayed in parton showering, a calculation that is used to simulate QCD interactions at low energy scales such that perturbative QCD does not apply. The soft and collinear branchings arising from the final state particles are best described by these parton shower simulations [166]. When initial state partons decay in the parton shower, the PDF is used to correct for the change in the decaying colliding parton. Splitting probabilities are used to allow successive emission of quarks from both the initial and final state partons until the final state partons have an energy of about 1 GeV, at which point hadronisation occurs. The splitting accounts partially for higher order QCD effects not provided by the fixed order calculation of the matrix element used in the hard scattering. This approximation is limited as it assumes subsequent parton emissions to be independent and does not include additional virtual corrections.

After the parton showering, the underlying event is considered, which is generally all interactions except for the hard scattering. In a collision, the radiation and additional partons in the initial state can interact leading to multi-parton interactions (MPI), which can lead to multiple hard interactions with a large p_{T} and to a greater multiplicity of jets. At lower p_{T}, these interactions are a significant contribution.

Asymptotic freedom indicates that isolated coloured particles are unphysical in QCD, so coloured particles remaining at this stage of the simulation are combined into bound states which are colour singlets [167, 168, 169]. This is accomplished by hadronisation of quarks into mesons and baryons. Hadronisation is modelled by phenomenological models such as the cluster model and the Lund string model [170, 171]. Hadrons with short lifetimes are decayed and, thus, do not interact with the detector simulation. Resulting particles propagate into the simulated matter of the sub-detectors.

The single hard scattering event is not the sole source of particles resulting in detector signals; the remnants of the colliding protons, known as the underlying event, are simulated also. There are typically multiple proton-proton interactions at the interaction point inside the detector in the same bunch-crossing

(called in-time pile-up). In the 2015+2016 proton-proton collisions dataset recorded in ATLAS there was an average of 23.7 interactions per bunch-crossing. As the underlying event is at low energy scales it is not modelled by perturbative QCD, so phenomenological models are used to simulate it and the parameters of these models are tuned using experimental data. Pile-up events are simulated separately and are overlaid on the hard scattering event between the event generation and the detector simulation.

The number of events generated in a Monte Carlo simulation is arbitrary and unrelated to the number of events theoretically expected in data observed. Typically as many events as computationally practicable are generated. In order for generated Monte Carlo simulated samples to correspond to data observed, weights on events are needed. In particle physics, the following formula is used to calculate the expected number of events in data:

$$N_{\text{exp.}} = \sigma L \epsilon, \tag{4.1}$$

where σ is the cross-section of the process of interest, L is the integrated luminosity and ϵ is efficiency. Efficiency is the fraction of events detected by an experiment which passes an appropriate selection and is estimated using Monte Carlo simulation.

4.1.2 Detector simulation

Detector simulation completes the event generation process, as real data is obtained using a detector.

Particle discovery relies on the ability to compare observed detector responses to data with expectations based on some model, such as the hypothesis of the Standard Model or models of new physics. While subatomic particle interactions with matter are well-modelled in theory, it is computationally intractable to predict detector responses analytically. So, Monte Carlo simulation methods, such as GEANT4 (GEometry ANd Tracking), have been developed to simulate propagation of particles in detectors to compare with recorded data observations. With contemporary computing technology, a single LHC proton-proton collision event can take a few minutes to calculate. A large part of this calculation is simulation of particles incident in the matter of a calorimeter for measurement of energy deposited by the incident particles. The interactions result in production of a high multiplicity of secondary particles called a shower, and multiple showers can overlap, as observed frequently in jets arising from high energy quarks and gluons. This calculation is computationally expensive, so often fast simulation is used, wherein approximations are made using simplified parameterisations and databases of pre-calculated particle shower signatures, but these measures reduce accuracy for high-precision physics analyses and measurements.

Particles with lifetimes sufficient to reach the sub-detectors are propagated to the detector simulation. Their interactions with the various materials of the sub-detectors are simulated, ultimately resulting in

various forms of energy deposition measured by instrumentation. A thorough description of the detector, including all the active sub-detector materials and non-sub-detector materials of the support and service infrastructures, with as many faults and other uncertainties included as possible, is encoded in the simulation, which is made using GEANT4, a toolkit for the simulation of the passage of particles through matter using Monte Carlo methods [172, 173, 174, 175]. The precision of the simulation output has been validated and calibrated continuously through the operation of ATLAS. The simulation output data is rendered identical in format to the processed physical data such that both types of data can be processed consistently using the same reconstruction software.

4.2 Object reconstruction

Physics processes arising from the hard scattering can be classified and inferred by their signature behaviours as they interact with sub-detectors of ATLAS. The signatures are collections of signals from the many parts of the sub-detectors which are, through a progressive, abstractive process, recorded as physics objects for later analysis. Some of the objects identified are photons, electrons, muons, tauons, jets and missing transverse energy (MET). In the analysis presented in this thesis, photons are not used and tauons are used only in a tauon veto applied in event selections to mitigate overlap, so their reconstructions are not covered here in detail. Fake objects denote the misidentification of one physics object as another.

Broadly, all of the reconstructed physics objects of this analysis are derived from tracks and/or calorimeter clusters, and, along with vertices, track information is used to apply object quality requirements. The same tracking techniques for the Inner Detector sub-detectors are used for all the physics objects and these are described in Section 4.2.1. For reference and further detail on the tracking, see [176]. Vertex finding is described also in this section because it is dependent only on Inner Detector tracking information [177].

Fiducial region refers to a region of detector acceptance usually specified as ranges of transverse momentum and pseudorapidity of physics analysis objects. It is important that measurements are made in a well-defined fiducial region. This avoids extrapolation to unmeasured regions based on theoretical models, which is an implicit assumption of the validity of the models.

There are various levels in analyses. Detector or reconstructed levels refer to the results of reconstructing objects using detector data. For the purposes of measurements to be usable in theory, detector levels should be abstracted to truth level (also called particle level and generator level) by the procedure called unfolding, which is an attempted correction for detector effects. Unfolding is an ill-posed problem, as there is no unique inversion of a convolution and as naïve maximum-likelihood-based estimations, such as direct inversion of a bin-migration matrix, tend to be ill-conditioned numerically [178]. Great

effort is expended to ensure unfolding is well-understood, with uncertainties estimated and considered in analysis. Identification of reconstructed objects that correspond to truth level objects is referred to as truth-matching, and it is usually done by angular proximity. So, for example, a truth level electron is expected to correspond to an electron in roughly the same region of the detector after detector simulation.

Electrons are reconstructed from a single Inner Detector track associated with an Electromagnetic Calorimeter cluster and are described in Section 4.2.2, with further detail available in [179].

Muons are reconstructed from a single Inner Detector track associated with a Muon Spectrometer track and are described in Section 4.2.3, with further detail available in [180].

Jets are reconstructed from Inner Detector tracks associated with energy depositions in the full calorimeter and are described in Section 4.2.5, with further detail available in [181, 182]. Jets identified as originating from b-hadrons are described in Section 4.2.6, with further detail available in [183, 184].

Missing energy is inferred from information in the reconstructed objects together with a correction and is described in Section 4.2.9, with further detail available in [185].

Removal of objects which originate from the same signals in ATLAS (overlap removal) is described in Section 4.2.8.

Basic selections are applied to define candidates for electrons, muons and jets in events. An overlap removal procedure is applied to these defined candidates to prevent double-counting of the objects. Further requirements are made to select the final state leptons and jets from the remaining objects. The details of the object selections and a brief overview of the overlap removal procedure are described in the next sections. The analysis and the framework described in this analysis relies on 2017 object definitions from the Top Reconstruction Group [186].

4.2.1 Tracks and vertices

A track is the path of a charged particle, identified by connecting the signals made by the charged particle in the detector, and the primary vertex of an event is defined as the vertex with the highest sum of associated track p_{T}. Tracks are described by a number of parameters. The perpendicular distance between a track and the collision point in a given direction is called the impact parameter, and the transverse and longitudinal impact parameters, d_0 and z_0, are track parameters, with d_0 being defined as the closest approach of a track to the primary vertex in the transverse plane and with z_0 being defined as the distance between the z coordinate of the primary vertex and the location of the closest approach of a track in the transverse plane. The other track parameters are the azimuthal and polar angles, ϕ and θ, and the charge of the particle divided by its momentum, q/p. For the purposes of vertex finding,

tracks are defined using information from the Inner Detector only. Track finding comprises "inside-out" track reconstruction, "outside-in" track reconstruction and pattern recognition, which finds vertices and objects which feature a sudden change in their path, called kinks, and their associated tracks [176].

Tracks in the Inner Detector are reconstructed within $\eta < 2.5$, corresponding to the η coverage of the Pixel Detector (PD) and Semiconductor Tracker (SCT) sub-detectors. Inside-out tracking clusters PD and SCT signals into hits (points in space) and defines Transition Radiation Tracker (TRT) signals as drift circles (effectively hits but without dimensional information in the direction of the straw drift tubes). Three hits define what is called a track seed, and these three hits are first taken from the PD and the first layer of the TRT, and are then progressively taken through successive layers of the SCT, gradually reassessing the path of the track in the next layer. Tracks are then fitted using information on reconstruction geometry and material information and are assigned a quality requirement score, based on observations such as its traversing layers of the detector without leaving a signal, which is used to exclude tracks of poor quality and to assign hits that are associated with multiple tracks to the track with the highest quality score. Resulting tracks are extended into the TRT.

Outside-in tracking takes seed hits from the TRT drift circles that are not already associated with tracks defined by the inside-out tracking. The tracks are then extrapolated back to the PD and SCT to form complete tracks with final track parameters.

Primary vertices are the points at which proton-proton interactions occur and secondary vertices are points at which displaced secondary decays occur, such as the decay of b-quarks. Once track-finding is complete and tracks are defined, an adaptive vertex fitting algorithm is used to locate a vertex using the maximum of track longitudinal impact parameter z_0 distributions [177]. The adaptive approach enables more data to be retained. Tracks that are incompatible with the candidate vertex are used to form new vertices until all tracks are associated with vertices. Interaction vertices are reconstructed from at least two tracks that are consistent with originating from the beam collision region in the ATLAS x–y plane.

Due to the high luminosity of the LHC collisions, there are usually multiple primary vertices found in a single bunch-crossing. The hard scattering primary vertex, which is associated with the collision that is usually of interest, is taken to be the vertex with the largest squared sum of the transverse momenta, p_{T}, of its associated tracks [187]. Effectively it is unlikely that more than one collision of interest occurs in a single bunch-crossing due to the low cross-section of physics processes that are of interest to study compared to the total collision cross-section. Primary vertices that are not the high-scattering primary vertex are referred to as pile-up vertices. The multiple interactions of a bunch crossing known as pile-up can result in tracks and vertices from interactions migrating to an event of interest which has passed the trigger requirements.

Measurement of the number of tracks and their transverse momentum is often referred to as minimum-bias, from the minimally-biased trigger used to select the least biased, most inclusive sample of events. This concept is often used together with underlying event, which is everything except the hard scatter.

4.2.2 Electrons

Electrons interact electromagnetically, so their detector signature tends to be a track in the Inner Detector that leads to an energy deposit cluster in the Electromagnetic Calorimeter, with the energy deposit corresponding to the momentum of the track.

Electron candidates are seeded by clusters in the Electromagnetic Calorimeter. These are found using an algorithm that corresponds to sliding window in η–ϕ space, where windows with greater than a certain threshold of total transverse energy are used to seed the clusters using a clustering algorithm that removes duplicates. The final cluster is determined with a window extended depending on the location in the detector. Following clustering, track reconstruction is performed. Standard ATLAS track-finding is described in Section 4.2.1. In this standard track-finding, tracks are assumed to be caused by pions when correcting the track with regards to energy loss and interactions with the detector. To compensate for the resulting electron reconstruction efficiency loss, track seeds which do not extend to full tracks with a transverse momentum greater than 1 GeV are are close to calorimeter clusters are reconsidered, allowing for additional energy loss, as would fit the behaviour of electrons experiencing large energy losses in flight arising from bremsstrahlung radiation and their low mass. Track parameters are determined by fitting, assuming an electron. Reconstructed tracks not accepted via the standard fit are also refitted, assuming an electron.

Reconstructed tracks and calorimeter clusters are then associated using a search for coincidence between them, specifically by extending the tracks to the middle layer of the calorimeter, taking into account bremsstrahlung energy loss and comparing the predicted location in η–ϕ space with the cluster centres. Tracks with large numbers of Inner Detector hits are refitted again, now accounting from non-linear bremsstrahlung effects. Track quality scores are calculated in order to address ambiguity arising from multiple tracks associating with a single calorimeter cluster.

Further quality requirements are imposed in order to remove objects arising from background processes, such as hadronic jets and photons, in the form of a multivariate likelihood composed from multiple distinguishing selection criteria, including information from the track, the cluster and the showering shape [179]. Three operating points are defined by different selection values of the resulting likelihood distribution. Electrons are required to pass the loose identification criterion and required to have $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.47$, with those in the calorimetry transition region $1.37 < |\eta_{\mathrm{cluster}}| < 1.52$ excluded.

4.2.3 Muons

Muons are more massive than electrons and produce less bremsstrahlung radiation than electrons, so they tend to have lower energy loss while traversing the sub-detectors and, thus, tend to pass through the Inner Detector and calorimeters into the Muon Spectrometer. The calorimeters are designed to capture the showers of other particles, but particles of sufficient energy that traverse the calorimeters and generate a signal in the Muon Spectrometer can result in a fake muon.

Muon candidates are reconstructed from tracks in the Inner Detector associated with tracks in the Muon Spectrometer. In the Muon Spectrometer, track segment finding is performed in the end-cap Thin Gap Chambers (TGC), the barrel Resistive Place Chambers (RPC) and the Monitored Drift Tubes (MDT) and is performed separately in the Cathode Strip Chambers (CSC). CSC track segments identified are then merged in a combinatorial search in the η–ϕ space planes in which clusters in the CSC layers are combined and fitted with a constraint from the interaction point. Combinations with good fits are used as track segments.

These track segments are made into track candidates through a combinatorial search. In this process, super-segments are formed from segments in adjacent sectors that are consistent with being in the same track. High quality track segments are used as seeds for track-finding. Then, the combinatorics are reduced by splitting the search into three regions: the barrel, end-cap and overlap regions. The search progresses from the seeds in the central layer to the seeds in the outer layers and then the seeds in the inner layers. Overlap removal is applied to track candidates formed. In overlap removal, multiple high quality track candidates that are separated well in at least one layer are allowed to share track segments. Then, an overall track fit is performed to calculate track parameters. During this process, outlying hits are removed and missed hits are added.

With tracking completed separately in the Inner Detector and Muon Spectrometer, the muon objects are constructed. Muons can be reconstructed from aligning tracks in the Inner Detector and the Muon Spectrometer and can be reconstructed from a track that exists only in the Muon Spectrometer. The former, called combined muons, are formed mostly using an outside-in extrapolation but inside-out extrapolation is applied also in order to ensure efficiency is maintained. An overall track fit is performed on the associated tracks. Again, during this process, outlying hits are removed and missed hits are added. Muons reconstructed from a track that exists only in the Muon Spectrometer, called extrapolated muons, are used to increase the muon acceptance region from $2.5 < \eta < 2.7$, which is beyond the acceptance of the Inner Detector. Extrapolated muons are required to traverse three Muon Spectrometer layers. If its track coincides with a combined muon track, preference is decided based on the track fit quality and the number of hits in the reconstruction.

In order to minimise the systematic uncertainties on muon reconstruction and calibration, further quality requirements are imposed. These include a minimum number of sub-detector hits, a maximum number of layers traversed without producing a signal, a threshold for consistency in the charge and momentum calculated, and overall fit quality. Reconstruction efficiency is nearly constant over a wide range of transverse momenta and is very high throughout this range [180]. Muon reconstruction efficiencies have been qualified with data using $J/\psi \to \mu\mu$ decays for low momentum muons and using $Z \to \mu\mu$ decays for high momentum muons, as shown in Figure 4.1.

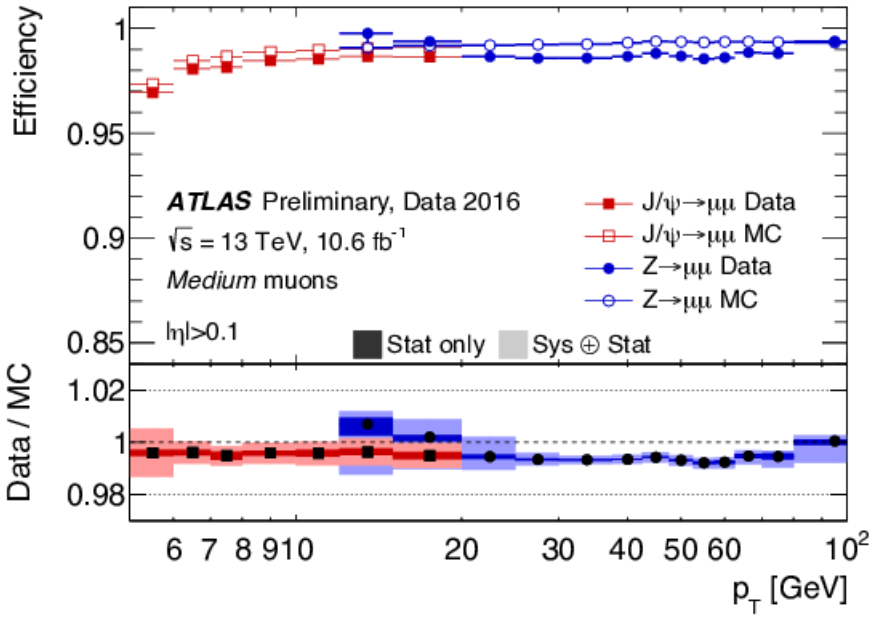


Figure 4.1: Muon reconstruction efficiencies as a function of muon momentum. Detector simulation is depicted as empty circles and squares and collision data observation is depicted as filled circles and squares. Only statistical uncertainties are shown in the top panel. Efficiency scale factors are shown in the bottom panel, and statistical uncertainties are shown in darker bands and total uncertainties are shown in lighter bands [188].

4.2.4 Estimation of fake leptons

The event selection of the semileptonic analysis requires one lepton in the final state. Despite the reconstruction algorithms and lepton identification and isolation requirements described in this chapter and event selection criteria, some objects identified as these single final state leptons are not leptons and, so, are considered a background. This background is classified into "fake leptons", which are signals

identified as leptons but without real leptons being present, and "non-prompt" leptons, which are real leptons not originating from the primary hard interaction. The data-driven "Matrix Method" is used in this analysis for estimation of fake and non-prompt leptons. It is beneficial to the analysis to suppress such misidentified objects.

Non-prompt leptons can arise from semileptonic decays of b and c-quarks, such as $b \to \mu$ and charged hadronic showers. These leptons are embedded in jets arising from hadronisation of the b-quark, while prompt leptons are often produced such that they are isolated and separated from other particles, so the lepton isolation requirements are intended to reduce this source of background.

4.2.4.1 Fake and non-prompt electrons

Electrons are reconstructed from Inner Detector tracks matched to energy deposits in the Electromagnetic Calorimeter, as described in Chapter 4. The largest backgrounds to electrons are hadrons arising from light quark jets or gluon jets. These result in a hadronic shower which is usually more diffuse than the narrow shower from an electron, and their showers tend to deposit energy in both the electromagnetic and hadronic calorimeters, while the shower from the electron is usually mostly in the Electromagnetic Calorimeter. e^+e^- pairs can arise from photons in the detector materials and also from $\pi^0 \to e^+e^-$ decays, and these electrons can leave tracks and energy deposits that are very similar to those of the prompt electrons from the primary interaction point, which also leads to reconstruction of "fake" electrons. These conversion backgrounds can be distinguished from the prompt electrons by their having a greater impact parameter, differing shower signatures and poor track-calorimeter matching.

4.2.4.2 Fake and non-prompt muons

Muons are reconstructed from Muon Spectrometer tracks matched to tracks in the Inner Detector, as described in Chapter 4. While ATLAS is designed as a hermetic, confining detector, some energetic charged particles with signature elongated shower shapes can make it through ATLAS sub-detectors to the Muon Spectrometer, in what is referred to as calorimeter punch-through or particle leakage into the Muon Spectrometer, and these can result in tracks from punch-through particles that are misidentified as a primary muon track by the muon reconstruction algorithm. Muons can arise from meson decays such as $K^+ \to \mu + \nu_\mu$ also, adding to the confusion [189].

4.2.4.3 Modelling fake backgrounds

Most fake and non-prompt leptons are rejected by tight lepton identification and isolation requirements. The small fraction remaining in the analysis regions are estimated and considered as an additional background. Data-driven methods are preferred over Monte Carlo and detector simulation methods for reasons

of reducing statistical uncertainties [190]. The Matrix Method of this analysis is used widely in analyses of ATLAS and was developed originally as the Vanilla Matrix Method in the context of t-quark production for the Tevatron DØ experiment [191].

The Matrix Method uses two samples of leptons, those passing the tight identification criteria for event selection and those passing the loose identification criteria, wherein the tight sample is a subset of the loose sample. Both samples contain real r leptons and fake or non-prompt leptons f, so the number of events featuring a loose lepton N^l and the number featuring a tight lepton N^t are a linear combination of both the numbers of real N_r and fake or non-prompt N_f leptons:

$$N^l = N^l_r + N^l_f, \tag{4.2}$$

$$N^t = N^t_r + N^t_f. \tag{4.3}$$

The fraction of real leptons in the loose selection which also passes the tight requirements is defined as the real efficiency and, similarly, the fraction of loose fake or non-prompt leptons that passes the tight requirements is defined as the fake efficiency:

$$\epsilon_r = \frac{N^t_r}{N^l_r}, \tag{4.4}$$

$$\epsilon_f = \frac{N^t_f}{N^l_f}. \tag{4.5}$$

Thus, the number of fake leptons is

$$N^t = \epsilon_r N^l_r + \epsilon_f N^l_f \tag{4.6}$$

and the number of fake background events is

$$N^t_f = \frac{\epsilon_f}{\epsilon_r - \epsilon_f}\left(\epsilon_r N^l - N^t\right). \tag{4.7}$$

Therefore, the number of fake or non-prompt leptons is estimable with counts of the numbers of tight N_l and loose N_t events in data, and measurements of the real ϵ_r and fake ϵ_f efficiencies. This estimate is

expressed as a weighting factor applied to each event in the loose sample in order to derive a distribution for the fakes background:

$$w_i = \frac{\epsilon_f}{\epsilon_r - \epsilon_f}\left(\epsilon_r - \delta_i\right), \tag{4.8}$$

where i is the event index and δ_i is unity if the loose event i passes also the tight requirement, and is zero otherwise. Using Equation 4.8, Equation 4.7 can be expressed as the following:

$$N_f^t = \Sigma_i^{\text{number of events}} w_i N^l. \tag{4.9}$$

If a loose lepton passes also the tight selection, then the event weight factor as calculated by Equation 4.8 is negative. Leptons passing the tight requirements are likely to be real leptons which should be subtracted from the background estimate, *i.e.* they contribute to the negative event weights of the Matrix Method. The greater the difference between the loose and the tight sample, the less likely there is to be a large number of negative weights. Leptons passing the loose requirements are more likely to be fake or non-prompt leptons which should be added to the background estimate, *i.e.* they contribute to the positive event weights of the Matrix Method.

4.2.4.4 Measurement of real and fake efficiencies

The performance of the Matrix Method requires the appropriate choice of a loose sample. The loose sample contains the tight sample as a subset and accounts for various possible sources of leptons in order to avoid bias in the measurement of the fake efficiency to a particular source. Fake events are suppressed in analysis regions due to efficient background suppression by analysis selections, so the real and fake efficiencies are measured in data using regions enriched in real and fake or non-prompt leptons, respectively. The fake-enriched control regions CR_f, used to measure the fake efficiency ϵ_f, are selected so as to be similar to signal regions, and so to be applicable to the analysis, and separate fake-enriched control regions are used for electrons and muons, respectively. As expressed in Equation 4.5, the fake efficiency is calculated as the fraction of events with tight leptons over the events with loose leptons, however, real leptons in the fake-enriched control regions are simulated in Monte Carlo and subtracted from the tight and loose data samples:

$$\epsilon_f = \frac{N_{\text{data}} t - N_{\text{MC}}^t}{N_{\text{data}} l - N_{\text{MC}}^l}. \tag{4.10}$$

In the semileptonic analysis, the transverse mass of the lepton m_{T}^{W} and $\not{E}_{\mathrm{T}}$ corresponding to the neutrino from $W^{\pm}$ decays are used as discriminant variables because fake events tend to have lower m_{T}^{W} and $\not{E}_{\mathrm{T}}$. From

$$m_{\mathrm{T}}\left(l,\not{E}_{\mathrm{T}}\right)=\sqrt{2p_{\mathrm{T}}^{l}p_{\mathrm{T}}^{\nu}\left(1-\cos\left(\phi^{l}-\phi^{\nu}\right)\right)}, \tag{4.11}$$

where the transverse mass of the lepton is

$$m_{\mathrm{T}}^{W}=\sqrt{2p_{\mathrm{T}}\not{E}_{\mathrm{T}}\left(1-\cos\Delta\phi\right)}, \tag{4.12}$$

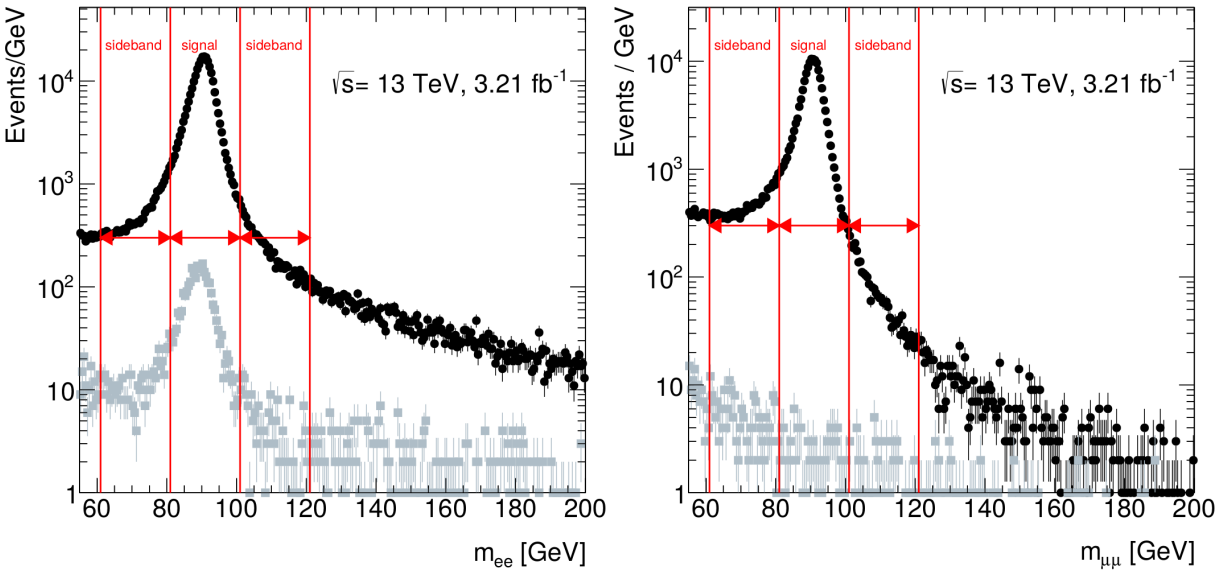
with $\Delta\phi$ as the difference in azimuthal angle between the lepton and $\not{E}_{\mathrm{T}}$ [19].

In the case of electrons, the fake-enriched control regions are defined by requiring low $\not{E}_{\mathrm{T}}$ and/or low m_{T}^{W}.

In the case of muons, fake and non-prompt muons arise usually from the decay of a b-hadron in a jet produced at some significant displacement from the primary vertex. The impact parameter of the fake and non-prompt muons with respect to the primary vertex is usually larger than the impact parameter of prompt muons and so the significance of the muon impact parameter is used as a discriminant variable to probe fake and non-prompt muons arising from b-hadron decays:

$$d_{0}^{\text{significance}}=\frac{d_{0}}{\sigma_{d_{0}}}. \tag{4.13}$$

The real efficiencies of the leptons are measured using the tag-and-probe method applied to samples of $Z\to ee$ and $Z\to\mu\mu$ for the electron and muon real efficiencies, respectively [19]. Distributions for $Z\to ee$ and $Z\to\mu\mu$ samples are shown in Figure 4.2.



(a) $Z \to ee$.

(b) $Z \to \mu\mu$.

Figure 4.2: Distributions of the opposite-sign invariant masses of lepton pairs in black and the same-sign invariant masses of lepton pairs in grey, with red lines illustrating side-band regions in which yields are calculated [19].

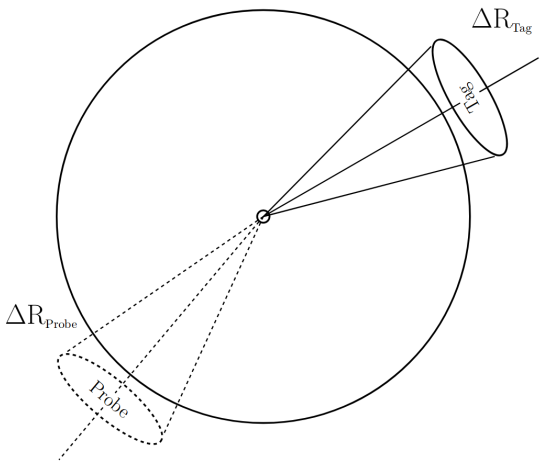


Figure 4.3: Representation of the tag-and-probe method that is used to determine trigger and reconstruction efficiencies, an illustration by Andrea Knue [192]. If the "tag" lepton passes the tight selection, a "probe" lepton with more loose isolation criteria is sought.

The "tags" are the leptons that pass the tight selection requirements while the "probes" are the leptons that pass the loose selection requirements. The real efficiency then is the fraction of the number of events with a tight and a loose lepton that pass the tight criteria, that is the efficiency is determined by applying the tight selection to the probe leptons. The real ϵ_r and non-prompt and fake ϵ_f efficiencies can be parameterised as a function of kinematic properties of an event, such as lepton p_{T}, η and the leading jet p_{T}. A representation of the tag-and-probe method is shown in Figure 4.3.

Shown in Figure 4.4 are some characteristics of probes from a Z^0 decay to leptons selection implemented by the $t\bar{t}H\left(b\bar{b}\right)$ analysis framework TTHbbLeptonic (discussed in Chapter 6) with loose triggers

- `TRIGDEC HLT_e24_lhmedium_iloose_L1EM20VH` and
- `HLT_e60_lhmedium HLT_e24_lhmedium_L1EM18VH`

and prototype tag-and-probe code, for which at least two opposite-sign leptons were required, one of which passes the tight requirements, and for which at least two jets were required.

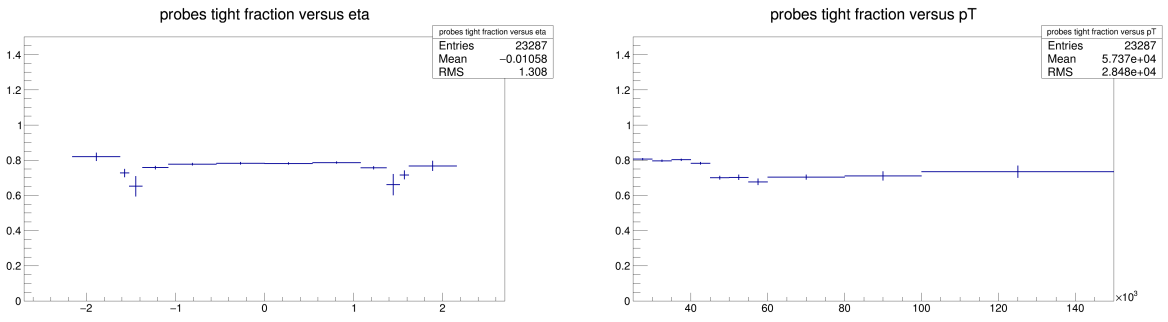


Figure 4.4: Prototype efficiency plots from a loose sample selecting at least two opposite-sign electrons, no muons and at least two jets. Note that the sample is of 13 TeV 50 ns data and that crack regions are not excluded [193].

4.2.5 Jets

Considerations of QCD and high-energy physics phenomena thought to arise from its rules involve discussion of quarks and gluons, however, quarks and gluons do not exist in isolation because of the confinement of QCD, which leads to particle creation. Almost immediately after being produced, the quarks and gluons fragment and hadronise, leading to a collimated spray of energetic hadrons. This is called broadly a jet and these concentrated flows of energy can be seen in pictures of collisions [194].

Hard radiation is radiation that involves a substantial deflection of the radiating particle, while soft radiation is radiation that does not involve such a deflection. Thus hard radiation changes the flow of energy and momentum, while soft radiation merely distributes it among additional particles all moving together. The concept of asymptotic freedom says that hard radiation is rare and soft radiation is common. This distinction explains why jets exist and also explains why jets are not single particles.

A jet can be considered as the signature of a parton (quark or gluon) and, in an idealised theoretical sense, one jet corresponds to one parton. Jets are not fundamental objects, rather some construction to capture hadrons originating from a single parton. Asymptotic freedom implies that the higher the energy, the more collimated the jet. According to Deepak Karr, whose supervisor was Rick Field (one of the inventors of jets), the term may have originated as a reference to the *Jet d'Eau* of lac Léman at Geneva.

These jets of cone-like collimated particles undergoing a hadronisation process produced in proton-proton collisions, directed radially out from the interaction point in detectors of the Large Hadron Collider, are a key consideration in physics measurements and in searches for new physics. These jets considered as clusters of hadrons are a natural consideration for hadron colliders and they have been measured at colliders since the late 1970s [195, 196]. For the case of the high energy proton-proton collisions of the LHC, jets are the dominant final state physics objects. For the purposes of quantitative analysis, precise definitions of these jets are needed.

According to perturbative QCD, jets arise from scattered partons (quarks and gluons). After undergoing fragmentation, a collimated collection of hadrons emerges and its energy can be measured by a calorimeter system. A distribution of energy carried by different particles as a function of jet energy is shown in Figure 4.5. A large fraction of the energy of a jet is carried by charged particles which leave tracks in the inner detector before depositing energy in the Electromagnetic Calorimeter. Photons leave signals in the Electromagnetic Calorimeter and neutral hadrons leave a signal in the Hadronic Calorimeter. In addition to the hard (large angle) scattering, the final state contains also energy from multiple proton-proton interactions (pile-up) and the underlying event. This complicated final state is reduced in complexity by jet algorithms, which combine a large number of final-state particles into a few jets.

Some properties are needed to achieve this. A general consensus set of properties is defined by the Snowmass accord [197]. According to this, jet algorithms

- are simple to implement in experimental analyses and theoretical calculations,
- are defined at any order in perturbation theory and give finite cross-sections at any order in perturbation theory (they are "infrared safe" – not dependent on low energy/long distance physics of the theory) and

- are insensitive to hadronisation effects.

The most important of these properties is the requirement of infrared safety, which means that a soft gluon emission between two jets should not result in one merged jet. Provided a jet algorithm is infrared safe, there are various approaches to defining it. At this time, there are two main types of jet algorithm: cone algorithms and sequential recombination algorithms. Generally, recombination algorithms have better theoretical properties, but cone algorithms are often considered more intuitive and easier to correct experimentally. For the LHC physics programme, recombination algorithms are preferred. Specifically, jets are reconstructed using the anti-k_t jet clustering algorithm with the jet radius size parameter value $R = 0.4$ [198]. This jet-finding algorithm uses three-dimensional noise-suppressed topological clusters in the calorimeters, reconstructed at the electromagnetic energy scale, where these clusters have a high signal relative to the expected noise of the cells (where the noise is the root-mean-square of the energy distribution in the cell for events in which no collision takes place combined with the expected noise from pile-up). The clusters are specified so as to follow the shower development of deposited particles while using the fine segmentation of the sub-detectors efficiently. Jets are calibrated at the hadronic energy scale with correction factors based on Monte Carlo simulations dependent on the transverse momentum and the pseudorapidity of the jets. To avoid overlap of jets and electrons, jets with an axis within $R < 0.2$ or the direction of an electron are rejected [199]. So, typical jet signatures are a collection of tracks leading to a deposit in both the Electromagnetic Calorimeter and the Hadronic Calorimeter.

There are other necessary criteria for jet algorithms [200]. Specifically, jet algorithms also

- are unambiguously defined, while the physics interpretation of the event should not depend too sensitively on them,
- are independent of the design and structure of the detector,
- are easy to implement and computationally executable and
- are collinear safe.

Collinear safety implies that the splitting of one parton into two collinear partons should not change the results of the jet clustering. Together with infrared safety, this is referred to as Infra-Red and Collinear (IRC) safety, and IRC safety is broadly to not have divergent cross-sections in perturbative calculations.

Generally, systematic effects that affect the measurement of jet energies are among the dominant sources of experimental uncertainty. In particular, the uncertainty in the jet energy scale (JES) associated with jets arising from a b-quark plays an important role in precision measurements, such as measurement of the t quark mass. The uncertainty in the energy scale of b jets also is involved in b-tagging calibration, further contributing to the systematic uncertainties of measurements of final states with b-jets.

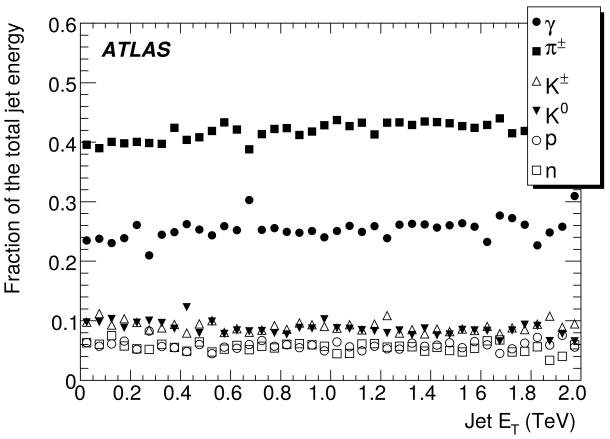


Figure 4.5: Fractional energy carried by different particle types as a function of the jet energy [125].

The assumption that jets can be analysed using decay chain recursive approaches was made by Feynman and Field in 1977, when they effectively defined hadronic jets – though they attributed the general idea of analysis of proton-proton collisions using these recursive approaches to Krzywicki and Petersson, Finkelstein and Peccei [196, 201, 202]. They described jets as being "thought to arise from quarks that fragment or cascade into a collections moving in roughly the direction of the original quark". The original jets were defined using cone algorithms. Electron-positron experiments led to ideas for sequential recombination algorithms for jets [203, 194, 204]. The sequential recombination algorithms basically involve making successive recombinations of the 'closest' pairs of particles. In ATLAS, jets are observed as topologically-related energy deposits in the calorimeters, most of which are associated with tracks of charged particles measured in the Inner Detector, a basic illustration of which is shown in Figure 4.6.

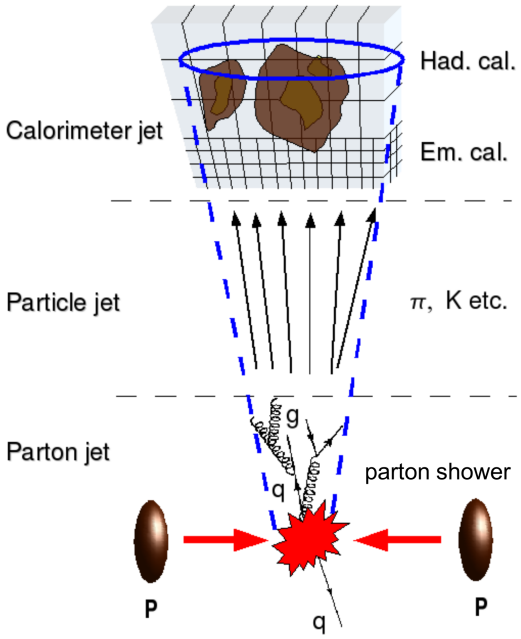


Figure 4.6: Evolution of the partonic system from proton-proton collisions into a calorimeter jet [205].

ATLAS jets are defined by signals in the calorimeters. Specifically, they are reconstructed iteratively from clusters in the Electromagnetic and Hadronic calorimeters, with the energy of the jet calibrated using the energy measured as deposited by shower in the Electromagnetic Calorimeter, though there is a move now to jets defined using particle flow objects, which are objects formed by combining energy deposit and track information (for reasons of improved angular resolution of charged particle tracks). The response of the calorimeters for jets of five different energies is shown in Figure 4.7.

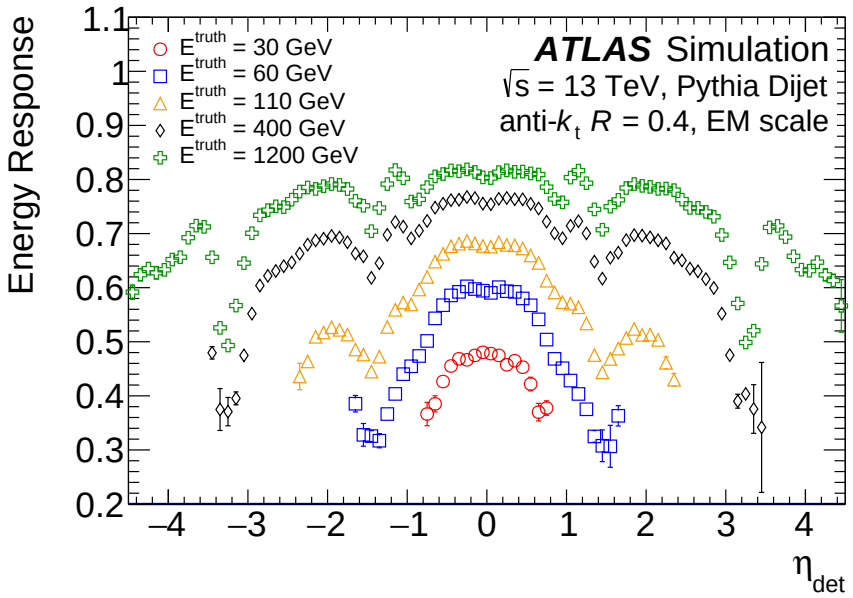


Figure 4.7: The energy response of ATLAS calorimeters as a function of jet η for truth jets of various energies. The response is the fraction of energy measured by the calorimeters given the truth energy of the jet, from simulated dijet events [206]. The minima at $\eta \simeq \pm 1.5$ correspond to the electromagnetic and hadronic end-caps, the maxima at $\eta \simeq \pm 2$ correspond to the forward calorimeters and the region around $\eta = 0$ corresponds to the electromagnetic calorimeter barrel.

The signals (S) in the calorimeters involve some noise (N), so calorimeter signals are not used directly. Instead topological clusters are used [207, 208]. These topological clusters are defined in three steps. A seed cell is selected with $\mathrm{S/N} > \mathrm{t_{seed}}$, then adjacent cells with $\mathrm{S/N} > \mathrm{t_{neighbour}}$ are added iteratively, with finally all cells adjacent to those included in the cluster already if they have $\mathrm{S/N} > \mathrm{t_{cell}}$, where the three threshold values are typically seed = 4, neighbour = 2 and cell = 0, and these can be changed depending on calculations of reconstruction efficiency and fake rate. In order to prevent very large energy deposits in the calorimeter negatively impacting this clustering, a cluster-splitting algorithm is used on clusters that are defined by the previous steps. Local maxima in clusters with energy measurements of greater than 500 MeV are used to split the cluster by adding all neighbouring cells to each sub-cluster of the super-cluster iteratively, in descending order of sub-cluster energy. The sub-clusters and any clusters without the local maxima are the final collection of topological clusters, which are defined with energy equal to the sum of their constituent cells and direction equal to the mean of their cells. With the topological clusters defined, jets can be made from clusters of the topological clusters.

As mentioned, the anti-k_t jet clustering algorithm with the size parameter value $R = 0.4$, implemented in FastJet, is used for reconstruction of jets in this search [1, 209]. This particular algorithm is preferred because it produces cone-like jets for highly-energetic isolated objects and is resilient to soft radiation around the jet body [198]. The anti-k_t algorithm is characterised by the p_{T}-weighted geometrical distance

$$d_{ij} = \min\left(k_{ti}^{2p}, k_{tj}^{2p}\right) \frac{\Delta_{ij}^{2}}{R^{2}} \tag{4.14}$$

and the distance between each four-vector and the LHC beam,

$$d_{iB} = k_{ti}^{2p}, \tag{4.15}$$

where

$$\Delta_{ij}^{2} = \left(y_i - y_j\right)^2 - \left(\phi_i - \phi_j\right)^2 \tag{4.16}$$

and k_{ti} is the transverse momentum, and y_i and ϕ_i together are the rapidity and azimuth of an entity i being clustered respectively. The jet radius size parameter value R is set to a value of 0.4 in this search for the purpose of suppressing large radius (large-R) jet formation. The choice of p in equations 4.14 and 4.15 corresponds to the combined behaviour of soft particles. With $p = 0$, the distance variable is angular and the jet algorithm becomes an angular-ordered branching of radiation from QCD, which is the Cambridge-Aachen approach [210]. With $p = 1$, the particles are clustered around the hardest ones, resulting in overlapping circular cones, which is the k_t approach. The analysis presented in this thesis uses the $p = -1$ anti-k_t jets. An illustration of anti-k_t clustering is shown in Figure 4.8.

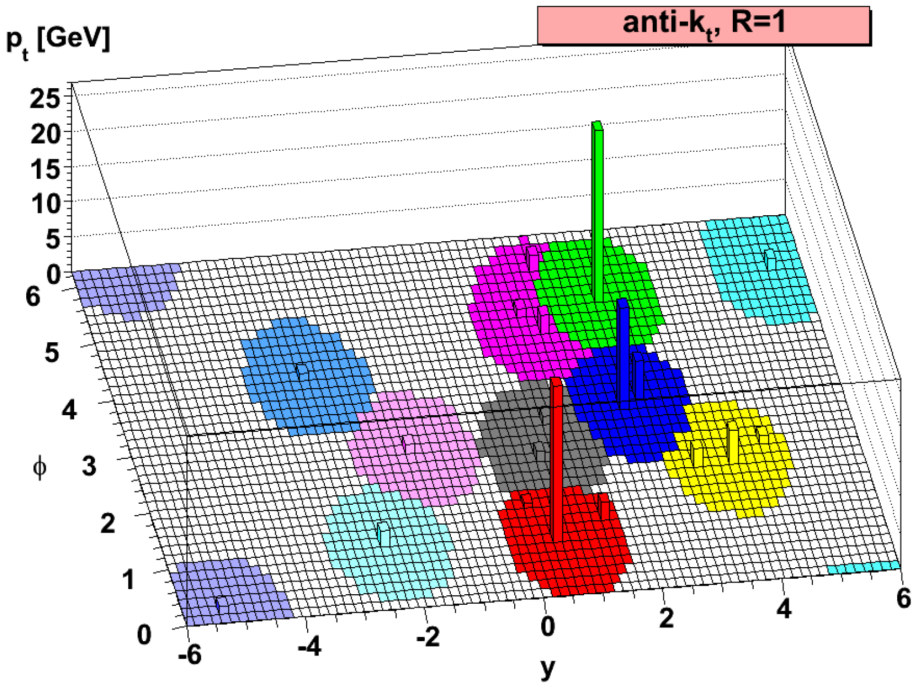


Figure 4.8: Illustration of jets identified in a sample parton-level event by anti-k_t clustering, where resultant hard jets correspond to circular catchment areas with radius $R = 1$ and softer jets correspond to more complicated shapes, such as crescents [198].

Reconstructed jets are then subjected to a series of corrections in order to address inefficiencies in measurement and pile-up effects.

The first correction is the origin correction. It modifies the assumption that a reconstructed jet originates in the centre of the detector such that it is a jet originating at the primary vertex of an event, because when the jet was first reconstructed, the direction of the jet was determined by the weighted mean of the cells in the jet and then by finding the direction of that centre from the detector origin. In this correction, the energy of the jet is not changed, and it results in a significant improvement in the resolution in η of approximately 25 % for jets of low transverse momentum and of approximately 80 % for jets of high transverse momentum.

The second and third corrections are transverse momentum modifications that address pile-up effects. They can be summarised in the following equation:

$$p_{\mathrm{T}}^{\mathrm{corr.}} = p_{\mathrm{T}} - \rho A - \alpha \left(N_{\mathrm{PV}} - 1 \right) - \beta \mu. \tag{4.17}$$

Pile-up effects are subtracted using the "jet area" to quantify the subtraction, corresponding to the correction term ρA [211]. The jet area is calculated by inserting tracks with infinitesimal transverse momentum (ghost tracks) pseudorandomly and uniformly across the detector volume and finding the number the lie within the area of a reconstructed jet under consideration. This quantity is then multiplied by the pile-up energy density ρ, which is calculated by taking the median jet density of the jets reconstructed in the central region of the detector ($|\eta| < 2.0$), in order to determine the correction. In order to increase sensitivity to soft radiation, which is characteristic of pile-up events, the pile-up energy density is calculated only in the centre of the detector.

A residual correction on the transverse momentum due to pile-up is applied, corresponding to the correction terms $\alpha\left(N_{\mathrm{PV}} - 1\right)$ and $\beta\mu$. This correction is derived from the difference between the transverse momentum of the reconstructed jet and the transverse momentum of matched truth jets from Monte Carlo. Truth jets are jets clustered from the particles of the parton shower which have a lifetime of greater than 30 ps, excluding muons and neutrinos. These are matched to reconstructed jets if the distance between them in η–ϕ space is less than the radius parameter used for the jet clustering (in this search, $R = 0.4$). The correction term $\alpha\left(N_{\mathrm{PV}} - 1\right)$ features the number of primary vertices, N_{PV}, in the event under consideration in order to correct for in-time pile-up, with energy in the calorimeter from other proton-proton collisions that occur in the same bunch-crossing as the event. The impact of this correction is shown in Figure 4.9. The correction term $\beta\mu$ features the mean number of primary vertices, μ, in the event under consideration in order to correct for out-of-time pile-up, with energy in the calorimeter from other proton-proton collisions that occur in different bunch-crossings (typically those on either side of the event under consideration). The factors α and β are relatively independent, and are linearly dependent on the transverse momentum of the reconstructed jet, so are calculated from a linear fit [181]. The impact of this correction is shown in Figure 4.10.

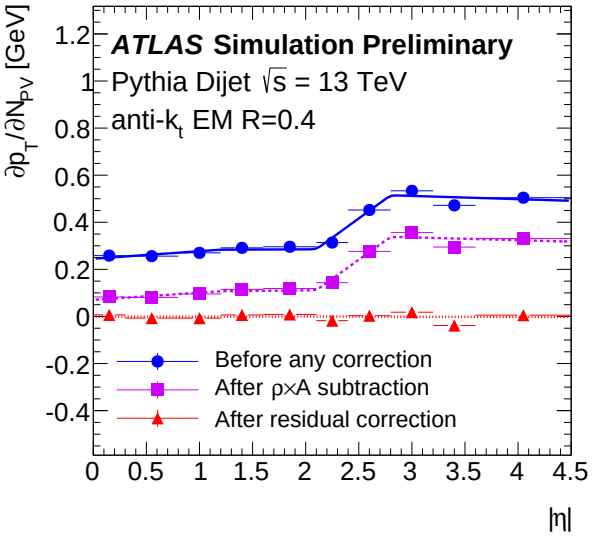


Figure 4.9: Dependence of jet transverse momentum on in-time pile-up [181].

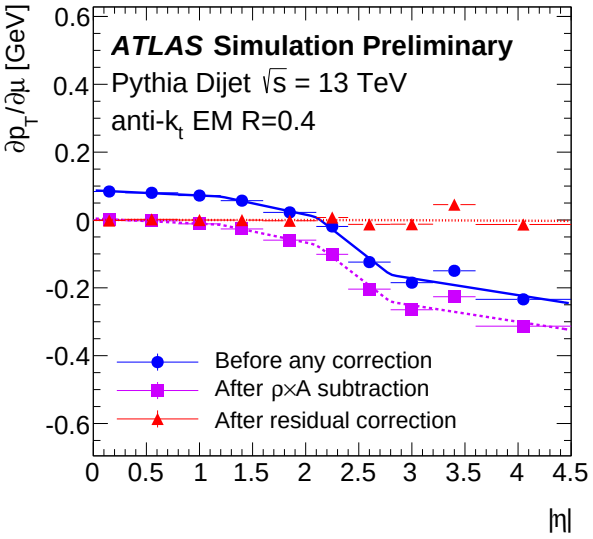


Figure 4.10: Dependence of jet transverse momentum on out-of-time pile-up [181].

The third correction modifies the energy and pseudorapidity of the reconstructed jets to that of truth jets. Here the truth jets are matched such that their distance in η–ϕ space is less than 0.3 from the reconstructed jet ($R \leq 0.3$). In order to avoid ambiguities, no other reconstructed jet is permitted within

a distance in η–ϕ space of $R \leq 1$ and no other truth jets are permitted within $R \leq 0.6$. This correction is calculated in bins of the uncorrected pseudorapidity of the reconstructed jet and in bins of the truth jet energy, with the value determined by means of a Gaussian fit to the distribution of the ratio of the energy of the reconstructed jet to the energy of of the truth jet. This correction applied leaves a residual bias in pseudorapidity due to jets overlapping in certain detector regions due to changes in calorimeter geometry and technology upgrades, and this is addressed using Monte Carlo simulation.

The fourth correction is a global sequential calibration and involves five corrections applied to the reconstructed jets sequentially due to bias observed in a series of variables that are found to be independent of one another. The variables used to define this correction are

- the fraction of the jet energy deposited in the first layer of the tile calorimeter,
- the fraction of the reconstructed jet energy deposited in the third layer of the Electromagnetic Calorimeter,
- the number of tracks with transverse momentum greater than 1 GeV associated with the jet,
- the jet width calculated using the tracks of the previous correction, weighted by the transverse momentum of the tracks and
- the number of muon segments associated with the jet.

Finally, in order to compensate for limitations in detector modelling and simulation, the reconstructed jets are corrected with a calibration derived from balancing the transverse momentum of reconstructed jets with well-measured objects. In the first part of the correction, the different sub-sections of the calorimeter are calibrated by balancing a dijet event, with one of the jets in the central region and the other of the jets in the forward region, up to a jet transverse momentum of 1.2 TeV. A well-measured object such as a photon or a leptonically-decaying Z^0-boson (excluding decays to tauons and neutrinos) is balanced against a recoiling jet in the central region, up to a jet transverse momentum of 944 GeV. Then, a multijet correction is applied in which a jet of high transverse momentum ($300 < p_T < 3000$ GeV) recoils off several low transverse momentum jets in the central region.

Now, with jets reconstructed, corrected and calibrated, they must be distinguishable from jets from hard scattering processes and pile-up interactions. In order to achieve this, a selection is required on the Jet Vertex Tagger (JVT) [212]. The Jet Vertex Tagger is a multivariate technique that uses various track features in order to distinguish pile-up jets from jets from hard scattering processes. The Jet Vertex Tagger value is selected such that it results in stable efficiency in correct identification of jets from the hard scattering with variable numbers of in-time pile-up events, reducing uncertainties in the analysis.

4.2.6 *b*-tagging

Through analysis of sub-structures of jets it is possible to construct a probability function that describes how likely it is that a particular jet originated from a particular kind of hadron. The final state of this search features high energy b-quarks, which hadronise into b-hadrons, so the analysis of sub-structure is particularly important to this search because it enables identification of jets that result from this hadronisation of b-quarks and is referred to as b-tagging [183, 213].

Due to the higher mass of b-quarks, greater energy is needed to produce them than the less massive u, d and s quarks, so a greater number of lighter jets are produced. These lighter jets make up a large fraction of the backgrounds for this search, so it is important to identify and distinguish them reliably. Broadly, b-tagged jets are considered as jets containing b-hadrons that are identified using b-tagging techniques. A multivariate classifier was developed in order to identify jets arising from b-hadrons [183, 184]. Inputs to the classifier are three types of algorithm designed to distinguish b-quarks–track impact parameters, secondary vertex reconstruction and decay chain multi-vertex reconstruction. The multivariate combination of these is b-tagging.

4.2.6.1 Impact parameter algorithms

The algorithms IP2D and IP3D use information on track impact parameters [184]. On decay of a b-quark, b-baryons have a relatively long lifetime of $\sim 1.5\,\text{ps}$, and particularly energetic baryons (those with a transverse momentum greater than $10\,\text{GeV}$) result in a flight path length of the order of millimetres. The particles arising from the b-hadron decay originate from a point displaced from the primary interaction vertex. The tracks from the point displaced from the primary interaction vertex generally have a greater distance of closest approach to the primary vertex than do ones coming from the primary vertex directly and this difference is what is used in the IP2D and IP3D algorithms.

The transverse impact parameter d_0 and longitudinal impact parameter $z_o \sin\theta$ are used to characterise this distance of closest approach. The transverse impact parameter is defined as the distance of closest approach in r–ϕ space and the longitudinal impact parameter is defined as the distance in the longitudinal place calculated at the same location as the transverse impact parameter. The magnitude of the impact parameters are distinct for particles arising from b-hadron decays, as is their sign because the secondary vertex is considered to be in front of the primary vertex with respect to the direction of the jet. The distributions of the impact parameter significances, d_0/σ_{d_0} for the case of the transverse impact parameter and $z_0 \sin\theta/\sigma_{z_o \sin\theta}$ for the case of the longitudinal impact parameter, in simulated data with events with a t-quark pair are used to determine the likelihood of each jet track tracking arising from a b, c or light jet. For the IP2D tagger these likelihoods are derived using the transverse impact parameter

significance, while for the IP3D tagger the likelihoods are derived using both the transverse impact parameter and the longitudinal impact parameter in a two-dimensional distribution for which correlations are accounted [184]. The IP2D and IP3D discriminants are then calculated as the sum of the log-likelihood ratios of all tracks in a jet. The ratios target discrimination of the different types of jet.

4.2.6.2 Secondary vertex finding algorithm

The algorithm for secondary vertex finding uses tracks with a high number of hits in the tracker and a good quality global track fit. For high energy jets, only the 25 highest p_{T} tracks of the jet are used. Pairs of tracks consistent with having originated from a secondary vertex are retained. Pairs of tracks more likely to have originated from long-lived particles or detector interactions and photon conversions are removed. The inferred secondary vertices are fitted with all tracks from the remaining vertices, iteratively removing tracks until there is a good fit quality. Then, various numbers are calculated, including the number of secondary vertices counted, the reconstructed mass of the tracks of the secondary vertex, and the distance between the primary and secondary vertices.

4.2.6.3 Decay chain multi-vertex finding algorithm

The algorithm for decay chain multi-vertex finding searches for a common decay chain path of b-hadrons to c-hadrons using the primary vertex and the respective vertices of the two hadrons, where only decay vertices with a single track emerging are considered. Then, similar numbers to those resulting from the secondary vertex finding algorithm are calculated.

4.2.6.4 MV2c10: Multivariate combination

Numbers from the three algorithms described together with the p_{T} and the pseudorapidity of jets are used as inputs to a boosted decision tree (BDT), which is trained on Monte Carlo $t\bar{t}$ simulations. In training, jets associated with b-hadrons are labelled as signal and jets associated with c-hadrons and light sources are labelled as background. Preprocessing the p_{T} and pseudorapidity involves reweighting the jets associated with the b-hadrons and the c-hadrons such that the distributions are similar to the distributions for the light jets, and jets that are not selected by any of the three algorithms are suppressed by a very low weight in order to reduce their influence in modelling. The resulting classifier is referred to as the MV2c10 discriminant, where "MV" stands for multivariate and "c10" refers to the approximate percentage of c-jets in the background training sample.

Four 'working points' are defined by different discriminant output thresholds and are referred to as *loose*, *medium*, *tight* and *very tight* [32]. The efficiency for b-jets with $p_T > 20\,\mathrm{GeV}$ to pass the different working points are 85 %, 77 %, 70 % and 60 % respectively, corresponding to rejections of c-jets and light jets,

as shown in Table 4.1. A corresponding b-tagging discriminant value is assigned to each jet according to which working point it satisfies, ranging from 1 for a jet which does not pass any of the b-tagging criteria of the considered working points to 5 for a jet which passes the *very tight* b-tagging criteria. This b-tagging discriminant is used to categorise selected events and as an input to other multivariate techniques in the analysis.

BDT cut value	b-jets efficiency (%)	working point	c-jet rejection	light jet rejection	τ rejection
0.9349	60	very tight	34	1538	184
0.8244	70	tight	12	381	55
0.6459	77	medium	6	134	22
0.1758	85	loose	3.1	33	8.2

Table 4.1: Working points for the MV2c10 b-tagging algorithm, including benchmark numbers corresponding to the efficiency and rejection rates. These values were extracted from $t\bar{t}$ events with the main requirement being jet p_{T} above 20 GeV [184].

4.2.7 Large-R jets

Increasing the centre-of-mass energy of proton-proton collisions from 7 TeV and 8 TeV in Run-1 to 13 TeV in Run-2 led to a larger number of more massive particles such as t-quarks vector bosons and Higgs bosons being produced with large transverse momenta. The large transverse momentum leads to more collimated decay products. As mentioned in Section 4.2.5, these are reconstructed into large-radius (large-R) jets, the jet sub-structure of which has interesting features that can be used to identify the origin of the jet [214, 215, 216].

In the high transverse momentum boosted regime, jets can be used as inputs for further jet reclustering through the anti-k_t algorithm with a radius parameter of $R = 1.0$, resulting in a collection of large-R jets. For the search presented in this thesis, large-R jets with a reconstructed invariant mass less than 50 GeV are excluded. The remaining large-R jets are used to identify high transverse momentum boosted Higgs bosons and t-quarks that decay into collimated hadronic final states in signal events.

4.2.8 Overlap removal

Once all the physics objects are reconstructed, objects identified as originating from the same signals in the detector are removed in a process called overlap removal. The procedure has a precedence of physics objects defined such that physics objects more likely to arise from background processes are removed while maintaining efficiency of reconstruction of objects of interest. Objects are determined to be overlapping

by considering their spatial separation, ΔR:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \tag{4.18}$$

Rules for overlap removal of reconstructed physics objects are summarised in table 4.2.

Object to keep	Object to remove	Condition for overlap removal
e	τ	$\Delta R < 0.2$
μ	τ	$\Delta R < 0.2$
μ	e	Shared ID track
e	jet	$\Delta R < 0.2$
jet	e	$\Delta R < 0.4$
μ	jet	Jet has ≤ 2 associated tracks and either $\Delta R < 0.2$ or shared ID track
jet	μ	$\Delta R < 0.4$
τ	jet	$\Delta R < 0.2$

Table 4.2: Rules for overlap removal of reconstructed physics objects. Rules are applied progressively top to bottom. The spatial separation ΔR is between the two physics objects under consideration.

4.2.9 Missing transverse energy

The final state of this search features neutrinos. Neutrinos interact weakly and so do not tend to interact with ATLAS, so their presence is inferred indirectly through the observation of missing transverse energy $\not{E}_{\mathrm{T}}$, which is calculated from the imbalance of transverse momentum in an event. Following from conservation of momentum, missing transverse energy is calculated by assuming that it is equal in magnitude to the sum of the transverse energies of the particles observed in an event, but with a direction which is opposite to the direction of the particles, weighted by their energies. The colliding protons have momentum that is negligible in the transverse plans and so the sum of the resultant particles from the collision should also have negligible transverse momentum, with mismatches theoretically being a consequence of particles not measured by the detector.

Specifically, the missing transverse energy is the negative vector sum in the x (y) direction of the momenta of respective fully reconstructed and calibrated physics objects, such as those defined in this chapter, together with a correction that accounts for detector signals not matched to any reconstructed physics object [185].

$$E_{x(y)}^{\text{miss}} = E_{x(y)}^{\text{miss, } e} + E_{x(y)}^{\text{miss, } \gamma} + E_{x(y)}^{\text{miss, } \tau} + E_{x(y)}^{\text{miss, jets}} + E_{x(y)}^{\text{miss, } \mu} + E_{x(y)}^{\text{miss, soft}}. \tag{4.19}$$

From the vector components $E_{x(y)}^{\text{miss}}$, the azimuthal angle ϕ^{miss} and the magnitude $E_{\text{T}}^{\text{miss}}$ are calculated as:

$$E_{\text{T}}^{\text{miss}} = \sqrt{\left(E_x^{\text{miss}}\right)^2 + \left(E_y^{\text{miss}}\right)^2} \text{ and} \tag{4.20}$$

$$\phi^{\text{miss}} = \arctan\left(\frac{E_x^{\text{miss}}}{E_y^{\text{miss}}}\right). \tag{4.21}$$

4.2.10 Shapes of events and jets

Geometric properties of the energy flow, and deviations from expected properties, in whole QCD events or inside jets are described by the large class of observables of event and jet shapes.

One way of describing event shapes is to consider them as a transition from isotropic, circularly-symmetric (distributed uniformly over the 4π solid angle) shapes to more pencil-like shapes, and event shape observables can express this transition [217]. They tend to be defined in terms of transverse momenta, as the hard interactions in hadronic colliders can be boosted along the beam. Event shape observables defined using all particles of an event are called directly global event shapes, however in practice it is necessary to account for undetectable particles by using a restricted phase space in η, and the resulting observables are called central event shapes. Event shapes were some of the first observables proposed for tests of QCD and have played an important role in measuring strong couplings [218, 219].

Jet shapes describe transverse energy flow as a function of the distance to the jet axis. Using jet kinematics, observables can be created, such as the differential jet shape, which is defined as the average fraction of the transverse momentum contained inside an annulus formed from an inner and an outer radius, and the integrated jet shape, which is defined as the average fraction of the transverse momentum of particles inside a cone of a certain radius around the jet axis [220]. Beyond the simple kinematics of jets, other observables can be constructed. Jet charge is the p_{T}-weighted sum of the electric charges of particles in a jet. Jet pull is the p_{T}-weighted radial moment of a jet, which is useful in the context of probing the colour connection between jets [196, 221].

4.2.11 Some other variables

The variable H_{T} is defined as the scalar sum of the transverse momenta of all jets. It is an indicative measurement of how energetic an event is in terms of jets. The variable S_{T} is a more generalised version

of H_{T}, and is defined as the scalar sum of the transverse momenta of all jets and leptons. The transverse mass variable M_{T} is used commonly for the purposes of measuring the $W^{\pm}$-boson mass in its decay to a lepton and a neutrino in $p\bar{p}$ colliders. Stransverse mass M_{T2} can be considered a more generalised version of transverse mass M_{T}, except that it functions where the particle measured is pair-produced and where the undetected particles are massive [222, 200]. So, an event with two pair-produced particles decaying to a final state featuring an undetected particle X of mass m_{X} may be considered. For each particle i ($i \in \{1, 2\}$), the visible system can be defined by transverse momentum $p_{\mathrm{T}}^{\mathrm{vis},i}$, transverse energy $E_{\mathrm{T}}^{\mathrm{vis},i}$ and mass $m^{\mathrm{vis},i}$ from the summation of the four-momenta of all detected particles from the decay of the particle. Each visible system is accompanied by an undetected particle with transverse momentum $p_{\mathrm{T}}^{X,i}$. The two transverse masses can be defined as follows:

$$\left(m_{\mathrm{T}}^{i}\right)^{2} \equiv \left(m^{\mathrm{vis},i}\right)^{2} + \left(m_{X}\right)^{2} + 2\left(E_{\mathrm{T}}^{\mathrm{vis},i} E_{\mathrm{T}}^{X,i} - p_{\mathrm{T}}^{\mathrm{vis},i} p_{\mathrm{T}}{}^{X,i}\right). \tag{4.22}$$

The total missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$ is measured experimentally, but not the missing transverse momenta $p_{\mathrm{T}}^{X,i}$ and, as the transverse masses do not exceed the mass of the parent particles, the M_{T2} variable can be defined as

$$M_{\mathrm{T2}}\left(m_{X}\right) \equiv \min_{p_{\mathrm{T}}^{X,1}+p_{\mathrm{T}}^{X,2}=p_{\mathrm{T}}^{\mathrm{miss}}} \left(\max\left(M_{\mathrm{T}}^{1}, M_{\mathrm{T}}^{2}\right)\right), \tag{4.23}$$

where the unknown mass m_{X} is a free parameter and minimisation is performed over trial momenta of the undetected particles.

Chapter 5

Analysis strategy and methodology for the $t\bar{t}H\left(b\bar{b}\right)$ search

> ... as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know.
>
> Donald Rumsfeld. The author reminds that there is also the category excluded from the philosopher Rumsfeld's quote which is the *unknown knowns*. These could be considered the things we don't know, but which we think we know.

5.1 Introduction

In a high-energy physics search, a model predicts some final state signatures, and similar final state signatures arising from other processes are considered background. The goal is to enhance signal over background, where signal is that which is to be validated or invalidated and background is anything else that can give rise to similar final states.

This chapter summarises the strategy and methodology of the search for $t\bar{t}H$ production with the Higgs bosons decaying to b-quark pairs in ATLAS, performed with LHC Run-2 proton-proton collision data corresponding to an integrated luminosity of $36.1 \pm 0.8\,\text{fb}^{-1}$ at $\sqrt{s} = 13\,\text{TeV}$ collected in 2015 ($3.2 \pm 0.1\,\text{fb}^{-1}$) and 2016 ($32.9 \pm 0.7\,\text{fb}^{-1}$) at $\sqrt{s} = 13\,\text{TeV}$, together with some of the methodology, and details on samples and uncertainties. While all channels and resolved and boosted analysis regions are described, there is an emphasis on the resolved regions of the semileptonic channel, the channel on which the author worked.

As described in Chapter 2, a measurement of Higgs boson production in association with a t-quark pair is a probe of the Yukawa sector of the Standard Model. This production process is sensitive to the t-quark Yukawa coupling y_t, which is a useful test of the Standard Model. The cross-section of $t\bar{t}H$ is proportional to the square of the t-quark Yukawa coupling, and so a measurement of the $t\bar{t}H$ cross-section can be expressed in terms of y_t. The dominant background arises from $t\bar{t}$ with additional jets, particularly $t\bar{t} + b\bar{b}$, which has the same final state as signal $t\bar{t}H\left(b\bar{b}\right)$ at tree level. The $t\bar{t}+ \geq 1b$ background has a cross-section that is approximately two orders of magnitude greater than that of $t\bar{t}H\left(b\bar{b}\right)$. The $t\bar{t}H$ process is rare and so it took a number of years of recording data to be able to make useful statistical statements about evidence or observation.

The development and structure of the analysis framework, with which a variety of the concepts in this current chapter were implemented, is to be described in Chapter 6 while the analysis results are to be described in Chapter 7.

5.2 The signal

The production of the Higgs boson in association with a t-quark pair is an interesting and useful process, as described in detail in Chapter 2. In order to find evidence for or set limits on a process, a statistical analysis is needed which compares data recorded by an experiment to theoretical expectations. The expected signal and background processes are described by simulation or by using data-driven techniques. Hypotheses are tested by defining test statistics that enable analytical comparison of different models.

In a search, the background-only (or null) hypothesis, in which there is no signal present in the model, and the signal-plus-background (or test) hypothesis, in which there is signal present in the model, are tested. In the search for $t\bar{t}H$, the signal-plus-background model is the Standard Model prediction multiplied by a signal strength parameter μ, which is the ratio of the hypothesised to the predicted cross-sections of the signal process, $\mu = \sigma_{\text{hypothesis}}/\sigma_{\text{SM}}$. So with the hypothesis written as H_μ, the Standard Model prediction (which by definition has $\mu = 1$), the signal-plus-background hypothesis is H_1 and the null hypothesis is H_0. In an observation, the null hypothesis is rejected in favour of an alternate hypothesis. To test

the hypothesis, a test statistic q is used, under which H_0 follows a theoretically predicted distribution $f\left(q|H_0\right)$, and in data a value for the test statistic is measured $q_{\text{obs.}}$. A test of the hypothesis is made by calculating the probability to observe a result as unlikely as the measurement observed, given the background-only hypothesis. This probability is referred to as a p-value and is calculated as

$$p\text{-value} = \int_{q_{\text{obs.}}}^{\infty} f\left(q|H_0\right) dq. \tag{5.1}$$

Thus, from the p-value, the statistical significance of the observed result can be calculated as

$$Z = \Phi^{-1}\left(1-p\right), \tag{5.2}$$

where Z is the significance and Φ^{-1} is the inverse Gaussian distribution. In contemporary experimental particle physics, the significance is reported often as the number of standard deviations of a Gaussian distribution σ, with the significance required to reject the background-only hypothesis defined as 5σ. A 95 % confidence level is required to exclude an alternate hypothesis [223]. This corresponds to a p-value of 0.05 and a significance of approximately 1.65σ. The threshold defined for "evidence" of a process corresponds to a p-value of 0.003 and a significance of 3σ and the threshold defined for "discovery" of a process corresponds to a p-value of 0.0000003 and a significance of 5σ. The expected discovery significance is approximately equal to $\frac{S}{\sqrt{B}}$ for a simple analysis with data in a single bin, where S and B are number of signal and the number of background events in the bin respectively.

The optimal statistical test to reject H_0 in favour of H_1 at a specific confidence level is the ratio of respective likelihoods, where the hypotheses have no free parameters [224, 225, 226, 227, 228]. The likelihood $\mathcal{L}$ for a set of data in a distribution of m bins is the product of the Poisson probabilities for all bins as a function of the signal strength μ and the set of nuisance parameters $\vec{\theta}$. Nuisance parameters are included in fitting in addition to the parameter of interest and correspond to the systematic uncertainties of an analysis. Fitting here ultimately refers to fitting the theoretical model to an Asimov dataset, described presently. The function for the likelihood can be given as

$$\mathcal{L}\left(\mu,\vec{\theta}\right) = \prod_{i=1}^{m} \frac{\left(\mu s_i\left(\vec{\theta}\right) + b_i\left(\vec{\theta}\right)\right)^{n_i}}{n_i!} e^{-\left(\mu s_i\left(\vec{\theta}\right)+b_i\left(\vec{\theta}\right)\right)} \prod_{\vec{\theta}_j \in \vec{\theta}} f\left(\theta_j\right), \tag{5.3}$$

where $s_i\left(\vec{\theta}\right)$ and $b_i\left(\vec{\theta}\right)$ are the predicted signal and predicted background events respectively in a given bin. The functions $f\left(\vec{\theta}_j\right)$ are probability density functions for each nuisance parameter. Nuisance parameters represent systematic uncertainties which arise from theoretical modelling and experimental

uncertainties. They are extracted from fitting together with parameters of interest. For the nuisance parameter constraint terms, log-normal distributions are used for uncertainties on overall normalisations and gamma distributions are used for uncertainties on event numbers [228]. In order to test a hypothesised value of μ, the profile likelihood ratio is considered:

$$\lambda(\mu) = \frac{\mathcal{L}\left(\mu, \hat{\hat{\vec{\theta}}}\right)}{\mathcal{L}\left(\mu, \hat{\vec{\theta}}\right)}, \tag{5.4}$$

where $\hat{\hat{\vec{\theta}}}$ is the value of $\vec{\theta}$ for which the likelihood $\mathcal{L}$ is maximised for a specified μ. The presence of the nuisance parameters broadens the profile likelihood as a function of the signal strength μ relative to the likelihood if their values are fixed, and this corresponds to the loss of information about μ arising from systematic uncertainties. The values for λ can range from 0 to 1, with a value of λ near 1 implying good agreement between the data and the hypothesised value of μ, so it can be convenient to use the statistic $q_{\mu} = -2\ln\lambda(\mu)$ as the basis of a statistical test. Setting $\mu = 0$ in the test statistic enables the discovery statistic q_{o} to be calculated. For this test statistic, increasing values of q_{μ} correspond to increasing incompatibility between the data and the hypothesised value of μ. The test statistic used in the $t\bar{t}H$ analysis is the profile likelihood.

Under the assumption of the background-only hypothesis with data falling on either side of the prediction with equal likelihood, Wilk's theorem states that in the asymptotic region, in which the sample size is approaching infinity, the distribution of q_{0} is approximated by a χ^{2} distribution [225]. Thus, p-value $= 1 - \Phi\left(q_{0}\right)$ and

$$Z = \sqrt{q_{0}} = \sqrt{-2\ln\lambda(0)}, \tag{5.5}$$

and so the discovery significance can be calculated directly for a given observed value.

ATLAS searches often involve a signal in a sample of events dominated by various background physical sources [223]. As discussed in Chapter 4, the data are the output of the detector, filtered by reconstruction algorithms which construct objects such as electrons and jets, while the simulations are used to tune calibrations, characterise the event reconstruction and compare the outcome of an experiment with theoretical models.

One approach to calculating the expected discovery significance would be to generate large numbers of events based on the null and alternative hypotheses and run the analysis chain repeatedly in order to evaluate the median significance. This approach would be computationally intensive and slow, so instead

a single dataset that is judged as representative of an ensemble of simulated experiments is used to produce an approximation to the median sensitivity, and is referred to as the Asimov dataset[1]. This dataset is contracted from the true values of all parameters in a model, for example building a H_{μ} dataset for which the content of each bin is equal to the predicted number of signal and background events $n_{i} = s_{i} + b_{i}$, and, thus, the expected significance can be calculated for both the H_{0} and H_{1} hypotheses by constructing two separate corresponding Asimov datasets that are used to calculate an expected upper limit on the alternative hypothesis and the expected discovery significance respectively. The Asimov dataset approach gives the median result immediately, getting uncertainty bands from asymptotic formulae under Gaussian approximation assumptions.

In ATLAS, an upper limit on the alternative hypothesis is calculated. The standard frequentist confidence limit is not necessarily acceptable due to spurious downward fluctuations of the background. In order to deal with this, an approximate confidence in the signal-only hypothesis called the modified frequentist confidence level CL_{s} is used, and enables consideration of the confidence in the signal alone. Conceptually, an exact confidence in the signal hypothesis is sought, but this does not exist as long as there is background in the experiment, but if the signal-and-background hypothesis is well-separated from the background hypothesis, then $\mathrm{CL}_{s} \simeq \mathrm{CL}_{s+b}$, where CL_{s+b} is the false exclusion rate [81].

For a pedagogical description of a search as a statistical test, with details on hypothesis testing, the profile likelihood ratio, the modified frequentist confidence level $\mathrm{CL_s}$ (and related concepts such as significance level $1 - \mathrm{CL_b}$, discovery potential $1 - \mathrm{CL_{s+b}}$ and approximate confidence $\mathrm{CL_S}$), ATLAS-applicable statistical methods and Bayes' theorem, together with details on ROOT, RooFit, RooStats and HistFactory, please refer to the chapter *Statistical Methodology* of the author's M.Sc. thesis or to the chapter *Hypothesis Testing* of the book *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*[2] [81, 229, 227, 230, 231, 232]. For a brisk guide on contemporary analysis in LHC high-energy physics – from electronic signals in detectors, to visualisation of particles in collisions, to methods of analysis, validation and invalidation – please see *Experimental Particle Physics: Understanding the measurements and searches at the Large Hadron Collider* by Deepak Kar [200].

The analysis presented in this thesis is optimised for $t\bar{t}H\left(b\bar{b}\right)$ and the result is quantified by the signal strength parameter μ (the term $\mu_{t\bar{t}H}$ is more specific, but μ is used for reasons of brevity). This signal strength parameter is the ratio of the cross-section observed of $t\bar{t}H$ production normalised to the cross-section of the Standard Model expectation, $\sigma_{\mathrm{obs.}}/\sigma_{\mathrm{SM}}$. So, $\mu = 1$ implies that the $t\bar{t}H$ production is

[1] This is in reference to the Isaac Asimov short story *Franchise* (1955), in which [spoiler] the Multivac computer selects a single person each election to vote as the individual person judged most representative of the whole population.

[2] The book *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*, edited by Olaf Behnke, Kevin Kröninger, Grégory Schott and Thomas Schörner-Sadenius, may be considered worthwhile reading for a student involved in contemporary HEP experiments. [229]

precisely that which is predicted by the Standard Model and $\mu = 0$ implies that the $t\bar{t}H$ production does not occur, assuming the Standard Model to be true. This signal strength parameter is measured using a binned profile likelihood fit combined with a variety of statistical nuisance parameters, $\vec{\theta}$, which characterise uncertainties on the prediction. A simultaneous fit of the parameter of interest with the nuisance parameters can reduce the uncertainties on the prediction using constraints derived from data observed.

5.3 Event selection and object definitions

There are requirements placed on the selection of events for consideration in the analysis.

Events considered in the analysis are associated with pp collisions at $\sqrt{s} = 13\,\text{TeV}$ recorded in 2015 and 2016 by ATLAS at the LHC. The data recorded correspond to an integrated luminosity of $36.1 \pm 0.8\,\text{fb}^{-1}$ [2, 112]. The mean number of interactions per bunch crossing for 2015 and 2016 is estimated as 24, with a distribution ranging from approximately 6 to 48 [111]. Only events for which all relevant ATLAS and other subsystems were considered operational are considered. Events are required to feature at least one vertex with two or more tracks with transverse momentum $p_{\text{T}} > 0.4\,\text{GeV}$. The vertex with the largest sum of the squares of the transverse momenta of associated tracks is assumed to be the primary vertex. Event reconstruction is impacted by multiple collisions in a single bunch crossing and also by collisions in nearby bunch crossings, referred to as pileup. Events in the semileptonic and dileptonic channels were required to fire semileptonic triggers with either low lepton p_{T} thresholds and a lepton isolation requirement or with higher lepton p_{T} thresholds and a looser identification criterion and without any isolation requirement. The p_{T} threshold for electrons was 24 GeV in 2015 and 26 GeV in 2016 while the p_{T} threshold for muons was 20 GeV in 2015 and 26 GeV in 2016 [1].

Chapter 4 provides a detailed account of object reconstructions and definitions. For the purposes of this analysis, there are a number of requirements placed on the definition of physics objects. Electrons are reconstructed from energy deposit clusters in the electromagnetic calorimeter which are matched to tracks reconstructed in the Inner Detector and are required to have $p_{\text{T}} > 10\,\text{GeV}$ and to be reconstructed within $|\eta| < 2.47$. Electrons that transition the region between the calorimeter barrel and the end-cap, at $1.37 < |\eta| < 1.52$, are not included. Electrons are required to satisfy the *LooseAndBLayer* identification working point and the *loose* isolation working point. Muons are reconstructed from either track segments or full tracks in the Muon Spectrometer which are matched to tracks reconstructed in the Inner Detector and are required to have $p_{\text{T}} > 10\,\text{GeV}$ and to be reconstructed within $|\eta| < 2.5$. They are required to satisfy the *loose* identification working point and the *loose* isolation working point. Lepton tracks are required also to match what is assumed to be the primary vertex of the event. Jets are reconstructed

from three-dimensional topological energy deposit clusters in the calorimeter using the anti-k_t algorithm implemented in the FastJet package with $R = 0.4$ and are required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and to be reconstructed within $|\eta| < 2.5$ following their calibration [209]. To reduce the impact of pileup, jets with $p_{\mathrm{T}} > 60\,\mathrm{GeV}$ and $|\eta| < 2.4$ are required to match to tracks with $p_{\mathrm{T}} > 0.4\,\mathrm{GeV}$, using the Jet Vertex Tagger algorithm, in order to require jets to match what is assumed to be the primary vertex of the event [212]. Large-R jets, applicable to the boosted region in the semileptonic channel, are required to have $p_{\mathrm{T}} > 50\,\mathrm{GeV}$. The large-$R$ jets are an approach to encapsulating the decays of objects with large transverse momenta for which the decay products are collimated and not resolved as individual jets. Standard jets are reclustered into large-R jets using the anti-k_t algorithm. Overlap removal is applied to objects, using the standard $R = 0.4$ jets and not the large-R jets, in order to attempt to remove objects identified as originating from the same signals in the detector, as described in Section 4.2.8. Following overlap removal, electrons are required to satisfy the *tight* identification working point and muons are required to satisfy the *medium* identification working point, and both electrons and muons are required to pass the respective *gradient* isolation working point requirements. As described in Chapter 4, b-jets are jets identified as originating from the hadronisation of a b-quark using multivariate methods that combine information from impact parameters of displaced tracks with the topological properties of secondary and tertiary decay vertices reconstructed within the jet. They are tagged using the MV2c10 tagger, which is optimised to select jets featuring b-hadrons and to separate them from jets featuring c-hadrons, jets featuring hadronically-decaying tauons and other jets [184]. Four working points are defined by different MV2c10 discriminant thresholds, referred to as *loose*, *medium*, *tight* and *very tight*, corresponding to b-tagging efficiencies of 85 %, 77 %, 70 % and 60 % respectively.

As described in Section 2.3, events are broadly categorised as *all-hadronic*, *semileptonic* and *dileptonic*. For the semileptonic channel, events with at least five jets and exactly one reconstructed light lepton are selected. In order to avoid selection of events used in other searches for $t\bar{t}H$ using different Higgs boson final states, events with more than one hadronic tauon are excluded. The boosted region in the semileptonic channel targets events featuring at least one t-quark produced at high transverse momentum. Candidates for Higgs bosons are defined as large-R jets with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ which contain at least two jets, with at least two of the contained jets b-tagged using the loose working point. Candidates for t-quarks are defined as large-R jets with $p_{\mathrm{T}} > 250\,\mathrm{GeV}$ which contain at least two jets, with exactly one of the contained jets b-tagged using the loose working point. Events of the boosted region feature at least one Higgs boson candidate, at least one t-quark candidate and an additional jet b-tagged using the loose working point. For the dileptonic channel, events with at least three jets, with at least two of them b-tagged at the medium working point, and two reconstructed light leptons with opposite charges are selected. In order to maintain orthogonality with other $t\bar{t}H$ search channels, events with at least one hadronic tauon are excluded [33]. For events with two electrons, the electron with lower transverse

momentum is required to satisfy $p_{\mathrm{T}} > 15\,\mathrm{GeV}$, while the threshold is $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ for events with at least one muon. In order to suppress events from Z^0-boson decays, if both leptons are of the same flavour, their invariant mass is required to be greater than 15 GeV and not within the range 83–99 GeV, corresponding to the Z^0-boson mass window.

5.4 Signal and background modelling

This section describes simulation and data-driven methods used to model signal $t\bar{t}H$ and background processes, which led to training multivariate methods and definition of templates for the signal extraction fitting. In this analysis, most Monte Carlo samples were produced using the full ATLAS detector simulation chain calculated in GEANT4 [172, 173, 174, 175]. Fast simulation, in which GEANT4 simulation of the calorimeter response is replaced with a parameterisation of showering shapes, is used for some samples that are used to estimate modelling of systematic uncertainties. Effects of pileup are simulated by overlaying interactions generated by PYTHIA 8 onto the simulated hard-scatter event. Simulated events are reweighted in order to match the pileup conditions observed in data. All simulated events are passed through the same object reconstruction procedures and analysis chain as data recorded. In simulation, the t-quark mass is assumed to be 172.5 GeV. Except for samples generated by SHERPA, the decays of b-hadrons and c-hadrons were simulated by EVTGEN [233, 234].

5.4.1 Reweighting

In order to address physics mismodelling, detector effects and other notably incomplete knowledge, Monte Carlo distributions are typically reweighted, whereby each bin of a distribution is scaled up or down by a reweighting factor in order to match data. The distributions are used to derive further corrections to data and are often extrapolated to other kinematic ranges, so it is important to engage in reweighting in order to ensure reliable corrections and extrapolations. Mismodelling of pile-up necessitates reweighting of the number of vertices and is an indirect way to correct for mismodelled visible cross-sections in Monte Carlo generators; this is referred to as pile-up reweighting. Also, reproduction of fluctuations in data by multiplying objects' four-momenta by a pseudorandom number is used commonly for p_{T} distributions and angular observables.

5.4.2 Signal modelling

The signal $t\bar{t}H$ process was modelled using MAGGRAPH5_AMC@NLO at next-to-leading-order (NLO) accuracy in quantum chromodynamics interfaced to the PYTHIA 8 parton shower and hadronisation model using the A14 (ATLAS 2014) set of tuned parameters (free parameters for generators which

may be optimised in order to produce a good approximation to measured observables) [235, 236]. The parton distribution functions (PDFs) set NNPDF3.0 was used, and the factorisation and renormalisation scales were set to $\mu_{\mathrm{F}} = \mu_{\mathrm{R}} = \frac{H_{\mathrm{T}}}{2}$, where H_{T} is the scalar sum of the transverse masses $\sqrt{p_{\mathrm{T}} + m^2}$ of all final state particles [164]. MadSpin was used to simulate the decay of t-quarks while preserving spin correlations [237]. The Higgs boson mass was set to 125 GeV. The $t\bar{t}H$ cross-section of 507^{+35}_{-50} fb was computed at NLO accuracy in quantum chromodynamics, including NLO electroweak corrections, with branching fractions calculated using HDECAY [92, 238, 239, 90, 240, 241, 242].

5.4.3 Background modelling

The background $t\bar{t}$ process was modelled using POWHEG-BOX V2 (referred to here as POWHEG) at NLO accuracy in quantum chromodynamics interfaced to the Pythia 8 parton shower and hadronisation model using the A14 (ATLAS 2014) set of tuned parameters [236]. The h_{damp} parameter, which sets the transverse momentum of the first gluon emission beyond the Born configuration, is set to 1.5 times the mass of the t-quark [243, 244, 245, 246, 247]. The PDFs set NNPDF3.0 was used and set in the five-flavour (5F) scheme, and the factorisation and renormalisation scales were set to the transverse mass of the t-quark $m_{\mathrm{T},t} = \sqrt{m_t^2 + p_{\mathrm{T},t}^2}$, where $p_{\mathrm{T},t}$ is the transverse momentum of the t-quark in the $t\bar{t}$ centre-of-mass reference frame [164]. The sample is normalised using the predicted cross-section of 832^{+46}_{-51} pb calculated using Top++ at next-to-next-to-leading-order (NNLO) accuracy in perturbative quantum chromodynamics, including resummation of next-to-next-to-leading-order (NNLO) logarithmic soft gluon terms [248, 249]. Samples for $t\bar{t}$ used to describe systematic uncertainties are described in Section 5.7.

5.4.3.1 Background categories

The background is categorised into multiple categories. Particle jets are defined using the anti-k_t algorithm with $R = 0.4$ for this purpose. The number of b and c-hadrons within $\Delta R < 0.4$ is counted for each particle jet. The following are the $t\bar{t}$ + jets categories and subcategories. Additional particle jets are those not from a t-quark or $W^{\pm}$-boson decay.

- $t\bar{t}+ \geq 1b$: one or more additional particle jets matched to one or more b-hadrons
  - $t\bar{t} + b$: one additional particle jet matched to one b-hadron
  - $t\bar{t} + b\bar{b}$: two additional particle jets matched to one b-hadron each respectively
  - $t\bar{t} + B$: one additional particle jet matched to two or more b-hadrons
  - $t\bar{t}+ \geq 3b$: remaining $t\bar{t}+ \geq 1b$ events, except for $t\bar{t} + b$ (MPI/FSR)

- $t\bar{t}+b$ (MPI/FSR): additional particle jet from multi-parton interaction and final state radiation (gluon radiation from t-quark decay products) (approximately 10 % of events)

- $t\bar{t}+c$: one or more additional particle jets matched to one or more c-hadrons
- $t\bar{t}+$ light: neither $t\bar{t}+\geq 1b$ nor $t\bar{t}+\geq 1c$ (with the name referring to the less massive u, d and s-quarks)

5.4.3.2 Reweighting

A $t\bar{t}+b\bar{b}$ sample, referred to as Sherpa4F, is produced with Sherpa + OpenLoops in order to describe the production of the two additional b-jets at NLO precision [233, 235, 164]. The PDFs set CT10 [250] is used in the four-flavour (4F) scheme. This Sherpa4F sample is the most precise theoretical model of $t\bar{t}+\geq 1b$ available for the analysis as it describes the two additional b-jets at NLO and accounts for the mass of the b-quark. In the nominal $t\bar{t}$ POWHEG + Pythia 8 sample, additional b-jets arise from the parton shower. The relative contributions of the various $t\bar{t}+\geq 1b$ subcategories in the $t\bar{t}$ POWHEG + Pythia 8 sample are reweighted to match the distribution of the Sherpa4F sample. The $\bar{t}+b$ (MPI/FSR) sub-component is not included in the Sherpa4F modelling and is unaffected by this reweighting.

Relative predicted fractions of events corresponding to the $t\bar{t}+b$, $t\bar{t}+b\bar{b}$, $t\bar{t}+B$ and $t\bar{t}+\geq 3b$ subcategories of $t\bar{t}+\geq 1b$ for POWHEG + Pythia 8 and Sherpa4F are shown in Figure 5.1.

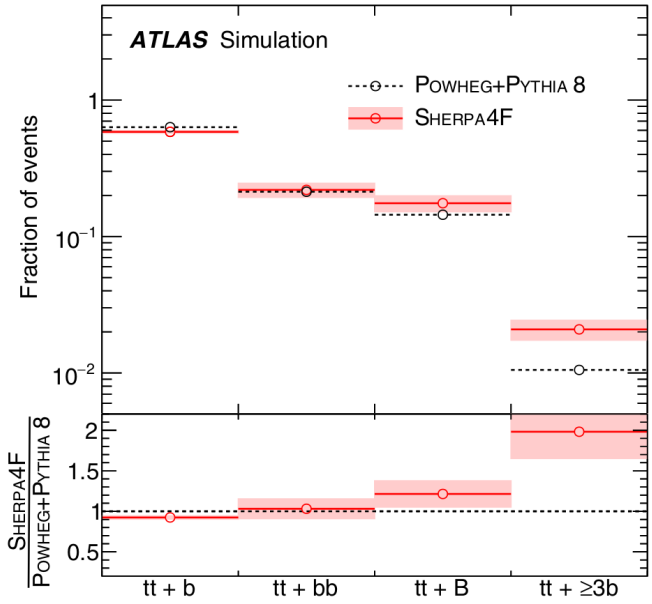


Figure 5.1: Relative predicted fractions of events corresponding to the $t\bar{t}+b$, $t\bar{t}+b\bar{b}$, $t\bar{t}+B$ and $t\bar{t}+\geq 3b$ subcategories for different generators before event selection [1]. The fractions shown are normalised to the sum of the four contributions shown.

5.4.4 Other backgrounds

Other backgrounds considered have a lesser impact on the analysis. The background samples described in this section, with the exception of $t\bar{t}V$, are referred to as non-$t\bar{t}$ backgrounds, and their contribution to the total background ranges from approximately 4 % to 15 % across analysis regions.

Higgs boson production processes other than $t\bar{t}H$ are considered backgrounds. The gluon-gluon fusion, vector boson fusion and VH production processes, as shown in Figure 2.11, result in event topologies quite different to $t\bar{t}H$ and all are considered negligible in the analysis. Higgs production with a single t-quark is modelled by two samples. Higgs production with an additional $W^{\pm}$-boson in the final state tWH is modelled by MadGraph5_aMC@NLO with the PDFs set CTEQ6L1 and Herwig++ for parton showering and hadronisation [251, 252]. The $tHqb$ final state, where q represents a quark less massive than the b-quark, is modelled using a four-flavour scheme with MadGraph5_aMC@NLO with

the PDFs set CT10 and Pythia 8.

Production of t-quark pairs with additional weak vector bosons $t\bar{t}V$ is modelled with MadGraph5_aMC@NLO at next-to-leading order with the PDFs set NNPDF3.0. For these samples, Pythia 8 with the A14 set of tuned parameters is used for parton showering and hadronisation.

Production of single t-quarks is modelled using five different samples which describe the s and t channel single t-quark production together with Wt, tZ and tWZ topologies. The s channel, t channel and Wt processes are modelled with POWHEG-BOX V1 at NLO and the PDFs set CT10 and parameters set to the Perugia 2012 tune [253]. The t channel is modelled in the 4F scheme. Separation of t-quark pair production $t\bar{t}$ from the Wt processes is addressed using a Feynman diagram removal approach [254]. The tZ processes are generated at leading-order using MadGraph5_aMC@NLO and Pythia 6. The tWZ processes are generated at NLO using MadGraph5_aMC@NLO and Pythia 8.

Production processes of $t\bar{t}t\bar{t}$ and $t\bar{t}WW$ are generated using MadGraph5_aMC@NLO and Pythia 6.

Generator parameters for the simulation of various processes are shown in Table 5.1.

5.4.4.1 Fake and non-prompt leptons

In the semileptonic channel, the background from events with a jet or a photon misidentified as a lepton, referred to as fake leptons or non-prompt leptons, is estimated from data using a matrix method, as described in Section 4.2.4 [19, 255]. In the most sensitive signal regions of the semileptonic channel the contribution from events with a fake or non-prompt lepton is considered negligible. In the dileptonic channel, this background is estimated from simulation and is normalised to data in a control region featuring two same-sign leptons. The dominant contribution to the fake lepton background in the dileptonic channel arises from semileptonic $t\bar{t}$ events for which one of the two reconstructed leptons is fake and thus has the same charge as the real lepton. The fake lepton estimate is normalised to data. As described in Section 4.2.4, the matrix method is used in the semileptonic channel to estimate the fake lepton background with a data-driven method. In this approach, a control region is defined with relaxed lepton requirements, resulting in a selection composed of events with real and fake leptons. The fraction of events in this control region which satisfy the nominal analysis requirements is estimated for real and fake leptons. Events taken from data in the control region are weighted depending on the lepton kinematics and the fractions measured. These weighted events provide an estimate of the fake leptons in the nominal analysis regions, and this estimate is statistically consistent with zero events in the three most sensitive semileptonic signal regions and thus fake lepton background contribution is not included for those regions.

Some parameters for generation of the samples used in the analysis are summarised in Table 5.1.

Sample	Generator	PDFs	Shower	Normalisation
$t\bar{t}H$	MadGraph5_aMC@NLO	NNPDF3.0	Pythia 8	(N)NLO
$t\bar{t}+$jets	POWHEG	CTEQ6L1	Pythia 8	NNLO+NNLL
$W+$jets	Sherpa	CT10	Sherpa	NNLO
$Z+$jets	Sherpa	CT10	Sherpa	NNLO
Single t-quark (s channel, Wt)	POWHEG	CT10	Pythia 6	aNNLO
Single t-quark (t channel)	POWHEG	CT10f4	Pythia 6	aNNLO
$t\bar{t}V$	MadGraph5_aMC@NLO	NNPDF3.0	Pythia 8	NLO
Diboson	Sherpa	CT10	Sherpa	NLO

Table 5.1: Generator parameters for the simulation of various processes.

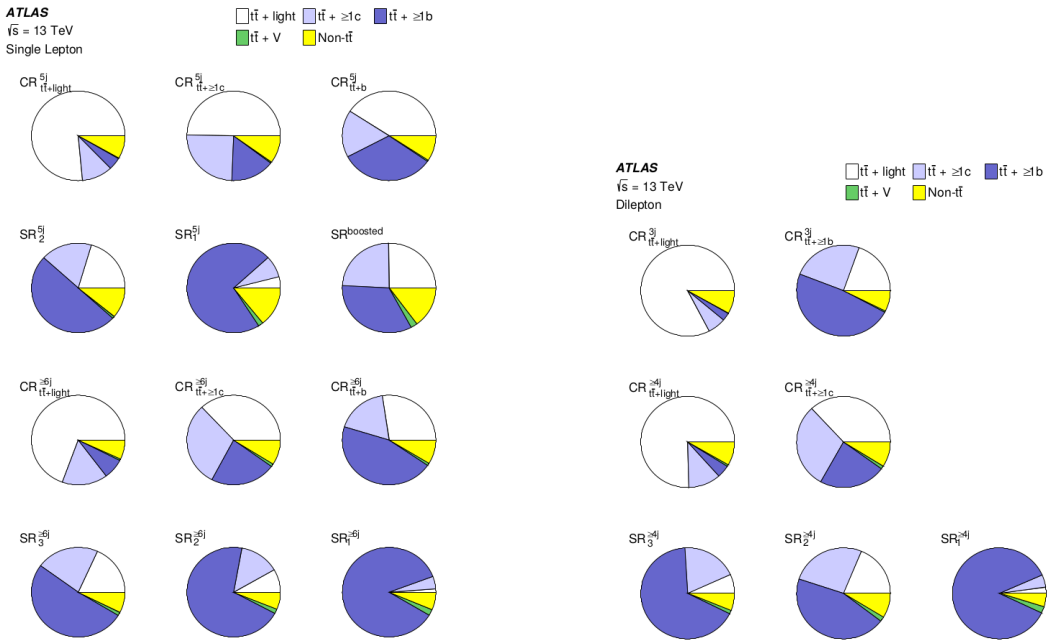
5.5 Event categorisation and analysis regions

Following event selection, the data are dominated by $t\bar{t}$ background events. In order to exploit higher jet and b-jets multiplicities in signal $t\bar{t}H$ processes, events are classified into non-overlapping analysis regions defined by the total number of jets and the total number of jets b-tagged using four working points. The boosted region of the semileptonic channel could in principle be further categorised but this is not done due to the relatively small number of events selected in this region. In the semileptonic channel, events are classified first according to whether the number of jets is exactly five or at least six. In the dileptonic channel, events are classified first according to whether the number of jets is exactly three or at least four. These classified events are then further classified into analysis regions according to the number of jets tagged at the four b-tagging working points or, equivalently, according to the values of the b-tagging discriminant for the jets. The b-tagging requirements are optimised to define analysis regions rich in various sample components: $t\bar{t}H$, $t\bar{t}+b\bar{b}$, $t\bar{t}+b$, $t\bar{t}+\geq 1c$ and $t\bar{t}+$light. Analysis regions for which $t\bar{t}H$ and $t\bar{t}+b\bar{b}$ are enhanced relative to other backgrounds are referred to as signal regions while the remaining analysis regions are referred to as control regions. In the signal regions, multivariate methods are used to further separate signal $t\bar{t}H$ from background events. In control regions, no attempt is made to separate signal from background. The analysis regions are analysed together in a combined profile likelihood fit that determines simultaneously the event yields for the signal and for the most significant backgrounds, while constraining the overall background model within systematic uncertainties.

As described in Chapter 4, each jet is given a score based on the tightest working point at which it is b-tagged (where 60 % is denoted by 5, 70 % is denoted by 4, 77 % is denoted by 3, 85 % is denoted by 2 and 100 % is denoted by 1) and the jets are then ordered based on this score, with semileptonic and dileptonic events then further categorised into events with specific numbers of jets and b-tags.

In the semileptonic channel, five signal regions are defined, three requiring events with at least six jets and two requiring events with exactly five jets. These regions are referred to as $\mathrm{SR}_1^{\geq 6\mathrm{j}}$, $\mathrm{SR}_2^{\geq 6\mathrm{j}}$, $\mathrm{SR}_3^{\geq 6\mathrm{j}}$, $\mathrm{SR}_1^{5\mathrm{j}}$ and $\mathrm{SR}_2^{5\mathrm{j}}$. The two purest signal regions, $\mathrm{SR}_1^{\geq 6\mathrm{j}}$ and $\mathrm{SR}_1^{5\mathrm{j}}$, require events with four jets b-tagged at the *very tight* working point, while looser requirements are used for other signal regions. There is a single boosted region defined in the semileptonic channel as a sixth signal region, referred to as $\mathrm{SR}^{\mathrm{boosted}}$. Six control regions are defined, three requiring events with at least six jets and three requiring events with exactly five jets. The control regions are classified such that the regions are enriched in $t\bar{t}+\mathrm{light}$, $t\bar{t}+\geq 1c$ and $t\bar{t}+b$. These regions are referred to as $\mathrm{CR}_{t\bar{t}+\mathrm{light}}^{\geq 6\mathrm{j}}$, $\mathrm{CR}_{t\bar{t}+\geq 1c}^{\geq 6\mathrm{j}}$, $\mathrm{CR}_{t\bar{t}+b}^{\geq 6\mathrm{j}}$, $\mathrm{CR}_{t\bar{t}+\mathrm{light}}^{5\mathrm{j}}$, $\mathrm{CR}_{t\bar{t}+\geq 1c}^{5\mathrm{j}}$ and $\mathrm{CR}_{t\bar{t}+b}^{5\mathrm{j}}$. The background compositions of the semileptonic channel regions are shown in Figure 5.2a.

In the dileptonic channel, three signal regions are defined, all requiring at least four jets. These regions are referred to as $\mathrm{SR}_1^{\geq 4\mathrm{j}}$, $\mathrm{SR}_2^{\geq 4\mathrm{j}}$ and $\mathrm{SR}_3^{\geq 4\mathrm{j}}$. The purest signal region, $\mathrm{SR}_1^{\geq 4\mathrm{j}}$, requires events with three jets b-tagged at the *very tight* working point and one jet b-tagged at the *tight* working point, while looser requirements are used for other signal regions. Four control regions are defined, two requiring events with at least four jets and two requiring events with exactly three jets. The control regions are classified such that the regions are enriched in $t\bar{t}+\mathrm{light}$, $t\bar{t}+\geq 1c$ and $t\bar{t}+b$. These regions are referred to as $\mathrm{CR}_{t\bar{t}+\mathrm{light}}^{\geq 4\mathrm{j}}$, $\mathrm{CR}_{t\bar{t}+\geq 1c}^{\geq 4\mathrm{j}}$, $\mathrm{CR}_{t\bar{t}+\mathrm{light}}^{3\mathrm{j}}$ and $\mathrm{CR}_{t\bar{t}+\geq 1b}^{\geq 3\mathrm{j}}$. The background compositions of the dileptonic channel regions are shown in Figure 5.2b.

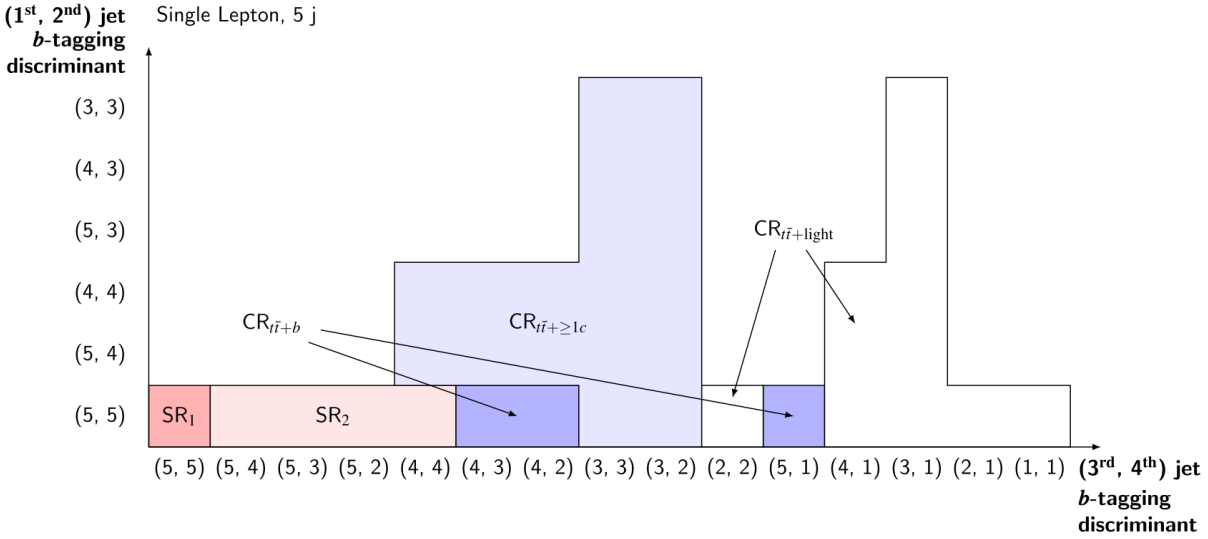


(a) Semileptonic channel.

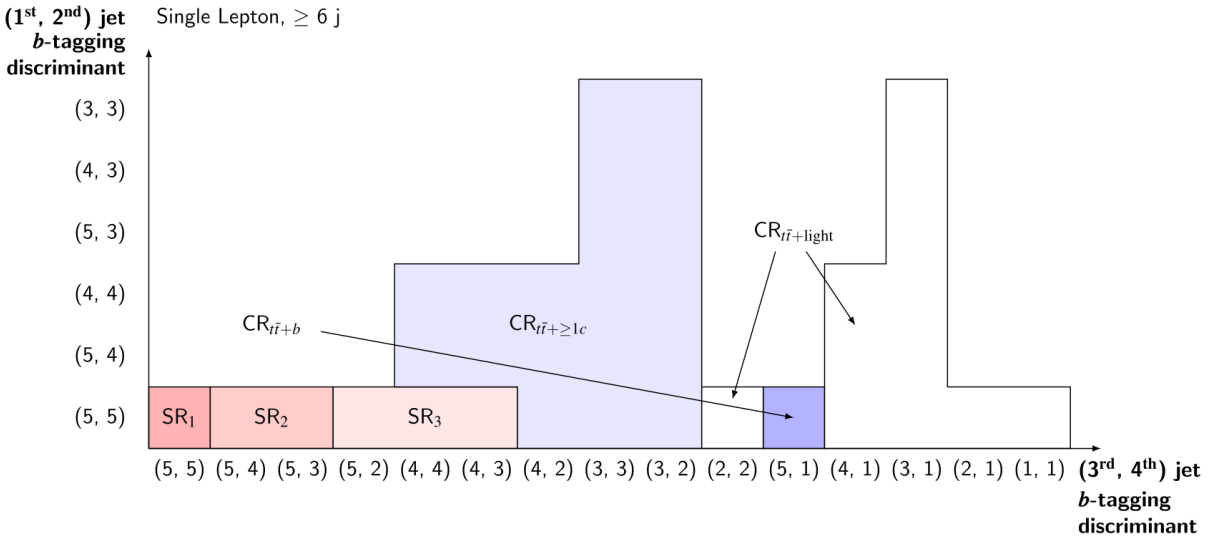
(b) Dileptonic channel.

Figure 5.2: Background compositions for each analysis region of both the semileptonic and dileptonic channels [1].

A representation of the analysis regions for the resolved semileptonic channel is shown in Figure 5.3. A similar representation also showing the background compositions for each analysis region of the resolved semileptonic channel is shown in Figure 5.4.

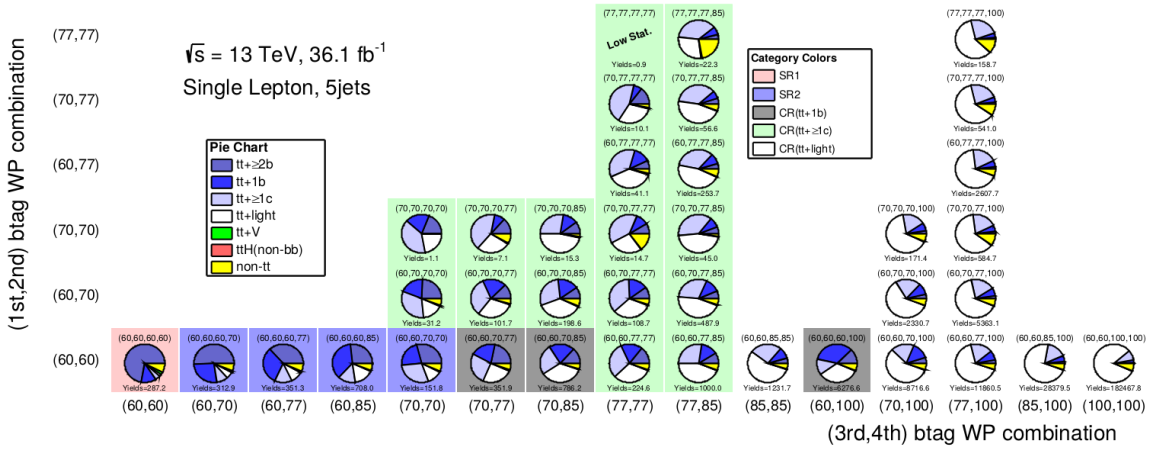


(a)

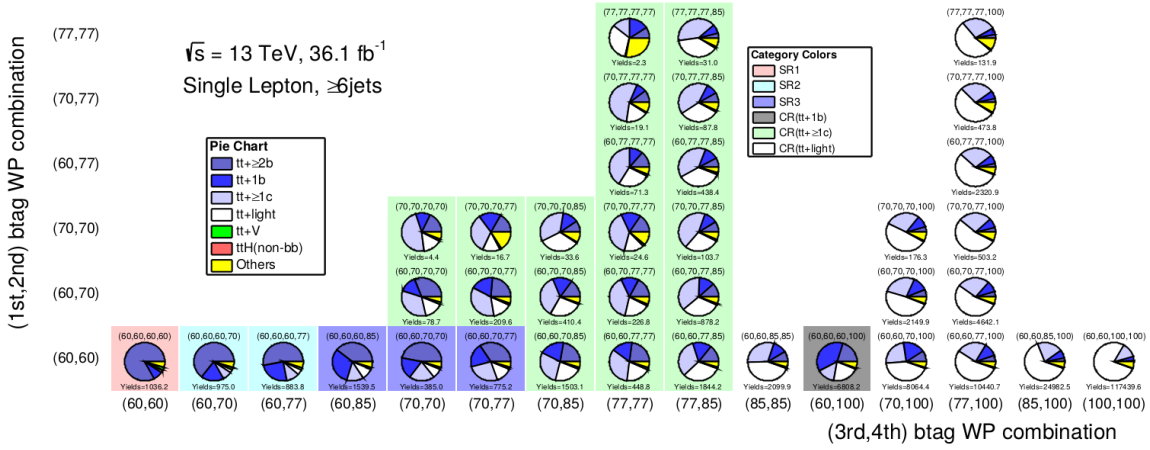


(b)

Figure 5.3: Analysis regions of the resolved semileptonic channel with events with exactly five jets or at least six jets [256]. The values of the b-tagging discriminant for the first and second jets are shown on the vertical axis, while the values of the b-tagging discriminant for the third and fourth jets are shown on the horizontal axis. Red colouring indicates signal-enriched regions while blue colouring indicates signal-depleted regions.



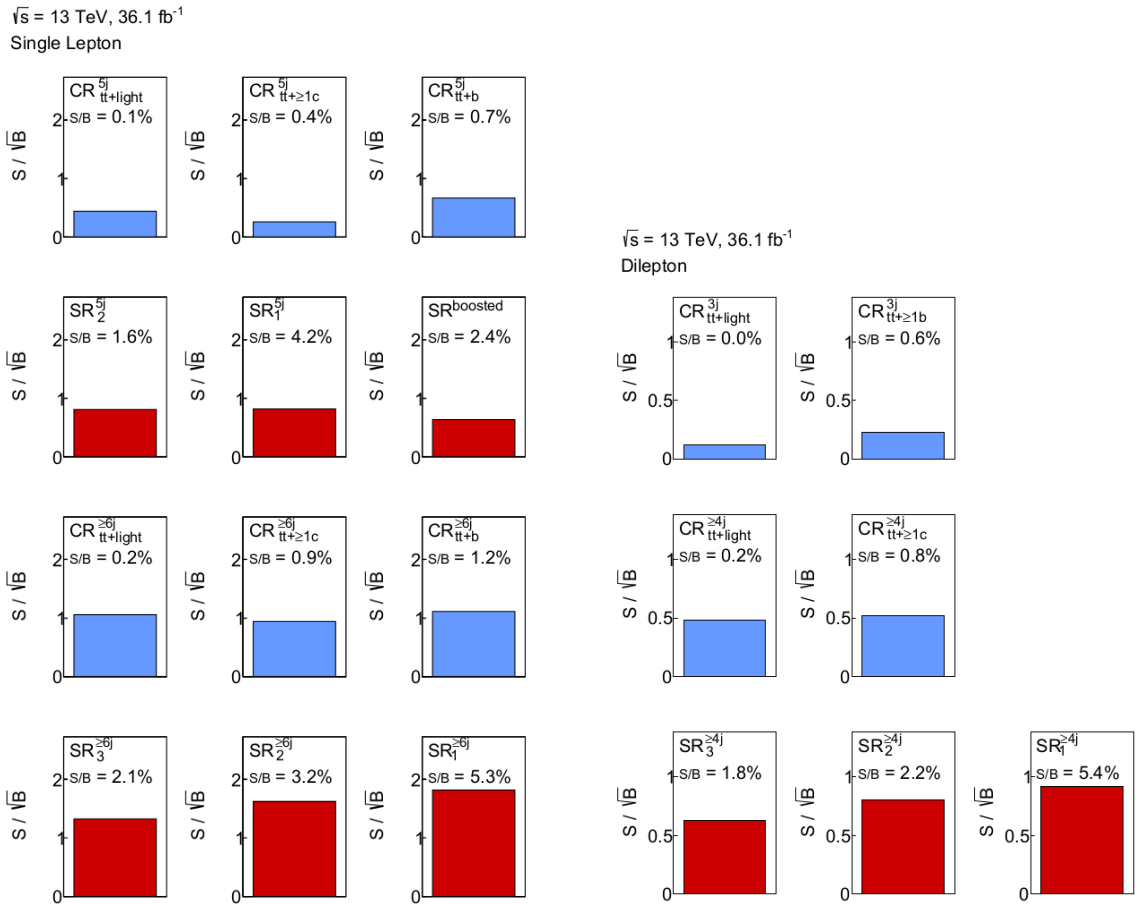
(a)



(b)

Figure 5.4: Sample composition for analysis regions of the resolved semileptonic channel with events with exactly five jets or at least six jets [256].

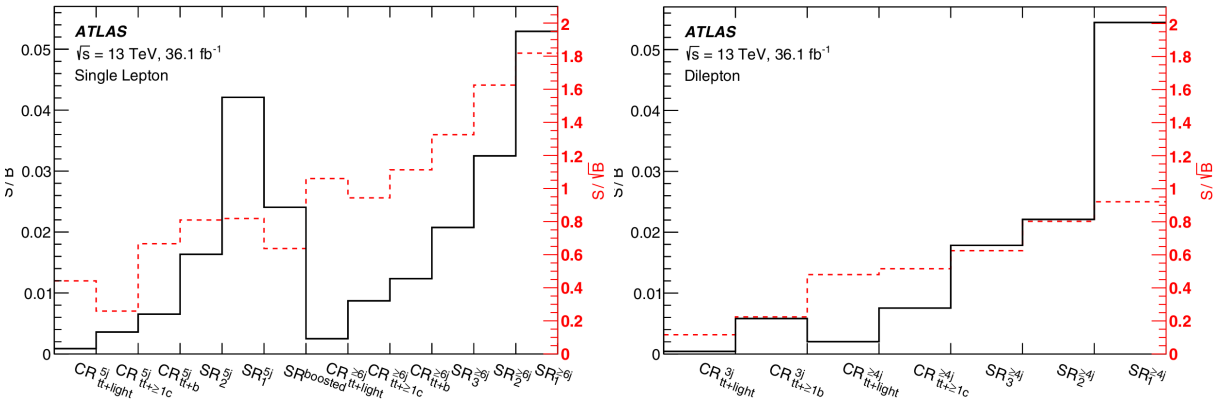
Representations of signal $t\bar{t}H$ purity for each analysis region of both the semileptonic and dileptonic channels are shown in Figures 5.5 and 5.6.



(a) Semileptonic channel.

(b) Dileptonic channel.

Figure 5.5: $S/\sqrt{B}$ and S/B ratios for each analysis region of both the semileptonic and dileptonic channels, where S is the number of signal events selected and B is the number of background events selected [190].



(a) Semileptonic channel.

(b) Dileptonic channel.

Figure 5.6: $S/\sqrt{B}$ (in black) and S/B (in red) ratios for each analysis region of both the semileptonic and dileptonic channels, where S is the number of signal events selected and B is the number of background events selected [1].

The expected fractions of signal events arising from the different Higgs decay processes are shown for all analysis regions in Figure 5.7. The $t\bar{t}H$ $(b\bar{b})$ decay corresponds to 96 % of signal $t\bar{t}H$ events in the signal regions of the resolved semileptonic channel, 86 % in the boosted semileptonic signal region and 89 % in the dileptonic channel. A summary of the semileptonic channel analysis regions showing expected yields and observed data is shown in Figure 5.8.

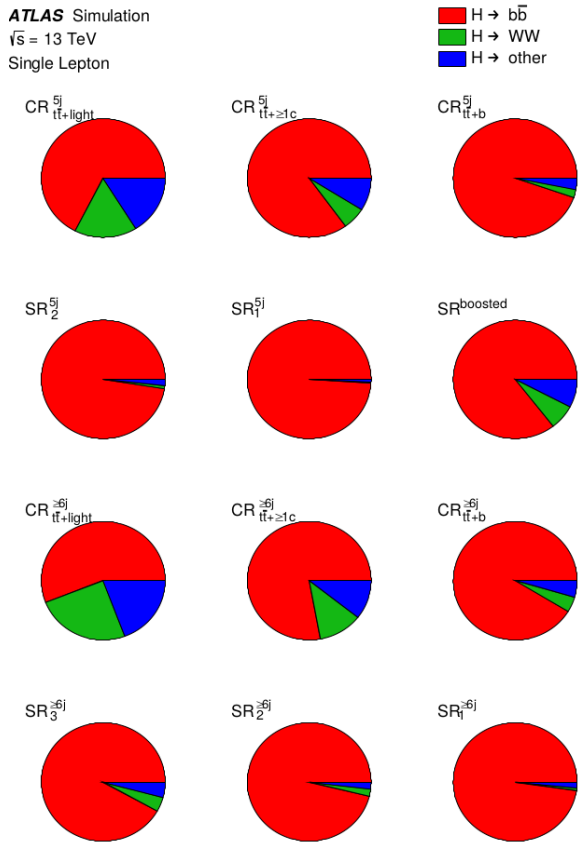


Figure 5.7: The fractional composition of the signal $t\bar{t}H$ in the semileptonic channel [257].

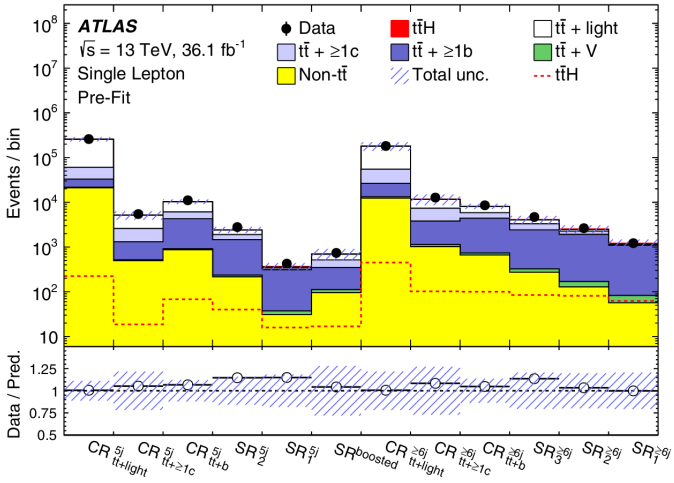


Figure 5.8: Expected yields and observed data for the semileptonic channel analysis regions [1]. The signal $t\bar{t}H$ contribution is normalised to the expected cross-section and is shown as solid red in each bin and as a dashed red line for clarity.

While there are some differences between the 2018 and 2020 analyses, a diagram of the 2020 analysis region selections is shown in Figure 5.9 to give a sense of the complexity of the selections typical of $t\bar{t}H$ analyses.

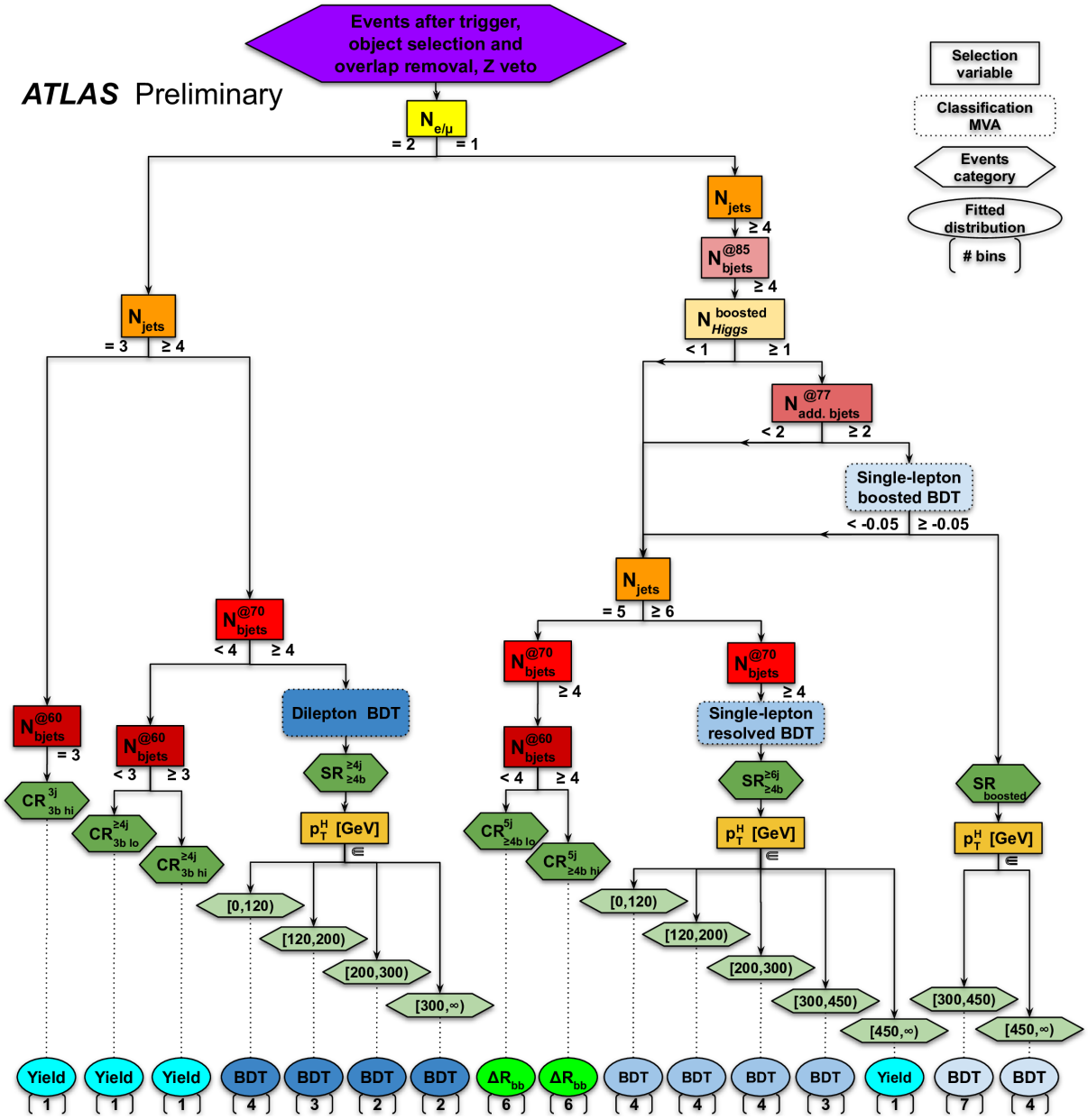


Figure 5.9: Signal and control regions for events with six or more jets and various b-tags in the semileptonic and dileptonic channels [258].

5.6 Multivariate analysis strategy

5.6.1 Overview

Following the selection of events and their categorisation into analysis regions, there are various separate analysis regions in which the signal and background are enhanced with regards to the reducible backgrounds. Various multivariate methods are used in order to further improve sensitivity. Variables with powerful discrimination between signal and background are defined using multivariate methods and these are used as inputs to classification boosted decision tree (BDT) methods (described in Chapter 8), the outputs of which are used as the final discriminants in all signal regions, combining multiple variables in order to have an enhanced separation of signal and background.

In the resolved analysis regions, the signal or background hypothesis is constructed by reconstructing the particles in the $t\bar{t}\left(H \to b\bar{b}\right)$ and $t\bar{t}+$ jets final states from the leptons, jets and missing transverse energy in the events. In the boosted analysis region, one of the large-R jets is tagged as a Higgs boson candidate in event selection, so no reconstruction method is needed for it.

Various multivariate methods are applied, each approaching the process reconstruction differently, for the overall purpose of discriminating signal from background. Their combination results in an overall stronger discriminant. Following this, the events are classified as signal and background by classification BDTs. Input variables to the classification BDTs comprise variables from the reconstruction methods and other variables in order to maximise sensitivity. The input variables are described and listed in Section 8.2.1.

5.6.2 Reconstruction BDT

Reconstruction BDTs are used in the resolved semileptonic and dileptonic channels in order to assign partons to jets in the $t\bar{t}\left(H \to b\bar{b}\right)$ final state, resulting in candidates for Higgs bosons, t-quarks and $W^{\pm}$-bosons in events. The candidates are built by combining leptons and jets. The use of b-tagging information reduces the total amount of permutations to be considered. For each permutation, invariant masses of candidate objects and combinations of them, as well as angular distances, are calculated and these are inputs to the reconstruction BDTs. The BDTs are trained on simulated events (for which the correct permutations are known). Two versions of the reconstruction BDTs are used. One version has all observables directly related to Higgs bosons removed in order to improve the discriminating power of variables directly related to the Higgs boson, such as the candidate Higgs boson invariant mass, at the cost of reducing reconstruction efficiency. In the semileptonic analysis region $\mathrm{SR}_1^{\geq 6j}$, the Higgs boson is reconstructed correctly in about 48 % of events using the reconstruction BDT with information about

the Higgs boson kinematics included, while in the dileptonic analysis region $\mathrm{SR}_1^{\geq 4j}$, the Higgs boson is reconstructed correctly in about 49 % of events [1]. In the boosted analysis region, the reconstruction is not needed as the Higgs boson and t-quark candidates are chosen as the selected large-R jets.

5.6.3 Likelihood discriminant

A likelihood discriminant (LHD) method considers the kinematics across possible jet-parton assignments. It is calculated using the kinematic differences between signal and background and is used in all resolved semileptonic signal regions [259]. It is calculated as a product of probability density functions for the signal and background hypotheses, with the particles in each hypothesis reconstructed from jet assignments to the Higgs boson and the t-quarks,

$$\mathrm{LHD} = \frac{p_{_s}}{p_{_s} + p_{_b}}, \tag{5.6}$$

where $p_{_s}$ and $p_{_b}$ are the probabilities for each event under the signal and background hypotheses respectively, calculated from the weighted average of the products of the probability density functions per combination. The probability density functions are built for various invariant masses, such as that of the reconstructed Higgs boson and t-quarks, and also for angular distributions using the leptons, jets and missing transverse energy of each event. All possible jet-parton assignments are used in the calculation, with combinations weighted by the b-tagging discriminants of individual jet assignments in order to penalise unlikely combinations, such as for cases in which b-tagged jets are matched to light quarks. While information from all possible combinations in an event is used to calculate the likelihood discriminant, no consideration is made of the correlations between variables within individual combinations.

5.6.4 Matrix element method

A matrix element method (MEM) that exploits the full matrix element calculation to separate the signal from background is calculated for events in semileptonic signal regions [260].

Whilst the final states of the signal and background processes are referred to as "irreducible", there are subtle differences between the processes. In signal $t\bar{t}H$ the Higgs boson is a scalar boson while in background $t\bar{t}b\bar{b}$ the gluon is a vector boson, and this results angular distributions with different structures. Theoretically, this information is encoded in the matrix elements of the processes. Leading-order matrix elements from MadGraph5 are used in the calculations. The matrix element method is an attempt to exploit this information, and results in an expression of the likelihood of a single event being generated by a particular process.

$$\mathrm{MEM}_{D1} = \log_{10}\left(\mathcal{L}_s\right) - \log_{10}\left(\mathcal{L}_b\right), \tag{5.7}$$

where $\mathcal{L}_s$ and $\mathcal{L}_b$ are likelihoods calculated for the event for the signal and background hypotheses. The matrix element method takes the four-vectors of the leptons, jets and missing transverse energy as inputs. The matrix element method is a powerful tool that combines theoretical models with event kinematics in an efficient way and it is possible to test multiple model hypotheses with likelihood ratios.

5.6.5 Classification BDT

In each of the signal regions a BDT is used to discriminate between signal $t\bar{t}H$ and background $t\bar{t}$. This is referred to as the classification BDT in the analysis. The output distributions of the classification BDT in the signal regions are used as the final discriminants for the profile likelihood fit. In the control regions, the overall event yield is used as an input to the fit, except in those regions enriched in $t\bar{t}+\geq 1c$ in the semileptonic channel: $\mathrm{CR}^{5j}_{t\bar{t}+\geq 1c}$ and $\mathrm{CR}^{6j}_{t\bar{t}+\geq 1c}$, for which H_{T}, the scalar sum of jet p_{T}, is used to further control the $t\bar{t}+\geq 1c$ background.

The $t\bar{t}H\left(b\bar{b}\right)$ final state features many jets from the Higgs boson and t-quark decays as well as from additional radiations. Many combinations of these jets are possible in the reconstruction of the Higgs boson and the t-quarks. In order to enhance the signal separation, the three multivariate methods described in Sections 5.6.2, 5.6.3 and 5.6.4 respectively are used prior to the classification BDT – the reconstruction BDT, the LHD and the MEM. The outputs of these multivariate methods are used as input variables to the classification BDT in various signal regions. The properties of the Higgs boson and the t-quarks from the reconstruction BDT are used in the definition of additional input variables for the classification BDT. The outputs of the reconstruction BDT, the LHD and the MEM were found to be the most powerful discriminant variables that are inputs to the BDT.

The classification BDT was trained to separate signal $t\bar{t}H$ from background $t\bar{t}$ on a sample that is assumed statistically independent of the sample used for the evaluation. As described and listed in Section 8.2.1, the classification BDT inputs exploit the different kinematics of signal and background events, as well as b-tagging information. General kinematic variables, such as invariant masses and angular separations of pairs of reconstructed leptons and jets are combined with the outputs of various multivariate methods and b-tagging discriminants of jets. In the boosted semileptonic signal region, kinematic variables are constructed from the properties of large-R jets and their constituents. The input variables were selected in a way intended to optimise performance of the classification BDT, and only well-modelled variables were used, and different combinations of input variables are used across the different signal regions.

Both the likelihood and matrix element method discriminants are the most powerful inputs to the classification BDT, followed by the reconstruction BDT output.

5.7 Systematic uncertainties

For full details on systematic uncertainties in the analysis, please refer to Reference [1].

Uncertainties might broadly be categorised into three forms. One is statistical uncertainty. This is a source of imprecision associated with the variability of a measurement process. Repeated measurements under similar conditions may return different results because of the stochastic nature of the features of the observed phenomena and the methods of measurement. One approach to addressing this form of uncertainty is averaging, whereby generally the variability of a measurement is reduced approximately by the square root of the observations. In general, this is the motivation for recording many measurements; the more data collected, the more precise the measurements made. Another form is systematic uncertainty. This is a bias resulting from imperfections in instrumentation and tools of measurement. Another form arises from the flawed assumption that there exists an exact measurement, or from the nature of a quantity under measurement being imperfectly defined. An example of this form of uncertainty can be seen in the investigation of the t-quark. This object is well-defined in quantum field theory, however its mass is not well-defined, as the particle cannot be taken in isolation from its environment. While statistical uncertainty can be addressed in general by the collection of more data, and while the uncertainty that arises from the ill-defined nature of the quantity under measurement is often considered small in high-energy physics, systematic uncertainties often dominate in measurements of high-energy physics analyses.

Most estimations of systematic uncertainties in high-energy physics are the result of a mixture of knowledge, experience, common sense and sometimes intuition; there are often no clear procedures to follow. Frequently, systematic uncertainties are neither defined clearly nor well-separated from statistical uncertainties. Sometimes two statistical uncertainties can be mixed together, for example in trigger efficiencies, which may be determined partially from statistical uncertainties of data. It is usual in contemporary high-energy physics experiments to consider stating for a measurement not only statistical uncertainties and traditional systematic uncertainties, but also other 'external' uncertainties, such as those arising from theoretical uncertainties. Results may be quoted with three, four and more uncertainties. While this may emphasise the impact of different sources on the total uncertainty, a long list of uncertainties does not necessarily enhance the understandability of a result. Just as in the case of higher order corrections in perturbative calculations, there is a limited degree to which we consider systematic uncertainties. In contemporary high-energy physics, the definition of systematic uncertainties might be considered to be

all uncertainties which are not directly due to the statistics of the data. As such, uncertainties arising from such sources as trigger efficiencies (which are derived from measured data) and detector acceptances (which are derived from Monte Carlo simulations) are considered as systematic uncertainties, where in other fields they might be considered statistical uncertainties. A justification for this could be that these uncertainties can be reduced following data-taking by more extensive Monte Carlo production or by new methods of measuring trigger efficiency or by reducing the statistical uncertainties on recorded data by loosening selection criteria. The uncertainty in the estimation of systematic effects (such as background, energy efficiency, energy etc.) is systematic uncertainty. The value of a systematic uncertainty is a Gaussian ~ 68 % confidence interval (as opposed to a 100 % safety tolerance). The line between statistical and systematic uncertainties is somewhat arbitrary, but a someway more pragmatic definition of systematic uncertainties in high-energy physics is measurement uncertainties which are not due to statistical fluctuations in recorded or simulated data samples. So, uncertainties arising from Monte Carlo statistics, such as detector acceptance uncertainties, or uncertainties arising from data statistics, such as trigger efficiency uncertainties, are not treated directly as systematic uncertainties, though in the final result these uncertainties may be expressed separately or may be added to the systematic uncertainty. Thus, sources of systematic uncertainties might include badly-known detector acceptances or trigger efficiencies, detector miscalibrations, badly-known detector resolutions, badly-known backgrounds, uncertainties in simulation or theoretical models, uncertainties in cross-sections, branching fractions, lifetimes, luminosity *etc.*, computational errors and software bugs, personal biases and errors, and other unknown effects on measurements. Some of these are particularly difficult to assess or constrain. A top-down, Aristotelian deduction approach can be taken, whereby an attempt is made to consider all possible sources of systematic uncertainties, together with a bottom-up approach, whereby an attempt is made to scrutinise an analysis for internal consistency, often involving comparisons between data and Monte Carlo distributions and comparisons between data sets recorded under different conditions or at different times.

Typical specifications of systematic uncertainties might be something like the following. For systematic uncertainties relating to theory, it might be stated that the systematic uncertainty is the difference between the outputs of the Pythia and Herwig generators. For systematic uncertainties relating to Monte Carlo simulations, the uncertainties are often unstated but are defined as a Poisson distribution with the mean corresponding to the number of events simulated. For systematic uncertainties relating to detector simulation, it might be stated that the uncertainty on the Jet Energy Scale is 5 % or that the b-tagging efficiency is 20 % for jets with $p_T < 40\,\text{GeV}$. A simple approach to evaluating the impacts of systematic uncertainties is to propagate the systematic uncertainties into results obtained. So, a nominal value μ_{nominal} for a measurement might be defined, $\mu_{\text{nominal}} = \hat{\mu}$, then the measurement at $\pm 1\sigma$ of some systematic uncertainty might be calculated, resulting in up and down shifts, $\mu_{\text{up}} = \hat{\mu}$ (systematic up) and $\mu_{\text{down}} = \hat{\mu}$ (systematic down). These are used to calculate the systematic uncertainty on the measurement

through numerical uncertainty propagation, $\sigma\,(\text{systematic}) = [\mu_{\text{up}} - \mu_{\text{down}}]/2$. This approach is repeated for all systematic uncertainties, with the summation in quadrature resulting in the measurement with the total systematic uncertainty, $\mu_{\text{measurement}} = \mu_{\text{nominal}} \pm \sigma\,(\text{systematic 1}) \pm \sigma\,(\text{systematic 2}) \pm ...$ [229]. In a more formalised approach involving incorporating systematic uncertainties likelihood functions for analysis, the statement of the Jet Energy Scale having a systematic uncertainty of 5 % would be elaborated upon, to a statement that there is a calibration parameter in the likelihood for which a true value is unknown, and for which the distribution of this parameter is Gaussian with a 5 % width, and that the effect of changing the calibration by 1 % corresponds to a coherent increase of 1 % in the energy of the jets of an event.

The $t\bar{t}H\left(b\bar{b}\right)$ analysis is impacted by many sources of systematic uncertainty. All of the sources generally can affect both the normalisation and the shape of the distributions on which they act. Exceptions to this are the luminosity uncertainty, as well as cross-section and normalisation uncertainties on the various processes in the analysis, which affect only the normalisation of the distributions on which they act. For cases of uncertainties impacting only the normalisation, a variation of the related nuisance parameter can still result in a shape variation to the distribution featuring a summation of processes. A nuisance parameter is introduced for each source of uncertainty.

5.7.1 Nuisance parameters

Nuisance parameters are implemented with a Gaussian constraint. A polynomial interpolation is made between the two templates defining the $\pm\sigma$ effect of the systematic variation corresponding to a nuisance parameter. Beyond this range, a linear extrapolation is made for the shape component and an exponential extrapolation is made for the normalisation component, preventing the total yield from samples being calculated as negative. The normalisation component thus behaves effectively as though the extrapolation were linear with a log-normal constraint.

Some nuisance parameters are defined by a variation of the nominal distribution in only a single direction. An example of this is the difference between the nominal $t\bar{t}$ simulation and a variation in which the Monte Carlo generator is replaced by an alternative generator. In the case of such nuisance parameters, the variation is defined as the $+1\sigma$ up effect and the variation is symmetrised in order to determine the template for the -1σ down effect of the nuisance parameter.

A smoothing procedure may be applied to templates defining the up and down effects, with the exception of most of the cross-section and normalisation uncertainties, in order to attempt to remove statistical uncertainties from the templates, which can result in unfair constraints in the profile likelihood fit.

If the impact of a nuisance parameter on a specific sample in a specific analysis region is negligible, it is

removed from the fit. Specifically, a normalisation effect in an analysis region is removed from the fit if it is below 1 %. A shape effect is removed from the fit if the shape size is below 1 %, where a shape size of 1 % corresponds to a change in the normalisation of the nominal distribution by 1 % compared with the mean effect, arising from any bin changing the normalisation.

Statistical uncertainties on the theoretical model correspond to the finite number of events simulated by generators and the finite number of events in the data-driven fake lepton estimate in the semileptonic channel, and these range up to a maximum of approximately 12 %.

5.7.2 Experimental uncertainties

5.7.2.1 Leptons

For the case of electrons, the systematic uncertainties arise from effects associated with the electron trigger, reconstruction, identification and isolation efficiencies, the calibration of the electron energy scale and resolution. There is a total of six nuisance parameters for electron systematic uncertainties.

For the case of muons, the systematic uncertainties arise from the muon trigger, the association of tracks with vertices, identification, identification of low-momentum muons and muon isolation, the calibration of the muon scale and resolution, and additional parameters. There is a total of 15 nuisance parameters for muon systematic uncertainties.

For the case of tauons, some nuisance parameters were considered and found to be negligible.

5.7.2.2 Jets

The systematic uncertainties for jets arise from the jet energy scale (specifically from calibration dependence on jet flavour, the jet position in η, pile-up, jets not in the calorimeters and high-momentum jets), jet energy resolution and jet vertex tagging. There is a total of 23 nuisance parameters for jet systematic uncertainties. As events considered in the analysis feature many jets, the contributions of jet systematic uncertainties are significant.

There are many systematic uncertainties arising from jet flavour tagging. The b-tagging efficiency and the mis-tag rates for c and light jets are measured for all working points used in the analysis and combined into a global calibration. Associated uncertainties consider correlations between different working points and are dependent on jet transverse momentum. Systematic uncertainties associated with b-tagging are split into 30 sources. The mis-tag rate for light jets is dependent on jet pseudorapidity. Systematic uncertainties associated with mis-tag rates for c-jets are split into 30 sources while systematic uncertainties associated with mis-tag rates for light-jets correspond to 80 nuisance parameters. The b-tagging

systematic uncertainty ranges from approximately 2 % to 10 % while the c and light-jet mis-tag rates correspond to systematic uncertainties in the ranges of approximately 5 % to 10 % and 10 % to 50 % respectively.

5.7.2.3 Missing transverse energy

Systematic uncertainties associated with the energy of objects used in calculating the missing transverse energy are propagated to the measured missing transverse energy and systematic uncertainties associated with calorimeter energy deposits used in the calculation of the missing transverse energy correspond to three nuisance parameters, and these uncertainties are not associated with any specific reconstructed objects.

5.7.3 Signal and background modelling

This section details systematic uncertainties associated with the nominal modelling.

5.7.3.1 Signal $t\bar{t}H$

The systematic uncertainty on the cross-section of the signal is comprised of QCD scale uncertainties and uncertainties associated with the PDF and strong coupling [92]. Uncertainties associated with Higgs boson branching fractions into $b\bar{b}$, $WW^{\pm}$ and other final states are accounted for with three nuisance parameters. The absolute uncertainty on the branching fraction for $b\bar{b}$ is $^{+1.2\ \%}_{-1.3\ \%}$. The various categories do not have a significant effect on the distribution shape, so only their normalisation effect is considered. An uncertainty associated with the parton showering and hadronisation modelling is introduced also.

5.7.3.2 Background $t\bar{t}$ + jets

There are many systematic uncertainties associated with the modelling of $t\bar{t}$ + jets. The systematic uncertainty on the cross-section of the background is 6 % and this is implemented as one nuisance parameter and correlated for all $t\bar{t}$ categories [248]. All other uncertainties associated with $t\bar{t}$ are implemented separately for the various $t\bar{t}$ components.

The normalisations of the $t\bar{t}+ \geq 1b$ and $t\bar{t}+ \geq 1c$ sub-components correspond to two nuisance parameters and these are free-floating values in the fit. Samples used to define other systematic uncertainties are reweighted such that the fractions of the $t\bar{t}+ \geq 1b$, $t\bar{t}+ \geq 1c$ and $t\bar{t}$ + light processes they contain correspond to the nominal $t\bar{t}$ sample and such that the $t\bar{t}+ \geq 1b$ sub-components match Sherpa4F generator predictions. Samples used for the dedicated uncertainties that describe the $t\bar{t}+ \geq 1b$ and sub-component fractions are not reweighted to match generator predictions.

Multiple alternative samples are generated for the $t\bar{t}$ background. Differences between the nominal samples and those of another generator are used as the systematic uncertainty associated with the choice of NLO event generator, a difference that varies the parton showering and hadronisation model also. Alternative samples are compared to quantify the uncertainty related to initial and final state radiation. These systematic uncertainties correspond to nine nuisance parameters that are split between $t\bar{t}$ categories.

A summary of the nuisance parameters for $t\bar{t}$ modelling is shown in Table 5.2. For thorough details of the systematic uncertainties associated with $t\bar{t}+$ jets, please refer to Reference [1].

Systematic sources	How evaluated	$t\bar{t}$ categories
$t\bar{t}$ cross-section	± 6 %	All, uncorrelated
$k\,(t\bar{t}+\geq 1b)$	Free floating $t\bar{t}+\geq 1b$ normalisation	$t\bar{t}+\geq 1b$
$k\,(t\bar{t}+\geq 1c)$	Free floating $t\bar{t}+\geq 1c$ normalisation	$t\bar{t}+\geq 1c$
Sherpa 5F versus nominal	NLO event generator choice	All, uncorrelated
Parton shower and hadronisation	POWHEG + Herwig 7 versus POWHEG + Pythia 8	All, uncorrelated
Initial and final state radiation	Variations of μ_R, μ_F and additional POWHEG and Pythia 8 parameters	All, uncorrelated
$t\bar{t}+\geq 1b$ Sherpa 4F versus nominal	Comparison of $t\bar{t}+b\bar{b}$ NLO (4F) versus POWHEG + Pythia 8 (5F)	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1c$ 3F versus 5F scheme	MagGraph5_aMC@NLO + Herwig++ versus POWHEG + Pythia 8	$t\bar{t}+\geq 1c$
$t\bar{t}+\geq 1b$ scale variations	Three categories	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ shower recoil scheme	Alternative model scheme	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ PDF (MSTW)	MSTW versus CT10	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ PDF (NNPDF)	NNPDF versus CT10	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ underlying event	Alternative set of tuned parameters for the underlying event	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 3b$ normalisation	± 50 %	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ multi-parton interaction	± 50 %	$t\bar{t}+\geq 1b$

Table 5.2: Summary of systematic uncertainties associated with $t\bar{t}+$ jets modelling [17, 18, 257].

5.7.3.3 Fake and non-prompt leptons

In the semileptonic channel, the data-driven fake lepton estimate corresponds to a 50 % systematic uncertainty. For this channel, it is split into 6 components, with electron events and muon events treated separately. A separate nuisance parameter is assigned for each type of analysis region: the boosted region, resolved regions with exactly five jets and resolved regions and six or more jets. In the dileptonic channel, the simulation-based fake lepton estimate corresponds to a single 25 % normalisation uncertainty. There are, thus, seven nuisance parameters associated with the fake lepton estimate.

Chapter 6

Analysis framework for the $t\bar{t}H\left(b\bar{b}\right)$ search

> While asleep, I had an unusual experience. There was a red screen formed by flowing blood, as it were. I was observing it. Suddenly a hand began to write on the screen. I became all attention. That hand wrote a number of elliptic integrals. They stuck to my mind. As soon as I woke up, I committed them to writing.
>
> Srinivasa Ramanujan

This chapter presents details on TTHbbLeptonic, the combined analysis and derivation software and computing framework of the $t\bar{t}H\left(b\bar{b}\right)$ analysis – details on its structure and functionality, historical development, and the organisation of its development. Also are presented details on TRExFitter, the software used for fitting (as described in Chapter 5), together with other details on software related to TTHbbLeptonic and TRExFitter and procedures on software validation and physics validation. Further, this chapter details more recent developments, and offers comments on a plan for high-energy physics software and computing research and development for the 2020s. The author created and led development and coordination of the combined $t\bar{t}H\left(b\bar{b}\right)$ analysis derivation framework TTHbbLeptonic, in addition to many other related contributions, which ranged from vital components to routine software and physics validations. The reader is urged to refer to Chapter 1 for details on specific author roles and contributions in order to avoid impractical and repetitious attributions to the author of extensive work detailed through this chapter.

6.1 Analyses steps

As described in Section 3.3 on abstractive transformations of ATLAS data, more basic forms of data, such as bytestream data from sub-detectors, are transformed progressively into more abstract forms that correspond to objects defined usually with reference to current fundamental physics theory. Events are selected by carefully investigated and tuned specifications and calibrations for the purposes of searching for and testing hypotheses, such as aspects of the Standard Model or BSM physics. These analyses then are reified via statistical analysis into concrete statements about the likelihood or the accuracy of a predictive theoretical model, such as the theory of the existence of the Standard Model Higgs boson. A high-level view of a typical flow of ATLAS data through derivations infrastructure (software which filters data, selecting events needed for different analyses) to analysis infrastructure (where data is divided into manageable sizes for analysis by analysis groups) was shown in Figure 3.27.

The selection of objects, as described in Chapters 4 and 5, is based largely on two considerations. First, the geometric characteristics and thresholds for measurement of energy and momentum are called detector acceptance, and this necessitates object selections based on p_{T} and η, which are incorporated in trigger configurations. Second, object selections should ensure a high reconstruction or trigger efficiency. Electrons, muons and photons are usually subject to isolation requirements together with quality requirements (loose, medium and tight) while jets are usually subject to quality requirements such that they are affected minimally by pile-up and calorimeter noise.

The selection of events can involve application of geometric cuts (such as imposing specific separations between objects), requiring events to have certain values or ranges of observables for the purpose of selecting events which feature the desired final state objects. It is usual to require that events were recorded as part of good runs – runs for which the detector was operational and there were no apparent problems, such as issues with detector components. Event selections that feature only events with specific processes of interest are called exclusive and event selections that feature specific processes of interest together with a collection of other processes are called inclusive.

6.2 ROOT

ROOT ("Rapid Object-Oriented Technology") is an object-oriented data analysis framework aimed at solving data analysis challenges in high-energy physics [261, 262]. It is the most commonly used software to perform high-energy physics analyses [200]. ROOT was developed in the context of the NA49 experiment at CERN and constituted a move to object-oriented programming, motivated in part by similarities possible between programming class hierarchies and hierarchies in physics simulation [263]. NA49 gen-

erated data of approximately 10 TB per run. This rate of data production provided a test environment for the development of ROOT, as the next generation of data analysis. RIVET (Robust Independent Validation of Experiment and Theory) and MADANALYSIS are worthy of mention as frameworks integral to high-energy physics analyses, however they are suited to theory more than experiment and tend to process simulated events as opposed to recorded data [264, 265].

There is a long history preceding ROOT, the development of which was started in 1994. Its immediate predecessor, from 1986, was Physics Analysis Workstation (PAW), which was conceptually similar to ROOT but was written in FORTRAN [266, 267]. The predecessor of PAW was the combined use of HTV (HBOOK and HPLOT) and the CERN Program Library (CERNLIB). It is notable that all of these systems were influenced by René Brun [263]. Some earlier history of computing for collider physics at CERN can be viewed in the video at Reference [268]. For other details on ROOT, please refer to chapter *Software* of the author's M.Sc. thesis [81].

The standard analysis flow in high-energy physics consists of running an experiment-specific analysis framework on data and Monte Carlo samples produced centrally. While the data formats and analysis frameworks are usually based on ROOT, experiment-specific software packages and databases are needed. The output of these frameworks is often flat n-tuples which are used to produce histograms, which are then processed further. In a simple case, this might involve comparisons of data and Monte Carlo simulated samples; in the case of a search there could be limits plots produced.

6.3 ATLAS analysis model

The original conception of the analysis model for ATLAS focused on analysis of derived AODs (Analysis Object Data) and small final n-tuples. In practice, however, for Run-1, AODs and Athena did not work quite as intended for analysis, with physicists writing large n-tuples and analyses being ROOT-based [21]. Large n-tuples are essentially more practical outputs from AODs. During Run-1, there were many different n-tuple formats, there were millions of CPU hours lost because of small mistakes, comparisons between different analyses were difficult, and there was unsustainable consumption of storage.

The format of xAOD data was defined as essentially that of an AOD which is readable by ROOT in order to attempt to address Run-1 software and usability problems [21]. It has a well-supported derivation framework model and has various tools for use in ROOT and Athena. For xAODs, a common Event Data Model is used, which means that there is a standard way of talking about and working with physics objects such as electrons, jets *etc.* in the xAOD data format.

The Run-2 derivation framework defines the conversion of input xAOD data to derived output DxAODs or n-tuples, which groups and users can process. The derivation framework can produce one or several

outputs for a specified input. The 2015 target size for a derivation was 1–10 TB. The data is reduced broadly by removing events (referred to as skimming) and by removing per-event information (referred to as slimming). A target upper limit threshold of 1 % of the xAOD per derivation was defined as a workable fraction of the xAOD input data. Assuming 50 derivations in total, this allows for approximately four versions of each derivation. The target CPU usage is < 40 ms per derivation or < 160 ms per group with a total budget of 2 s per event.

AnalysisBase was developed as the Run-2 common analysis framework of ATLAS, centred around xAODs, and there are analysis frameworks for certain classes of ATLAS analyses, such as AnalysisTop, which are derived from AnalysisBase.

6.4 TTHbbLeptonic Run-2 analysis framework

TTHbbLeptonic is a derivations and analysis framework for the $t\bar{t}H\left(b\bar{b}\right)$ analysis [269]. It is built upon AnalysisTop and integrated with the RootCore build system, the EventLoop package for looping over xAOD physics events, and many other components besides.

Initially, the TTHSemileptonic framework was created by the author for the purposes of the semileptonic $t\bar{t}H\left(b\bar{b}\right)$ analysis. When it became apparent that there was wider appeal for this framework, the more generalised TTHbbLeptonic framework was created by the author, and this framework was further developed as a result of the union of the efforts of the teams of the $t\bar{t}H\left(b\bar{b}\right)$ semileptonic analysis, the $t\bar{t}H\left(b\bar{b}\right)$ dileptonic analysis, the $t\bar{t}H\left(b\bar{b}\right)$ boosted analysis, the $t\bar{t}H\left(b\bar{b}\right)$ all-hadronic analysis, the $H \to 4b$ analysis and the $H^+ \to tb$ analysis [270]. Combined analysis efforts are beneficial, as they can result in modular, easy-to-use common tools, cross-compatibility of analyses and the ability to run all event selections at once. Clear, understandable, accessible and reliable infrastructure invites others to contribute. TTHbbLeptonic was developed to have good, clear code with clear, robust validation procedures (including diagnostics scripts, output comparisons tools *etc.*), branch-based development, per-version documentation that is clear and detailed, to be welcoming to new users and analysis groups, and to be released early and often, and linked tightly with AnalysisTop releases. The $t\bar{t}H\left(b\bar{b}\right)$ analysis relies on object definitions from the Top Reconstruction Group [186]. These objects were implemented in TTHbbLeptonic, together with event selections, variables and tools. Subject to the ATLAS Data Access Policy, all of the software developed was released as free/libre and open source [271]. TTHbbLeptonic is shown in the context of the overall analysis infrastructure in Figure 6.1.

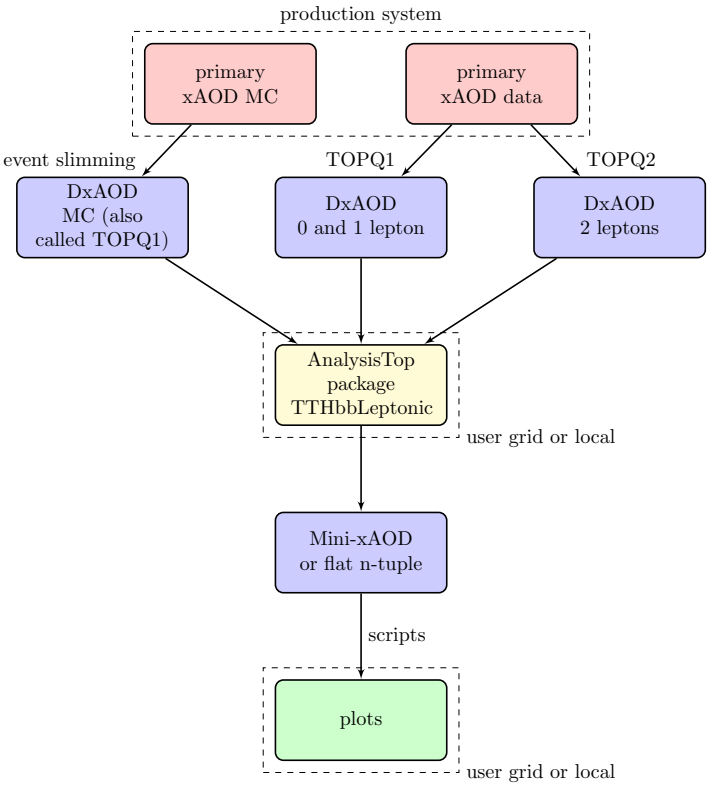


Figure 6.1: TTHbbLeptonic shown in the context of the analysis infrastructure.

6.4.1 Tools for producing outputs, utilities for running over outputs, and documentation

TTHbbLeptonic was built on AnalysisTop with many additional tools added for selections, studies, plotting *etc*. It features a pre-selection tool for variables, a variables calculation tool, which serves as an accessible way to calculate and add new variables, curated sets of output variables corresponding to various analyses, such as the semileptonic and dileptonic analyses, various multiplicities and flags, and $t\bar{t}H$ $(b\bar{b})$ Run-1 and Run-2 variables (among many other components). A variety of these variables is listed in Chapter 8, and there are many other variables besides, such as variables corresponding to PDG ID information (from the PDG particle numbering scheme which assigns a unique code to each type of particle) [43]. It features a Monte Carlo reweighting tool, multivariate analysis tools to loop over jet combinatorics (to save combinations of variables for use with multivariate methods), an implementation of the Tag Rate Function (TRF) method, and various tools for the boosted analysis (such as definitions of large-R jets and jet substructure variables). It features dynamic keys (for accessible changing of specifications such as b-tagging algorithms and b-tag cut), custom object loaders, and TMVA and

NeuroBayes integration. It features a general reweighting tool with which a distribution in a ROOT file can be reweighted according to a configuration. It features a jet truth-matching tool that matches jets to partons (a tool that is generator-dependent and so may be run only on some samples). TTHbbLeptonic also interfaces with multivariate methods such as the reconstruction BDT and the classification BDT for the semileptonic and dileptonic channels.

TTHbbLeptonic also features a dedicated Python virtual environment and a number of Python-based tools, such as an LHC computing grid data output merging script (to 'graft' ROOT output trees together as needed), many plotting scripts for histograms, comparisons, stack plots, palettes *etc.*, a run report generator and an automatic timing measurement method. As part of an effort to make the framework accessible for new users and groups, extensive, per-version documentation was written and kept up to date as an essential part of the release checklist. Also as part of this effort, computer environment examination (to enable developers to briskly assess problems for users), installation and setup scripts were developed, as was a basic graphical user interface, as shown in Figure 6.2. There were also scripts for automatic generation of cuts configuration files, example datasets *etc.* There was also a tool, ttH-offline-wrapper, for conversion of TTHbbLeptonic output to a format suitable for the TRExFitter fitting infrastructure. There were developments to ensure compatibility with CERN Git and Subversion version control systems. Collaboration of the author with Lukas Heinrich and Kyle Cranmer resulted in containerisation of the analysis framework using Docker and associated software.

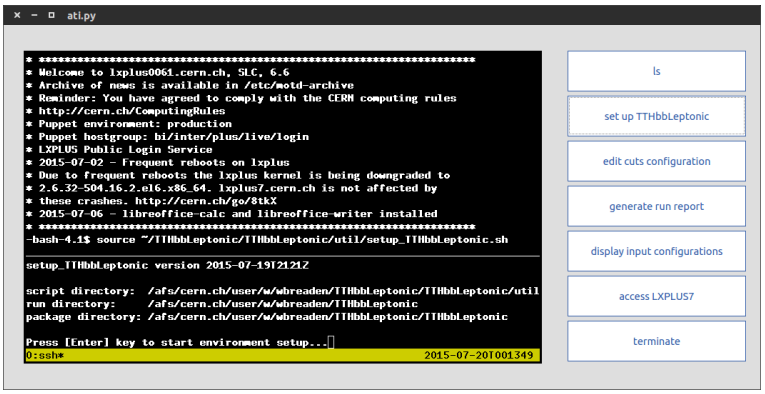


Figure 6.2: TTHbbLeptonic Analysis Terminal Interface (ATI), a sort of GUI terminal, intended particularly for new users.

Included in documentation were tutorials and videos intended to guide new users and new developers. An example framework usage video is at Reference [272].

TTHbbLeptonic features sample lists and large number of definitions of variables, which are added by

TTHbbLeptonic to the output.

Tag Rate Functions (TRFs), which are functions of η and p_{T}, represent event weights equivalent to probabilities that events contain a desired number of b-jets, exclusively or exclusively. The TTHbbLeptonic method of computation of TRF weights is dependent on the truth flavour of jets, so this method can be applied only in simulation. If indirect/TRF tagging is used, TRF weights are applied as multiplicative factors to overall event weights. This procedure then replaces the explicit b-tagging requirement. The net effect of the indirect tagging is that Monte Carlo statistics are recovered while the b-tagging selection efficiency is mimicked by TRF reweighting. TRF-based tagging is supported for the $t\bar{t}H\left(b\bar{b}\right)$ resolved semileptonic and dileptonic analysis.

TTHbbLeptonic features tools for running production on the LHC computing grid. It also features a tool to run production across many LXPLUS nodes as a backup to the grid.

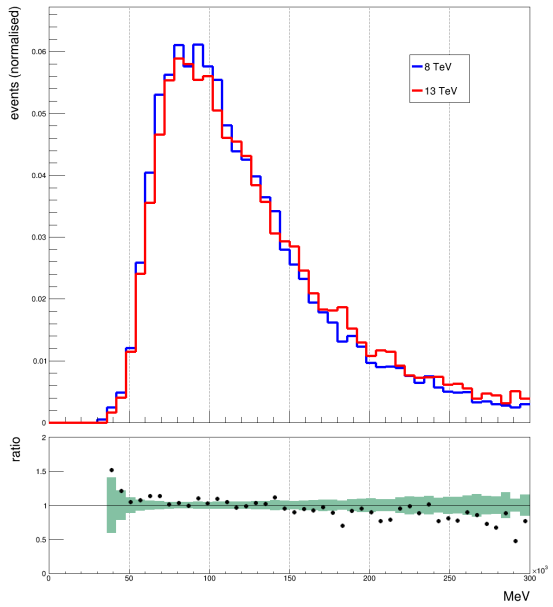
Output detailing the numbers of events remaining following various cuts implemented by TTHbbLeptonic on a sample for the purposes of a cut-flow challenge validation (one of many conducted as both procedural and software validations) is shown in Table 6.1.

Cut	Events remaining
INITIAL	40000
GRL	40000
GOODCALO	40000
TRIGDEC	9841
EL_N 25000 >= 1	7900
EL_N 25000 == 1	7639
MU_N 25000 == 0	7039
TRIGMATCH	7012
JETCLEAN LooseBad	7011
JET_N 25000 >= 1	6976
JET_N 25000 >= 2	6714
JET_N 25000 >= 3	5677
JET_N 25000 >= 4	3814
JET_N 25000 == 4	1937
MV2C20_N -0.4434 == 2	810
MV2C20_N -0.4434 == 3	88
MV2C20_N -0.4434 >= 4	2
JET_N 25000 == 5	1133
MV2C20_N -0.4434 == 2	528
MV2C20_N -0.4434 == 3	76
MV2C20_N -0.4434 >= 4	5
JET_N 25000 >= 6	744
MV2C20_N -0.4434 == 2	345
MV2C20_N -0.4434 == 3	63
MV2C20_N -0.4434 >= 4	6

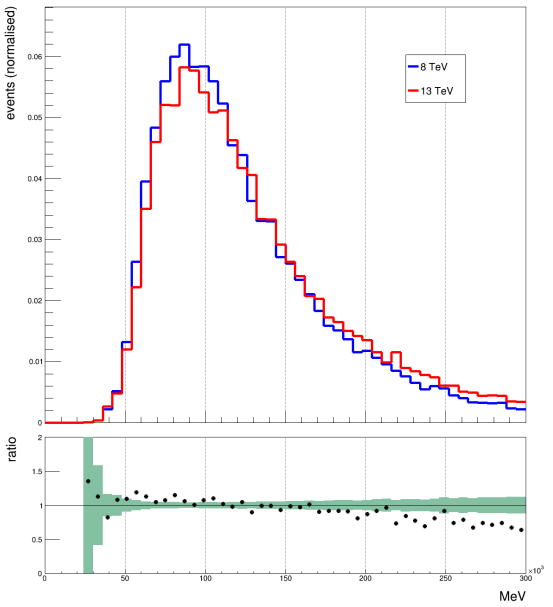
Table 6.1: Example of cut-flow event challenge validation from 2015.

There were many extensive checks of TTHbbLeptonic in order to continuously validate it over its formative initial development and during the years following this. Such validations were completed on a rota by various people in the development team, including the author. Plots associated with some representative examples of these validations completed by the author are included in this chapter.

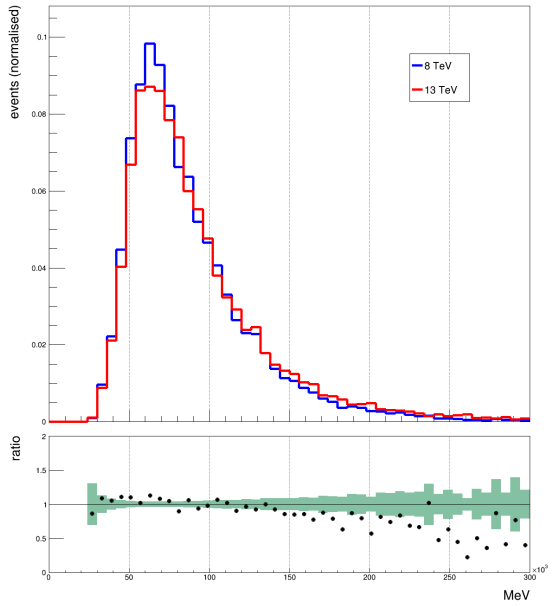
One typical validation involved comparison, for the purposes of consistency, of 8 TeV derivations with 13 TeV derivations, some observables plots of which are shown in Figures 6.3 and 6.4.



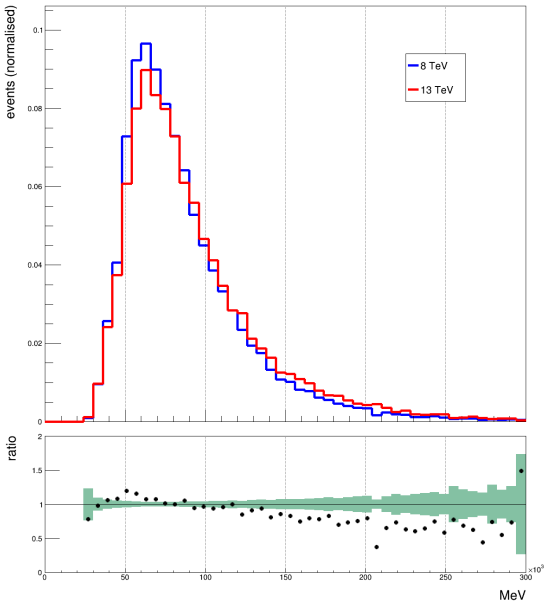
(a) Comparison of p_{T} of jet 1 for electron selection at 8 TeV and 13 TeV.



(b) Comparison of p_{T} of jet 1 for muon selection at 8 TeV and 13 TeV.

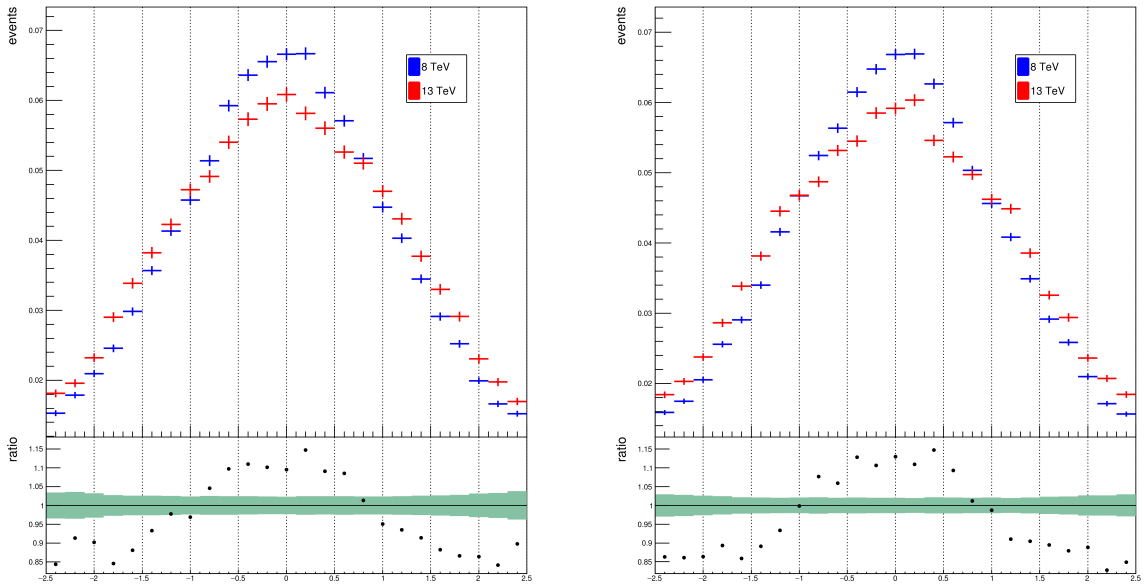


(c) Comparison of p_{T} of jet 2 for electron selection at 8 TeV and 13 TeV



(d) Comparison of p_{T} of jet 2 for muon selection at 8 TeV and 13 TeV

Figure 6.3: Comparison of jet p_{T} for semileptonic channel electron and muon selections at 8 TeV (shown in blue) and 13 TeV (shown in red) [273].

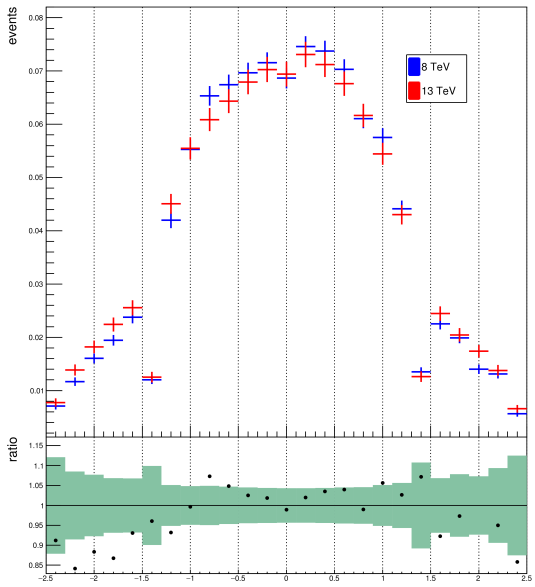


(a) Comparison of jet η for electron selection at 8 TeV and 13 TeV.

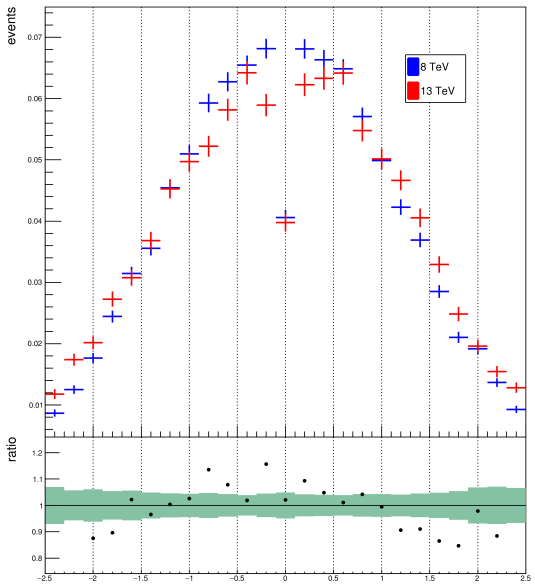
(b) Comparison of jet η for muon selection at 8 TeV and 13 TeV.

Figure 6.4: Comparison of jet η for semileptonic channel electron and muon selections at 8 TeV (shown in blue) and 13 TeV (shown in red) [273, 274, 275].

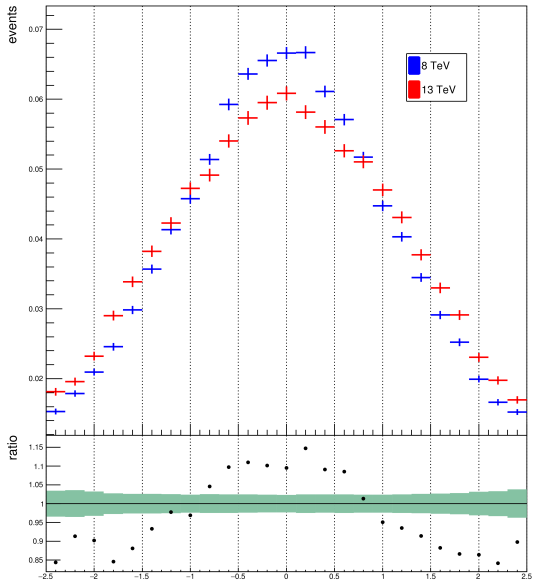
In the change from 8 TeV to 13 TeV there is a cross-section increase by a factor of almost 3.5. Various distributions associated with leptons and jets are compared with respect to 8 TeV and 13 TeV in Figures 6.5, 6.6, 6.7 and 6.8. The Jet Vertex Fraction (JVF) shown is a discriminant derived by combining tracks and their primary vertices with calorimeter jets. It expresses a probability that a jet originated from a particular vertex.



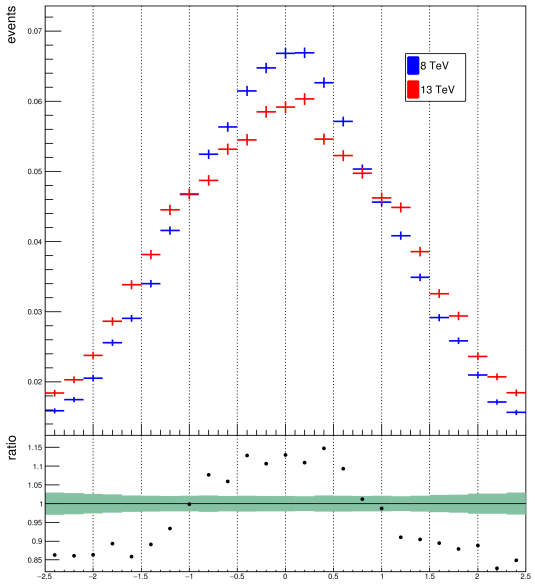
(a) Electron η for electron selection.



(b) Muon η for muon selection.

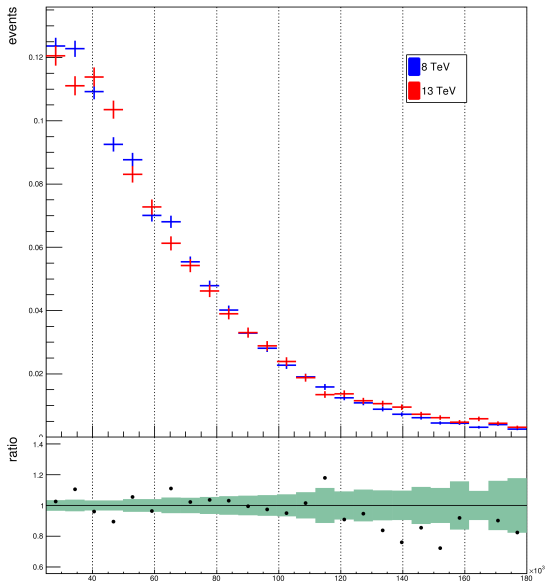


(c) Jet η for electron selection.

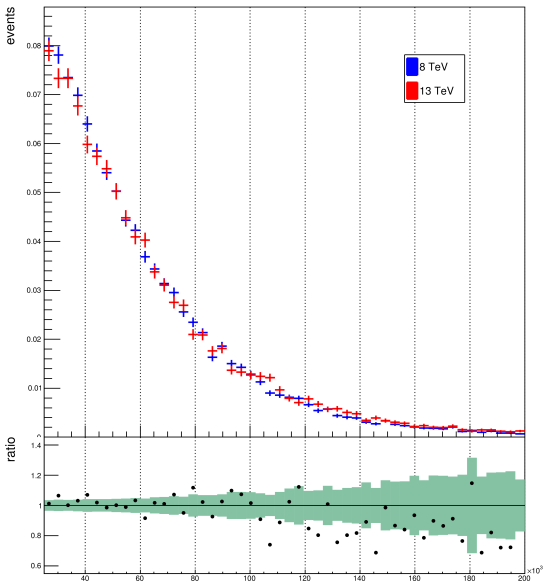


(d) Jet η for muon selection.

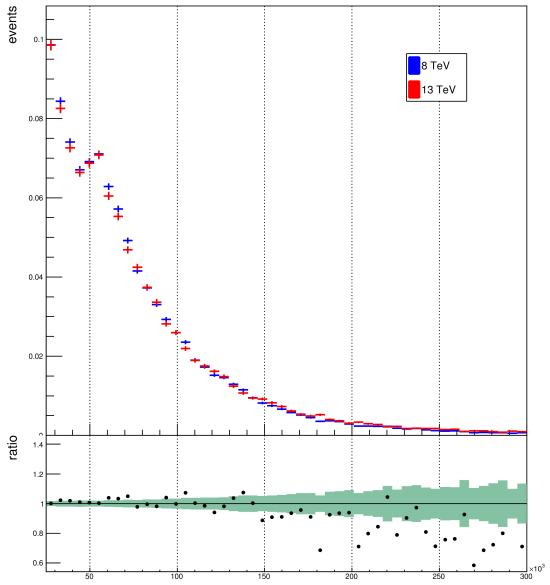
Figure 6.5: Comparisons of electron, muon and jet η for semileptonic channel electron and muon selections at 8 TeV (shown in blue) and 13 TeV (shown in red).



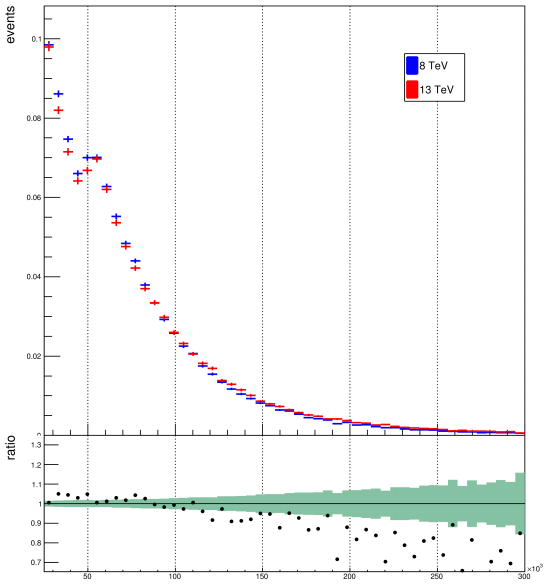
(a) Electron p_{T} for electron selection.



(b) Muon p_{T} for muon selection.

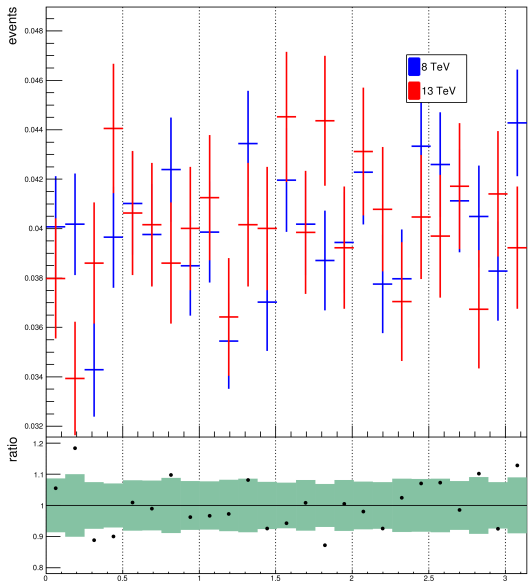


(c) Jet p_{T} for electron selection.

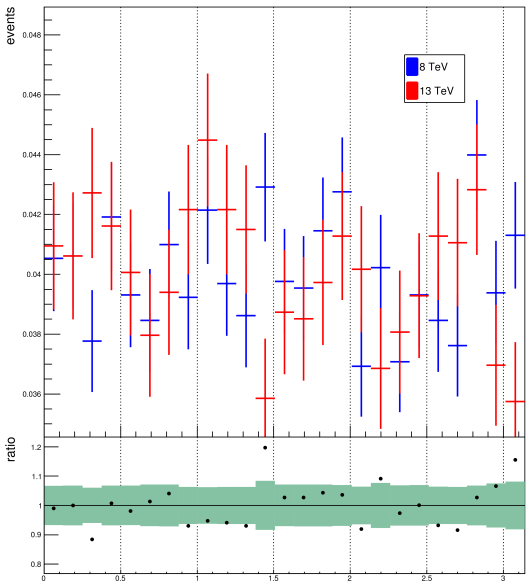


(d) Jet p_{T} for muon selection.

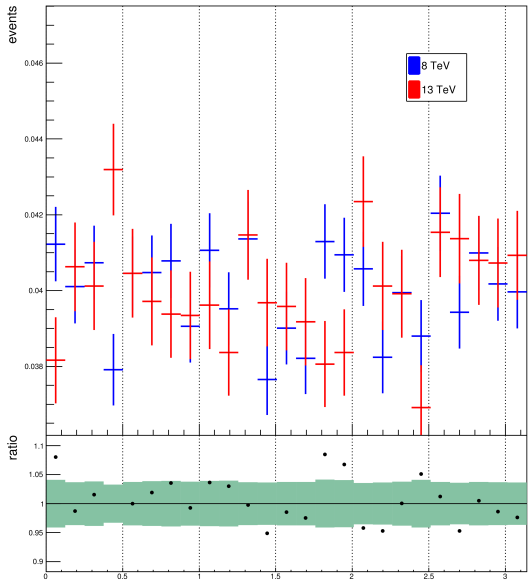
Figure 6.6: Comparisons of electron, muon and jet p_{T} for semileptonic channel electron and muon selections at 8 TeV (shown in blue) and 13 TeV (shown in red).



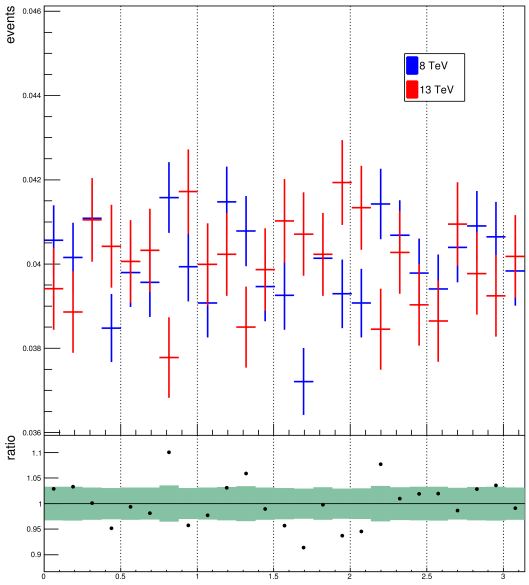
(a) Electron ϕ for electron selection.



(b) Muon ϕ for muon selection.

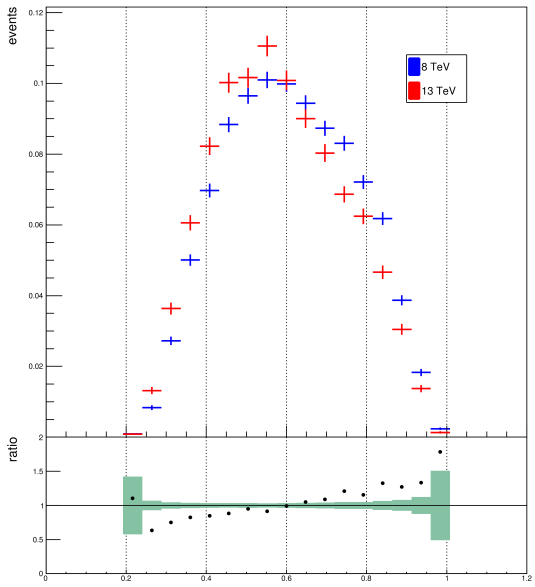


(c) Jet ϕ for electron selection.

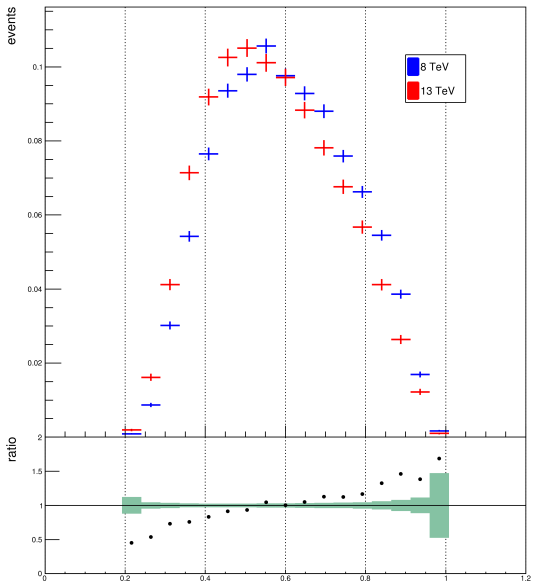


(d) Jet ϕ for muon selection.

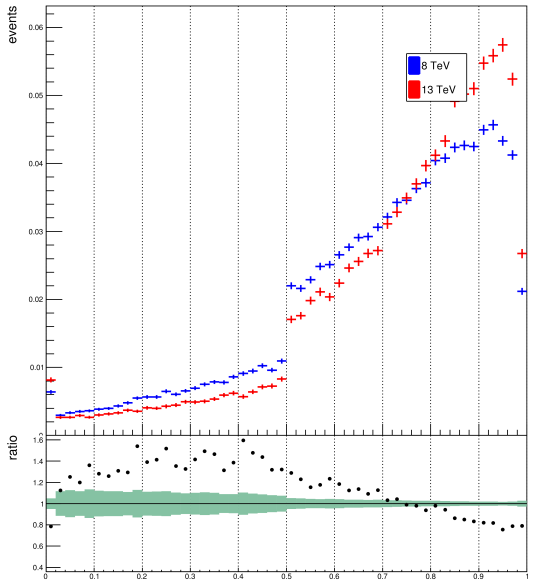
Figure 6.7: Comparisons of electron, muon and jet ϕ for semileptonic channel electron and muon selections at 8 TeV (shown in blue) and 13 TeV (shown in red).



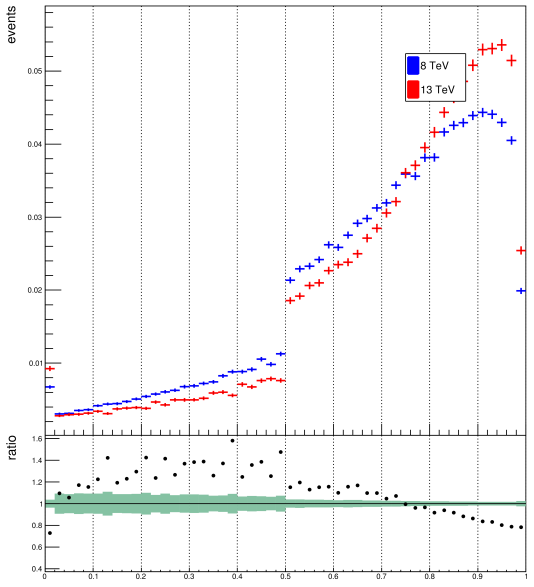
(a) Centrality for electron selection.



(b) Centrality for muon selection.



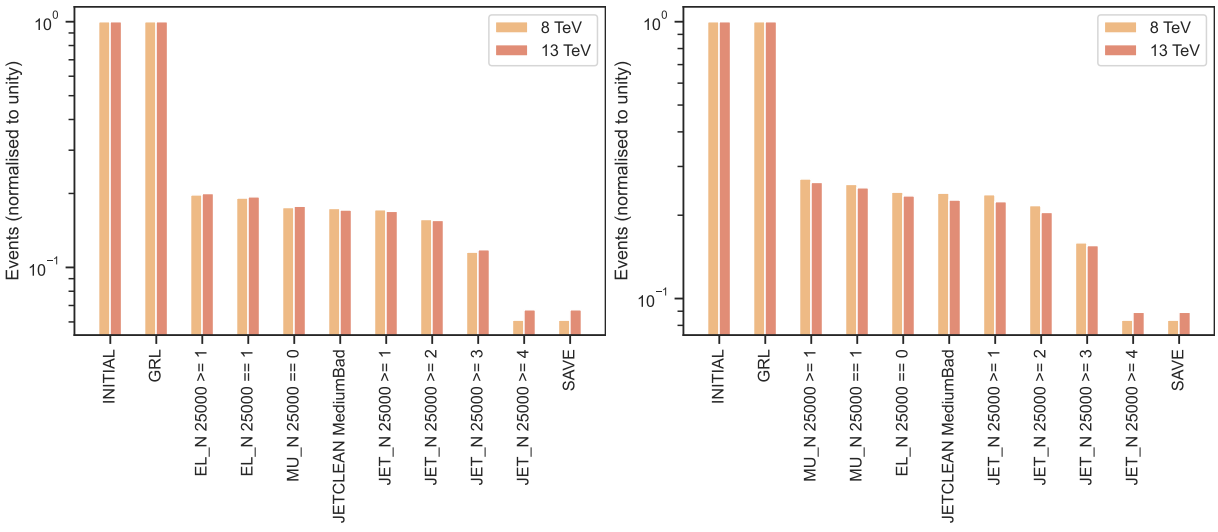
(c) JVF for electron selection.



(d) JVF for muon selection.

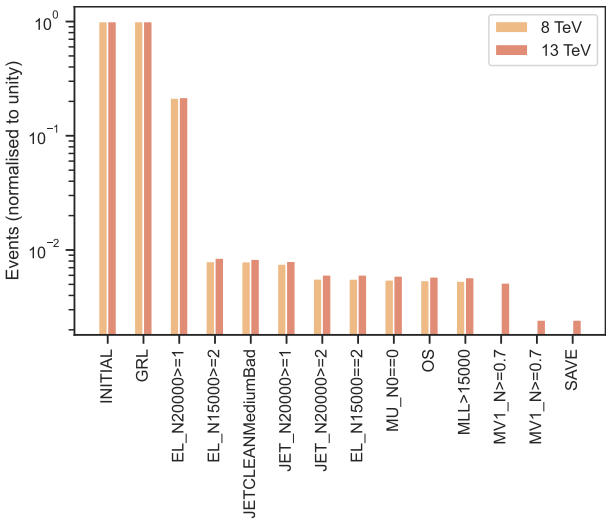
Figure 6.8: Comparisons of centrality and Jet Vertex Fraction, for semileptonic channel electron and muon selections at 8 TeV (shown in blue) and 13 TeV (shown in red).

A regular validation ensured that event yields through cut-flows for 8 TeV and 13 TeV samples were consistent or distinct in expected ways. Figure 6.9 shows a representative example validation of this form for both the semileptonic and dileptonic channels, in this case showing agreement for the most part in the event yields arising from the respective cut-flows (and also showing some small discrepancies, possibly accounted for by mismodelling of t-quark and $t\bar{t}$ systems) [276, 277]. In this case, there had been a new generation of POWHEG-Pythia samples. A baseline comparison was needed prior to a development that constituted modularising the TTHbbLeptonic framework event saver formalism defined by AnalysisBase such that it accommodated multiple analysis channels and new analyses in a way that would make further development more efficient and less likely to induce errors.

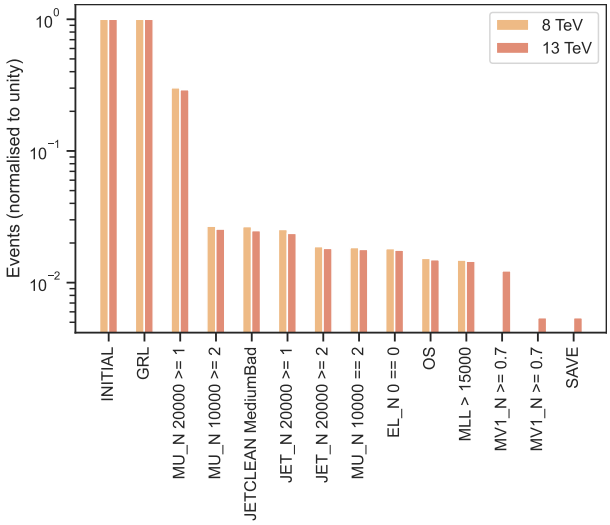


(a) e+jets.

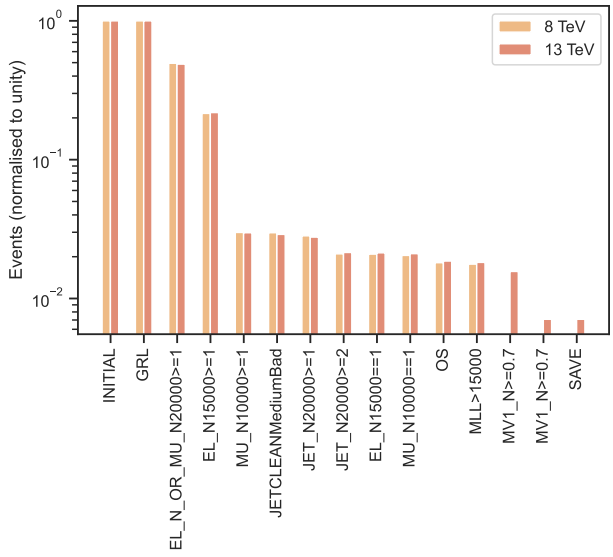
(b) μ+jets.



(c) ee+jets.



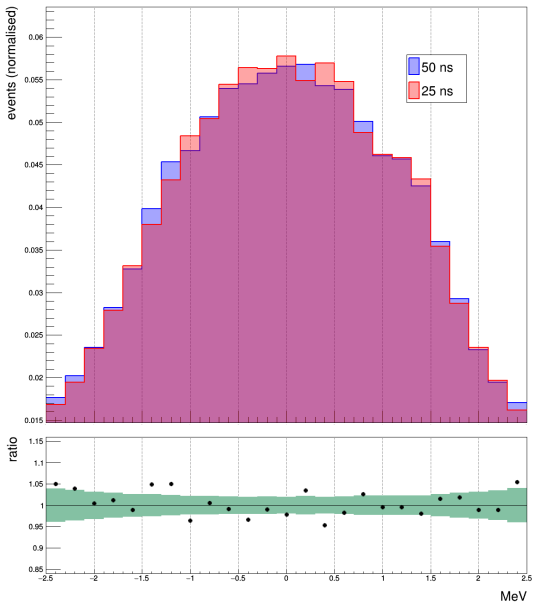
(d) $\mu\mu$+jets.



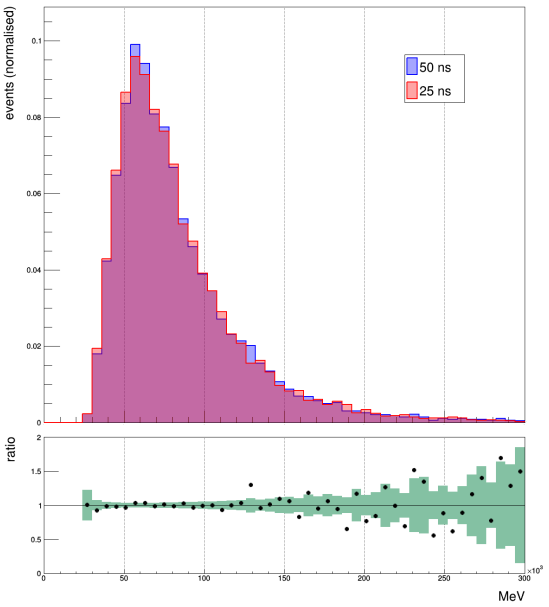
(e) $e\mu$+jets.

Figure 6.9: Cut-flow checks for TTHbbLeptonic with AnalysisTop 2.1.31 for semileptonic and dileptonic channels with 8 TeV and 13 TeV POWHEG-Pythia samples.

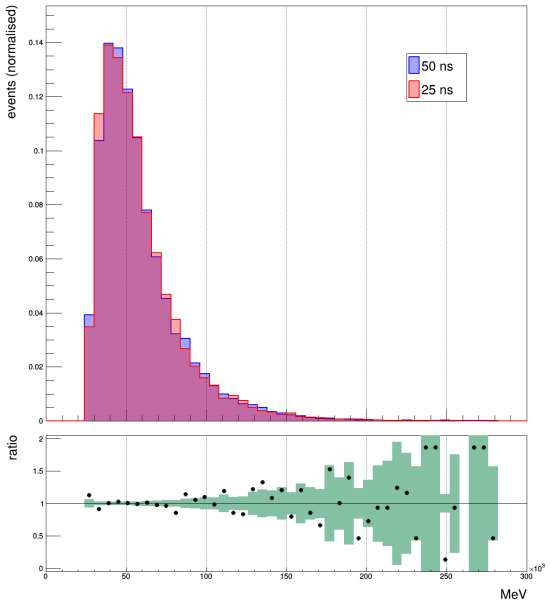
Changing from a bunch spacing of 50 ns to a bunch spacing of 25 ns doubled the number of proton bunches in the LHC to approximately 2,800. Another validation of TTHbbLeptonic involved the comparison, for the purposes of consistency, of a 13 TeV Monte Carlo 2015 (MC15) sample for a bunch spacing of 50 ns with an 13 TeV MC15 sample for a bunch spacing of 25 ns, as shown in Figure 6.10.



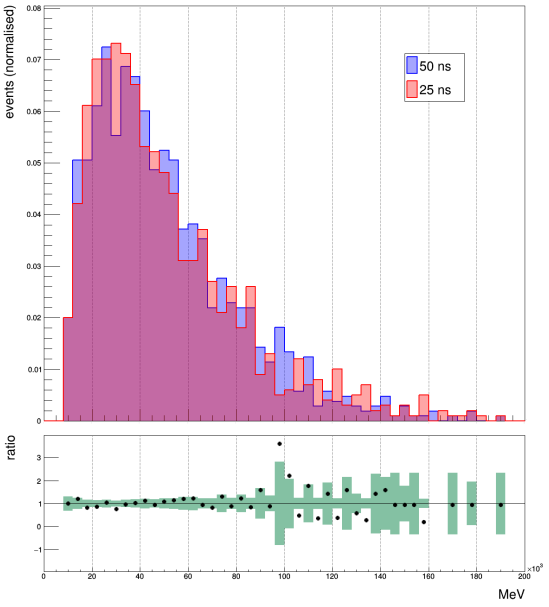
(a) Jet η from an electron selection.



(b) Jet 2 p_{T} from an electron selection.



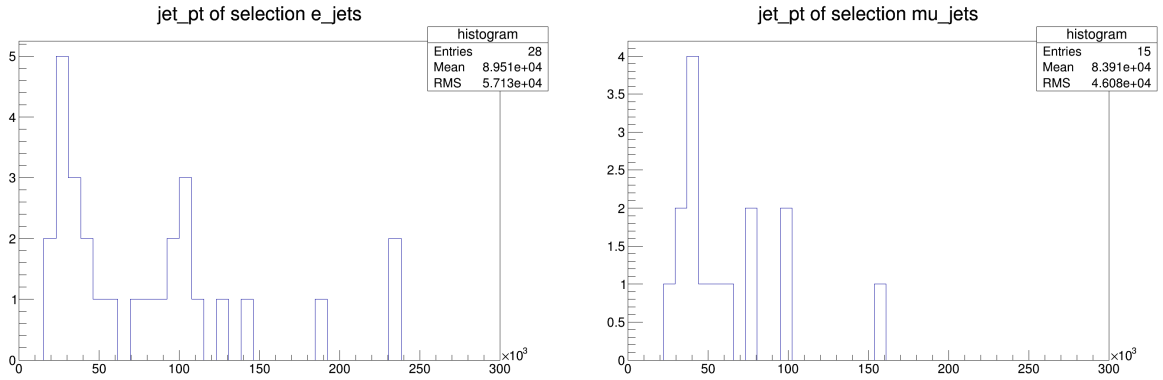
(c) Jet 3 p_{T} from an electron selection.



(d) Muon p_{T} from a muon muon selection.

Figure 6.10: Comparison of a 13 TeV MC15 sample with 50 ns with a 13 TeV MC15 sample with 25 ns.

As described, the TTHbbLeptonic framework was developed out of the TTHSemileptonic framework, also created by the author. The earliest versions of TTHbbLeptonic featured event selector, object loader, object definition and tool loader facilities. Having been developed and validated extensively beyond these early features, TTHbbLeptonic was ready for use at the very start of LHC Run-2. Some early plots by the author of jet p_{T} derived from the first data processed from Run-2 are shown in Figure 6.11.



(a) Jet p_{T} of single electron selection.

(b) Jet p_{T} of single muon selection.

Figure 6.11: Early (earliest?) (2015-06-10) analysis plots of jet p_{T} in LHC Run-2.

6.5 Development of analysis framework and checklists

TTHbbLeptonic was developed by a team of very active contributors. For a video animation of the formative development of TTHbbLeptonic through 2015, please see Reference [31]. This video is a visualisation of the repository commits of the development team. The extensive contributions of the author are represented in the video as a red icon labelled 'wbreaden'. The root directory is shown at centre, directories are represented by line branches and files are represented by circle leaves. Brief projections from user icons to files and directories represent repository commit actions.

The Rivet ethos suggests that "user analysis code should be extremely clean and easy to write – ideally it should be sufficiently self-explanatory to in itself be a reference to the experimental analysis algorithm – without sacrificing power or extensibility" [264]. The TTHbbLeptonic analysis framework was intended to follow such a form for the purposes of ease of use and comprehension, and for the purpose of being accessible to users and to various analyses. By dictat, the ROOT code base master branch trunk was intended to always work and the author followed this approach when launching development of TTHbbLeptonic. Contributors and joining analyses groups were encouraged to develop in branches

with "experts" acting as gatekeepers for merging such developments with the trunk master branch. The trunk master branch was treated with commit procedures, checklists and validations in order to attempt to ensure it was always bug-free and functional to such a degree that TTHbbLeptonic could always be released from this master branch. There were coding styles defined and scripts developed to attempt to clearly and automatically modify written code in order to match a defined coding style. With reference to *The ATLAS Data Management Software Engineering Process*, robust checklists were developed for validation of the code and configurations and documentation, before repository commits, tag releases, production runs *etc.* [278].

The TTHbbLeptonic development team coordinated projects via JIRA and met approximately weekly at a meeting with conversation emphasised over presentations, and did not involve slide presentations. The team presented, on personal rotation, to the main HTop group approximately bi-weekly or when appropriate (such as following a significant release). The development team responded briskly to issues raised via dedicated communications channels and there were sub-groups of people in the team dedicated to specific tasks, such as production runs. The checklists featured greater and lesser degrees of automation depending on the application.

As an example, a simplified checklist for a tag release of TTHbbLeptonic was as follows:

- Ensure a Good Runs List (GRL) is specified in the cuts configurations. (A GRL is used as a mask to use only data which was taken at a time when all detector functions were working correctly.)
- Ensure that the GRL and lumicalc (a tool to extract the luminosity associated with a particular GRL) are current in the cuts configurations.
- Ensure TTHbbLeptonic compiles.
- Ensure all unit tests run successfully.
- Run validation scripts. Ensure comparison plots are identical to the previous release or changed in expected ways.
- Recreate the Python virtual environment as appropriate and commit it.
- Update the Doxygen documentation.
- If necessary, update details on cuts configuration files in the documentation.
- If necessary, update details on included samples lists in the documentation.
- If necessary, include details on new output variables in the documentation.
- Update the documentation otherwise as appropriate.

- Create a new CERN Go link for post-release notes and add this link to the documentation.
- Define a unique name for the release.
- Add a summary of details on the new tag release to the documentation.
- Compile the documentation.
- Commit and then release the tag and send an e-mail to the HTop group mailing list.

6.6 TRExFitter Run-2 statistical inference framework

The statistical approach to a search is outlined in Section 5.2. TRExFitter is a framework used in many ATLAS analyses for statistical inference via profile likelihood fits [279]. It evolved from Run-1 scripts, to the TTHFitter framework, to its current form as the more general TRExFitter, "Top Related Experiment Fitter: for powerful and dangerous fitting". In the profile likelihood method, there is a simultaneous fit of parameters of interest and nuisance parameters encoding the effects of systematic uncertainties. It relies on a global likelihood function including all bins and all parameters and the fit minimises the global negative log-likelihood. This enables reduction of systematic uncertainties by effectively performing in-situ calibrations.

For Run-1, there were many different tools used for fitting, limits-setting and various validation checks. TTHFitter was developed initially as a single repository of tools and code for these purposes and for automated histogram, table and final publication plots generation. As the project grew in complexity and support from groups outside of $t\bar{t}H$ efforts, the framework was renamed to TRExFitter.

TRExFitter is designed to handle many components needed for the statistical part of analysis. Inputs of various formats together with a configuration file are provided to TRExFitter and the framework proceeds to produce a RooFit workspace [280]. It performs profile likelihood fits on multiple analysis regions with multiple bins and calculates limits and significance using RooStats [281, 282]. The configuration file parsing is designed to accept details on the general details of the fitting, such as the point of interest (*e.g.* $t\bar{t}H$ signal strength), file paths *etc.*, regions that define the distributions considered in the fit, together with cuts and labels, and samples and systematic uncertainties. The configuration steers the building of histograms and a HistFactory workspace. The workspace is built from histograms, and thousands of these histograms are made – with one histogram per analysis region, per sample, per systematic variation.

TRExFitter ultimately produces plots, such as data over Monte Carlo and yields per region, backgrounds composition, and tables and provides tools to interpret the fitting behaviour, such as best-fit

values of all nuisance parameters, the ability to exclude nuisance parameters or set them to certain values, and expressions of the correlations of fit parameters. Pre-fit plots are generated for each analysis region featuring the total uncertainty of all sources evaluated and visualised, using binning algorithms to bin as appropriate (particularly applicable to multivariate analysis output distributions). Systematic plots are generated to visualise the effects of all systematic variations per analysis region, per sample, per variation. Smoothing algorithms can be applied to the systematic variations in order to attempt to mitigate artificial effects from statistical fluctuations. There is some basic automation that is designed to flag problematic distributions, but more automation would be desirable. The workspace is designed to include all information needed for fitting. Pruning is performed by removing small systematic variations from the fit model. Post-fit plots visualise the fitting results projected onto all analysis regions and can validate post-fit modelling in regions not used in the fit. A sense of the impact on TRExFitter fitting time with the progressive inclusion of systematic uncertainties is shown in Figure 6.12. Following fitting, a variety of capabilities are available, such as CL_s limits, calculation of significance, nuisance parameter ranking by impact and the combination of different fits.

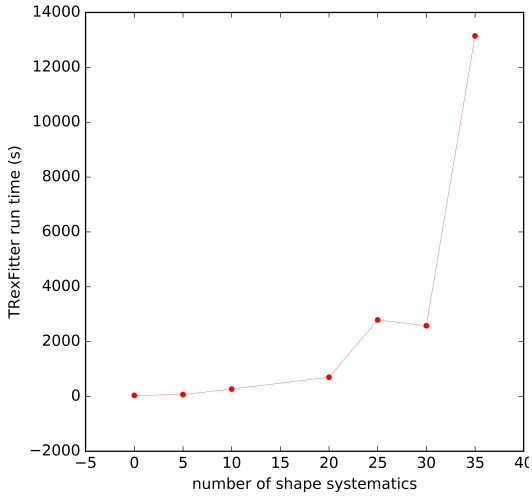


Figure 6.12: TRExFitter run times for various numbers of shape systematic uncertainties included in fitting.

6.7 Outlook on high energy physics software and computing Research and development and on the analysis framework

6.7.1 Research and development for the High-Luminosity LHC

Research, development and exploration of particle physics in coming decades is ambitious, broad and challenging. In planning for the High-Luminosity LHC (HL-LHC) upgrade, which is described in Chapter 10, investment in the research and development of software to acquire, manage, process, and analyse the much larger amounts of data to be recorded is required. A broad plan for the research and development of High Energy Physics (HEP) software and computing infrastructure was completed at the end of 2018 in order to enable collaborative and complimentary efforts to be agreed upon carefully [283].

6.7.2 Future development of the analysis framework

The progression of the framework from TTHSemileptonic to TTHbbLeptonic and the many unions of analyses and tools into TTHbbLeptonic was a gradual, validated generalisation of the framework that increased its effectiveness while reducing the onus on both developers and users of the software. Since the $t\bar{t}H\left(b\bar{b}\right)$ analysis result published in 2018, this generalising advancement has continued with the refactoring and updating of TTHbbLeptonic to form the more general TTHbbAnalysis [33]. While this framework has been validated and has worked well by being an essential part of the analysis that led to evidence of $t\bar{t}H$ (which subsequently led to observation), and is expected to persist for the time being, it is the view of the author that occasionally there should be a substantial rewrite or renewal of infrastructures in order for advancement to continue, in this case to make physics understanding better and to make physics analyses easier [284, 2, 3].

Chapter 7

Results of the $t\bar{t}H\left(b\bar{b}\right)$ search

> The Large Hadron Collider (LHC) is a big round thing that has protons in it and the protons smash together making smaller things and then we look at the smaller things.
>
> Euan Mclean

7.1 Introduction

This chapter details the results of the statistical analysis for the search for $t\bar{t}H\left(b\bar{b}\right)$. Following the application of the TTHbbLeptonic framework and associated software, the statistical analysis to maximise the profile likelihood ratio, corresponding to minimising the test statistic, was completed with the application of the TRExFitter infrastructure. Specifically, the likelihood functions for the model described in Section 5.4 together with the uncertainties described in Section 5.7 were defined within HistFactory, with histograms that were created by TRExFitter specifying distributions relevant to definition of the likelihood functions [285]. Minimisation of the test statistic was performed by Minuit2 and uncertainties for unconstrained parameters were determined using the Minos algorithm [286, 287]. The profile likelihood test statistic was constructed as described in Section 5.2. Systematic uncertainties were implemented as described in Sections 5.2 and 5.7. Upper limits were calculated with the modified frequentist confidence level CL_s and the result of the search for $t\bar{t}H\left(b\bar{b}\right)$ was quantified by the signal strength parameter $\mu_{t\bar{t}H\left(b\bar{b}\right)}$, which is the ratio of the cross-section observed of $t\bar{t}H\left(b\bar{b}\right)$ production normalised to the cross-section of the Standard Model theoretical expectation, $\sigma_{\mathrm{obs.}}/\sigma_{\mathrm{SM}}$ – a measurement of the probability that data observed is compatible with the background-only hypothesis. The uncertainty of the best-fit value of the signal strength parameter $\mu_{t\bar{t}H\left(b\bar{b}\right)}$ was estimated by varying the test statistic by one

unit.

7.2 Fit model

The fit model is described by the variables selected to build the template distributions, together with the list of systematic uncertainties and their correlations across the 19 analysis regions. The BDT outputs of the signal analysis regions, described in Section 5.6, are combined in the profile likelihood fit in order to test for the presence of signal $t\bar{t}H\left(b\bar{b}\right)$ while simultaneously determining the normalisation and constraints for the differential distributions of the dominant background components. In the control regions, the event yields are used, with the exceptions of $\mathrm{CR}^{5\mathrm{j}}_{t\bar{t}+\geq 1c}$ and $\mathrm{CR}^{\geq 6\mathrm{j}}_{t\bar{t}+\geq 1c}$, for which $H_{\mathrm{T}}^{\mathrm{had.}}$ is used. The discriminants used for the 19 analysis regions of the semileptonic and dileptonic channels are listed in Table 7.1.

Channel	Analysis region	Discriminant
Semileptonic	$\mathrm{SR}^{\geq 6\mathrm{j}}_{1}$	BDT
Semileptonic	$\mathrm{SR}^{\geq 6\mathrm{j}}_{2}$	BDT
Semileptonic	$\mathrm{SR}^{\geq 6\mathrm{j}}_{3}$	BDT
Semileptonic	$\mathrm{SR}^{5\mathrm{j}}_{1}$	BDT
Semileptonic	$\mathrm{SR}^{5\mathrm{j}}_{2}$	BDT
Semileptonic	$\mathrm{SR}^{\mathrm{boosted}}$	BDT
Semileptonic	$\mathrm{CR}^{\geq 6\mathrm{j}}_{t\bar{t}+\mathrm{light}}$	BDT
Semileptonic	$\mathrm{CR}^{\geq 6\mathrm{j}}_{t\bar{t}+\geq 1c}$	$H_{\mathrm{T}}^{\mathrm{had.}}$
Semileptonic	$\mathrm{CR}^{\geq 6\mathrm{j}}_{t\bar{t}+b}$	event yield
Semileptonic	$\mathrm{CR}^{5\mathrm{j}}_{t\bar{t}+\mathrm{light}}$	event yield
Semileptonic	$\mathrm{CR}^{5\mathrm{j}}_{t\bar{t}+\geq 1c}$	$H_{\mathrm{T}}^{\mathrm{had.}}$
Semileptonic	$\mathrm{CR}^{5\mathrm{j}}_{t\bar{t}+b}$	event yield
Dileptonic	$\mathrm{SR}^{\geq 4\mathrm{j}}_{1}$	BDT
Dileptonic	$\mathrm{SR}^{\geq 4\mathrm{j}}_{2}$	BDT
Dileptonic	$\mathrm{SR}^{\geq 4\mathrm{j}}_{3}$	BDT
Dileptonic	$\mathrm{CR}^{\geq 4\mathrm{j}}_{t\bar{t}+\mathrm{light}}$	event yield
Dileptonic	$\mathrm{CR}^{\geq 4\mathrm{j}}_{t\bar{t}+\geq 1c}$	event yield
Dileptonic	$\mathrm{CR}^{3\mathrm{j}}_{t\bar{t}+\mathrm{light}}$	event yield
Dileptonic	$\mathrm{CR}^{\geq 3\mathrm{j}}_{t\bar{t}+\geq 1b}$	event yield

Table 7.1: Analysis regions of each channel with their respective discriminant.

The distributions of the classification BDT output in the resolved semileptonic channel signal regions, both before the fit to data and after, are shown in Figure 7.1. The distributions are well-modelled pre-fit within uncertainties. No distinction is made in the fit between the signal and control regions, other than by a different selection of discriminant variables. The degree of agreement is improved post-fit through the adjustment of nuisance parameters in the fit, with the best-fit normalisation factors of $k\,(t\bar{t}+\geq 1b)$ and $k\,(t\bar{t}+\geq 1c)$ being 1.24 ± 0.10 and 1.63 ± 0.23 respectively. Uncertainties on these normalisation factors do not include the theory uncertainty of the corresponding $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ cross-sections.

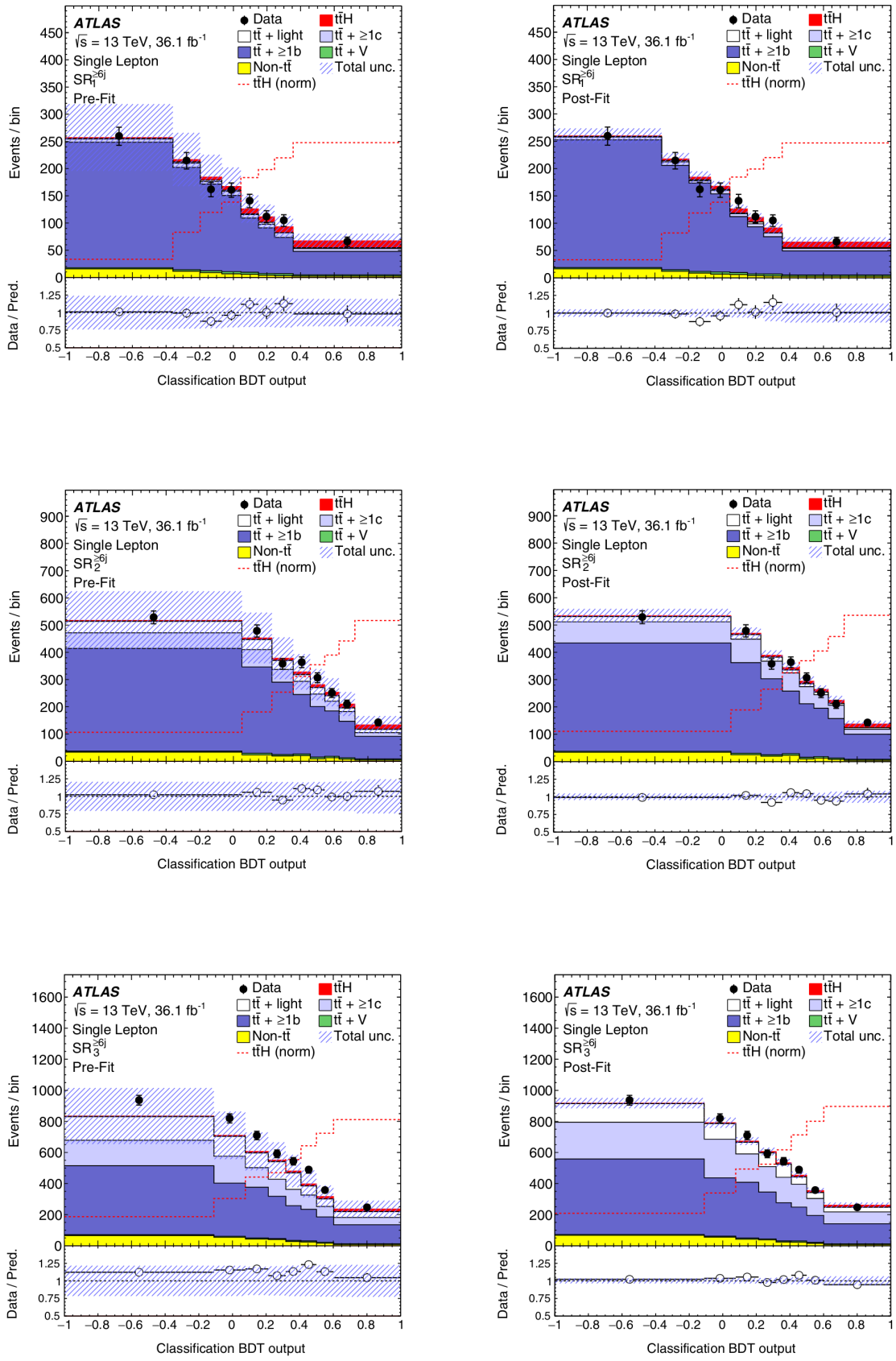


Figure 7.1: Comparison of data and prediction for BDT outputs, pre-fit (on the left) and post-fit (on the right), for the semileptonic channel signal regions of events with at least six jets [1]. The signal $t\bar{t}H$ $(b\bar{b})$ distribution is normalised to the total background prediction and shown in red. Pre-fit plots do not feature uncertainties for the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ normalisation factors as these are free-floating parameters in the fit.

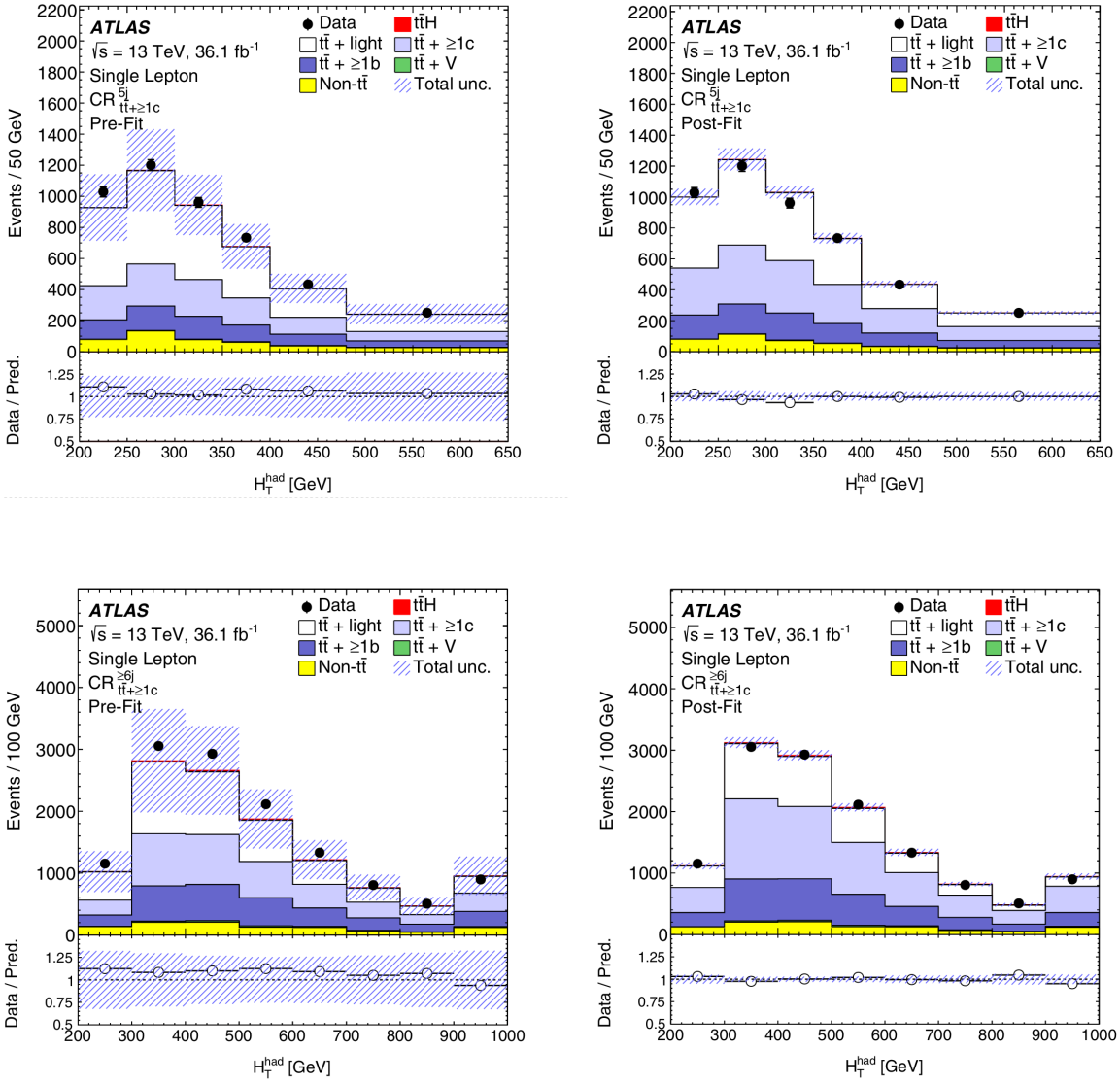


Figure 7.2: Comparison of data and prediction for $H_{\mathrm{T}}^{\mathrm{had.}}$, pre-fit (on the left) and post-fit (on the right), for semileptonic channel control regions [1]. The signal $t\bar{t}H$ $(b\bar{b})$ distribution (a relatively small contribution in control regions) is normalised to the total background prediction and shown in red. Pre-fit plots do not feature uncertainties for the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ normalisation factors as these are free-floating parameters in the fit.

The binning of distributions used for fitting is optimised for sensitivity while constraining statistical uncertainties associated with the model to less than 20 % in order to avoid a possible bias arising from statistical fluctuations in the model. Background-dominated regions are used to improve the knowledge of the background through constraints of the nuisance parameters or resulting correlations. In control regions in which the $H_{\mathrm{T}}^{\mathrm{had.}}$ distribution shape is not well-modelled, a single bin is used in the fit to data in order to avoid propagating mismodelled effects from control regions to signal regions, and also to avoid arbitrary pulls of the nuisance parameters in order to correct such mismodelling.

Signal strength is a free-floating parameter in the fit, with no prior, together with the normalisation factors of the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ backgrounds. It was found that Monte Carlo simulations tend to underestimate $t\bar{t}+\mathrm{HF}$ fractions relative to data observed, thus the normalisation of $t\bar{t}+\mathrm{HF}$ components are free-floating in the fit with no prior. Other backgrounds have their normalisation controlled through nuisance parameters that correspond to the theoretical knowledge of the respective cross-sections or uncertainties on methods of estimation, such as is the case for the fake lepton background estimated using the matrix method. The Asimov fit is defined such that nuisance parameters corresponding to systematic uncertainties are centred at zero and the normalisation factor is centred at unity.

The performance of the analysis is calculated using the representative Asimov dataset approach described in Section 5.2. The Asimov fit for the semileptonic channel results in a signal strength of

$$\begin{aligned}\mu_{\text{Asimov data}} &= 1.00 \pm 0.32\,(\text{stat.})^{+0.60}_{-0.57}\,(\text{syst.}) \\ &= 1.00^{+0.68}_{-0.65},\end{aligned} \tag{7.1}$$

which corresponds to an expected significance for the $t\bar{t}H\left(b\bar{b}\right)$ signal of 1.5σ assuming the expected Standard Model $t\bar{t}H\left(b\bar{b}\right)$ production. Uncertainty on the signal strength is dominated by systematic uncertainties. Here the statistical uncertainty includes the uncertainty on the normalisation factors of $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$.

The known systematic uncertainties that most impact the uncertainty on the signal strength are shown in Figure 7.3. The sensitivity of the search for $t\bar{t}H\left(b\bar{b}\right)$ is limited by the modelling of the dominant $t\bar{t}+\geq 1b$ background, which results in an uncertainty on the signal strength of $^{+0.49}_{-0.48}$, in addition to an uncertainty of $^{+0.12}_{-0.14}$ from $t\bar{t}+\geq 1b$ normalisation. Background Monte Carlo statistics and the fake lepton estimate result in an uncertainty on the signal strength of $^{+0.29}_{-0.31}$. This could be reduced by increasing the number of events generated in the phase space where the signal $t\bar{t}H\left(b\bar{b}\right)$ is present. The modelling of $t\bar{t}H$ results in an uncertainty on the signal strength of $^{+0.24}_{-0.03}$. The four most highly ranked nuisance parameters all are related to the $t\bar{t}+\geq 1b$ background model and are constrained to 0.5σ. The highest ranking nuisance parameter arises from differences between the events generated by the generator for the nominal sample

and by a generator for an equivalent sample. This highest ranking nuisance parameter corresponds to a post-fit uncertainty of $^{+0.45}_{-0.43}$ on the signal strength and is constrained to 0.47σ. As shown in Figure 7.3, some nuisance parameters are shifted in the fit from their nominal values. The shifts in the nuisance parameters were found mainly to correct the theoretical predictions of the $t\bar{t}$ background to data observed. The variations in signal strength corresponding to the shifts are quantified by fixing nuisance parameters to their pre-fit values, repeating the fit, and then comparing the resultant signal strength value with the value from the nominal fit. The variations were found to be smaller than the uncertainty in the signal strength. Further, independent signal strength values resulting from fits comprised of different sets of analysis regions and from the semileptonic and dileptonic channels were found to be compatible also.

Theoretical predictions for the $t\bar{t}+ \geq 1b$ background are impacted by large uncertainties, as reflected in the differences between alternative simulated samples used to model this background. The corresponding systematic uncertainties are large and a crucial limiting factor for this search. The choice of nuisance parameters corresponding to these systematic uncertainties was studied extensively in order to attempt to ensure flexibility in the fit that is sufficient to correct for mis-modelling of this background and to avoid any bias in the measured signal strength. 13 independent nuisance parameters are used to describe the $t\bar{t}+ \geq 1b$ background modelling uncertainties. The capability of the fit to correct for mis-modelling (beyond that apparent in fit input distributions) was assessed by comparing the predictions of all input variables to the classification BDT obtained post-fit to data. No significant deviations of the predictions from data were found, and the agreement is improved post-fit. As an additional validation check on the robustness of the fit, a pseudo-dataset was built from simulated events by replacing the nominal $t\bar{t}$ background with an alternative sample not used in the definition of any uncertainties in the analysis. This was a POWHEG + Pythia 6 sample similar to the sample used previously in the $t\bar{t}H\left(b\bar{b}\right)$ analysis for Run-1 [93]. The fit with this substitution did not reveal any bias in signal extraction.

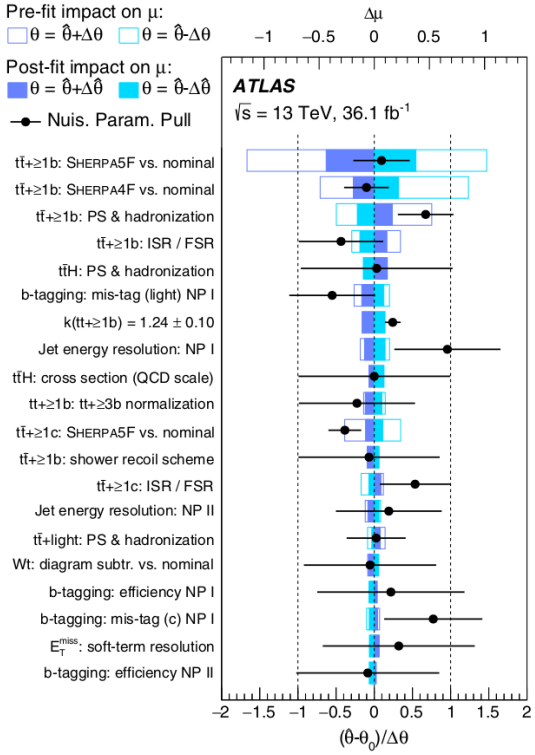


Figure 7.3: Nuisance parameters ranked by their impact on the observed signal strength [1]. Nuisance parameters representing Monte Carlo statistical uncertainties are not included. Empty blue rectangles represent the pre-fit upper impact on the signal strength while filled turquoise rectangles represent the post-fit upper impact on the signal strength. Black points represent the pulls of the nuisance parameters relative to their nominal values θ_0. The impact of each nuisance parameter $\Delta\mu$ is calculated by comparing the nominal best-fit value of the signal strength with the result of the fit when the nuisance parameter under consideration is fixed at its best-fit value $\hat{\theta}$, shifted by its pre-fit uncertainties $\pm\Delta\theta$ and its post-fit uncertainties $\pm\Delta\hat{\theta}$. Experimental uncertainties that are decomposed into several independent sources are represented by NP 1 and NP 2, which correspond to the first and second nuisance parameters, ranked by their impact on the observed signal strength.

Contributions to uncertainties in signal strength $\mu_{t\bar{t}H(b\bar{b})}$ are shown in Figure 7.2.

Source of uncertainty	$\Delta\mu_{t\bar{t}H(b\bar{b})}$	
$t\bar{t}+\geq 1b$ modelling	+0.46	-0.46
Background model statistical uncertainty	+0.29	-0.31
b-tagging efficiency and mis-tag rates	+0.16	-0.16
JES and JER	+0.14	-0.14
$t\bar{t}H$ modelling	+0.22	-0.05
$t\bar{t}+\geq 1c$ modelling	+0.09	-0.11
JVT and pile-up modelling	+0.03	-0.05
Other background modelling	+0.08	-0.08
$t\bar{t}+$ light modelling	+0.06	-0.03
Luminosity	+0.03	-0.02
Light lepton (e, μ) identification, isolation, triggering	+0.03	-0.04
Total systematic uncertainty	+0.57	-0.54
$t\bar{t}+\geq 1b$	+0.09	-0.10
$t\bar{t}+\geq 1c$	+0.02	-0.03
Intrinsic statistical uncertainty	+0.21	-0.20
Total statistical uncertainty	+0.29	-0.29
Total uncertainty	+0.64	-0.61

Table 7.2: Contributions to uncertainties in signal strength $\mu_{t\bar{t}H(b\bar{b})}$, calculated post-fit [1]. Background model statistical uncertainty refers to statistical uncertainties in Monte Carlo events and in data-driven determination of non-prompt and fake lepton background components in the semileptonic channel. The total statistical uncertainty is calculated after fixing all nuisance parameters in the fit except for the free-floating normalisation factors for the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ background components. Uncertainties in the normalisation factors of $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ are included in the total statistical uncertainty rather than in the total systematic uncertainty component. The statistical uncertainty evaluated after fixing the normalisation factors of $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ and is then quoted as an "intrinsic statistical uncertainty". Other uncertainties are calculated by repeating the fit after fixing a subset of nuisance parameters and subtracting in quadrature the resulting total uncertainty in signal strength from the uncertainty of the full fit. The same approach is used to calculate the separate effects of the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ normalisations. The total uncertainty is distinct from the sum in quadrature of the various components due to correlations between nuisance parameters in the fit.

7.3 Results

An excess of events over the theoretically expected background was found with an observed significance of 1.4σ, compared to an expected significance of 1.6σ. A signal strength larger than 2.0 was excluded at the 95 % confidence level. The expected significance and exclusion limits were calculated using the post-fit background estimate and are shown in Figure 7.4.

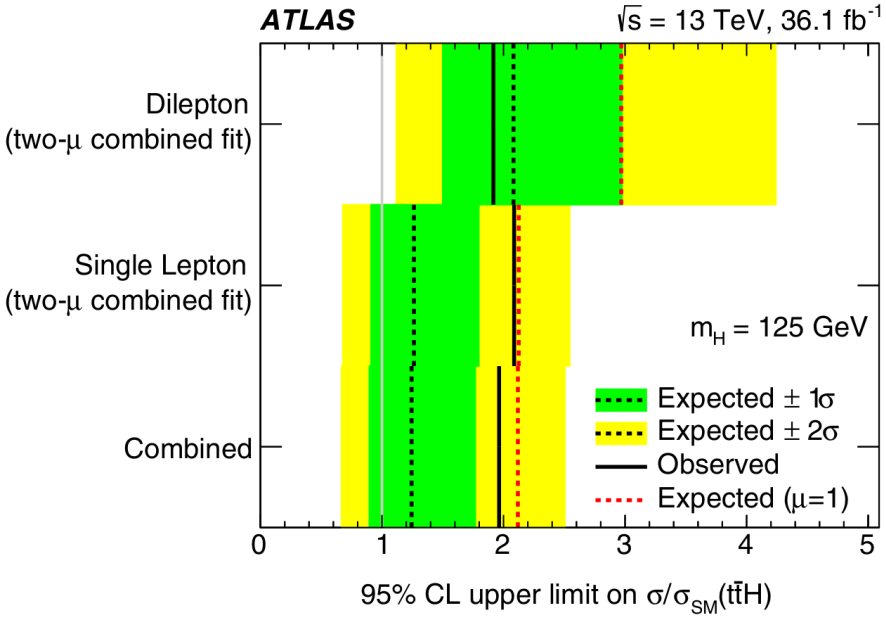


Figure 7.4: Brazil-band plot of the 95 % confidence level on the $t\bar{t}H$ $(b\bar{b})$ cross-section relative to the Standard Model prediction in the semileptonic and dileptonic channels and their combination. The observed limits are shown (in solid black lines) together with expected limits corresponding to both the background-only hypothesis (in dotted black lines) and the Standard Model hypothesis (in dotted red lines). Uncertainty bands corresponding to 1σ and 2σ on the background-only hypothesis are shown [1].

The data event yields are compared to the post-fit theoretical predictions for all events entering the analysis selection and are shown, grouped and ordered by the signal-to-background ratios of the final discriminant bins, in Figure 7.5. Theoretical predictions are shown for both the background-only hypothesis and the signal-plus-background hypothesis, where the signal is scaled independently to the observed signal strength for the upper limit on the signal strength.

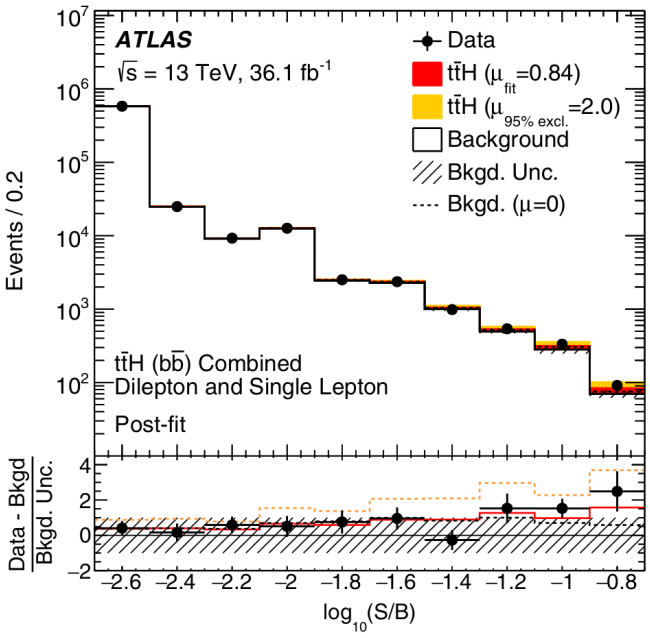


Figure 7.5: Post-fit data event yields of signal S and background B as a function of $\log_{10}\frac{S}{B}$ compared with post-fit theoretical predictions [1]. Final discriminant bins (with the first bin including underflow) in all semileptonic and dileptonic analysis regions are combined into $\log_{10}\frac{S}{B}$ bins, with signal normalised to the Standard Model theoretical prediction used for the calculation of $\log_{10}\frac{S}{B}$. The signal $t\bar{t}H\left(b\bar{b}\right)$ is shown normalised to the best-fit value $\mu = 0.84$ and to the best-fit value excluded at the 95 % confidence level $\mu = 2.0$, in both cases summed to the post-fit theoretical background prediction. The lower pane shows the pull (residual divided by its uncertainty) of each bin of the data observed relative to the post-fit theoretical background prediction. These pulls for the data observed are compared with the pulls of the post-fit signal-plus-background theoretical prediction, assuming a signal strength equal to the best-fit value (shown as a solid red line) and equal to the exclusion limit (shown as a dashed orange line). The post-fit background and its pulls under the assumption of the background-only hypothesis are shown also in the upper pane (shown as a dashed black line mostly obscured). Evidence of a $t\bar{t}H\left(b\bar{b}\right)$ signal-like excess over background is visible at the higher values of $\log_{10}\frac{S}{B}$.

The combined best fit value for the signal strength of $t\bar{t}H\left(b\bar{b}\right)$, assuming a Standard Model Higgs boson of 125 GeV, was found to be

$$\begin{aligned}\mu_{t\bar{t}H(b\bar{b})}\left(m_H = 125\right) &= 0.84 \pm 0.29\,(\text{stat.})^{+0.57}_{-0.54}\,(\text{syst.}) \\ &= 0.84^{+0.64}_{-0.61},\end{aligned} \tag{7.2}$$

a value obtained from the combined fit in all signal and background analysis regions in the semileptonic and dileptonic channels, with the large uncertainty found to arise largely from the uncertainty on Monte Carlo modelling of the $t\bar{t}b\bar{b}$ background.

An alternative combined fit was performed, for which the semileptonic and dileptonic channels were assigned two independent signal strengths. The corresponding best fit values of the signal strength of $t\bar{t}H\left(b\bar{b}\right)$ were found to be $0.95^{+0.65}_{-0.62}$ in the semileptonic channel and $-0.24^{+1.02}_{-1.05}$ in the dileptonic channel. The combined signal strength and the two independent signal strengths are shown with their statistical and systematic uncertainties in Figure 7.6.

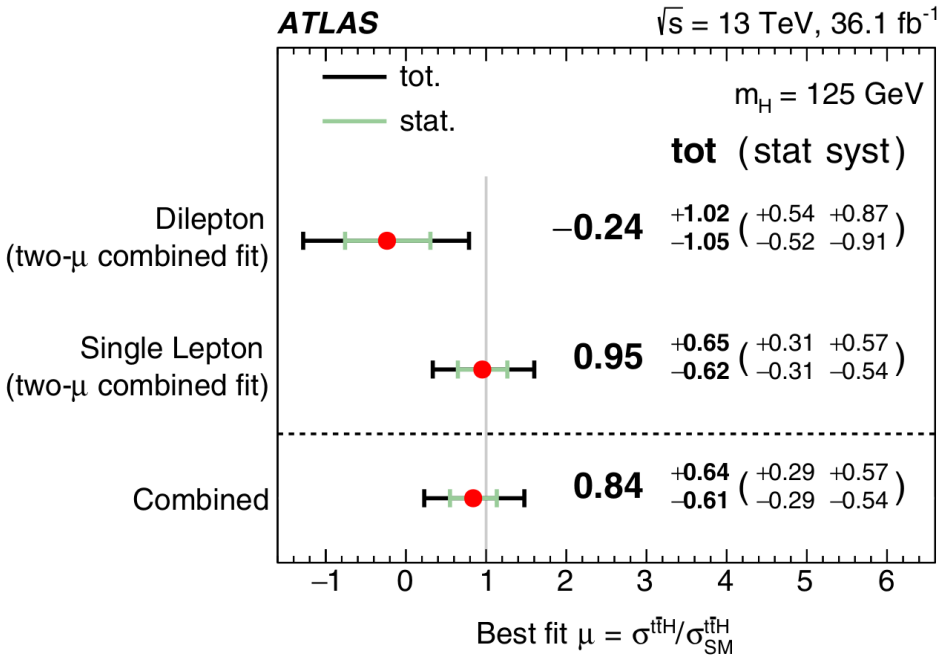


Figure 7.6: The combined signal strength of $t\bar{t}H$ together with the signal strengths from the two channels together with their respective statistical and systematic uncertainties [1].

The data observed are consistent with both the background-only hypothesis and with the theoretical

Standard Model $t\bar{t}H\left(b\bar{b}\right)$ prediction. An excess of 1.6σ is predicted in the presence of a Standard Model Higgs boson and an excess of 1.4σ above the expected background is observed.

The results of this search were approved and published by the ATLAS Collaboration in 2018 [1, 32]. The $t\bar{t}H$ process is rare and so it took a number of years of recording data to be able to declare evidence of it. The search was made possible through the functioning of the LHC and ATLAS and through the efforts of the analysis teams who devised and validated the analysis and who came up with methods to squeeze every last bit of information from data recorded.

Following this finding, work continued beyond the $t\bar{t}H\left(b\bar{b}\right)$ analysis (and the direct work of this thesis) and resulted in observation of $t\bar{t}H$ using data corresponding to an integrated luminosity of $79.8\,\text{fb}^{-1}$ and a combination of results from Higgs boson decays to $b\bar{b}$, WW^*, $\tau^+\tau^-$, $\gamma\gamma$ and ZZ^* with an observed significance of 5.8σ, compared to an expected significance of 4.9σ [2, 3]. Combined with $t\bar{t}H$ search results using data recorded at centre-of-mass energies of $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$, the observed significance was 6.3σ, compared to an expected significance of 5.1σ. Assuming Standard Model branching fractions, the $t\bar{t}H$ production cross-section at $\sqrt{s} = 13\,\text{TeV}$ was measured as $670 \pm 90\,(\text{stat.})^{+110}_{-100}\,(\text{syst.})\,\text{fb}$, which is in agreement with the Standard Model theoretical prediction of $507^{+35}_{-50}\,\text{fb}$.

The significances of the individual and combined analyses relative to the background-only hypothesis are shown together with the corresponding observed $t\bar{t}H$ cross-sections in Table 7.3 while their ratios to the Standard Model theoretical predictions are shown in Figure 7.7. The $t\bar{t}H$ production cross-sections measured in pp collisions at centre-of-mass energies of 8 and 13 TeV and compared with Standard Model theoretical predictions are shown in Figure 7.8.

Analysis	Integrated luminosity (fb^{-1})	$t\bar{t}H$ cross-section (fb)	Observed significance	Expected significance
$H \to \gamma\gamma$	79.8	$710^{+210}_{-190}\,(\text{stat.})^{+120}_{-90}\,(\text{syst.})$	4.1σ	3.7σ
$G \to$ multilepton	36.1	$790 \pm 150\,(\text{stat.})^{+150}_{-140}\,(\text{syst.})$	4.1σ	2.8σ
$H \to b\bar{b}$	36.1	$400^{150}_{-140}\,(\text{stat.}) \pm 270\,(\text{syst.})$	1.4σ	1.6σ
$H \to Z^* \to 4l$	79.8	900 (68 % CL)	0σ	1.2σ
Combined (13 TeV)	36.1–79.8	$670 \pm 90\,(\text{stat.})^{+110}_{-100}\,(\text{syst.})$	5.8σ	4.9σ
Combined (7 TeV, 8 TeV, 13 TeV)	4.5, 20.3, 36.1–79.8	-	6.3σ	5.1σ

Table 7.3: $t\bar{t}H$ production cross-sections, and observed and expected significances relative to the background-only hypothesis [2]. Results of individual analyses and combinations are shown. As no events in $H \to Z^* \to 4l$ were observed, an observed upper limit is set at the 68 % confidence level on the $t\bar{t}H$ production cross-section in that channel using pseudo-experiments.

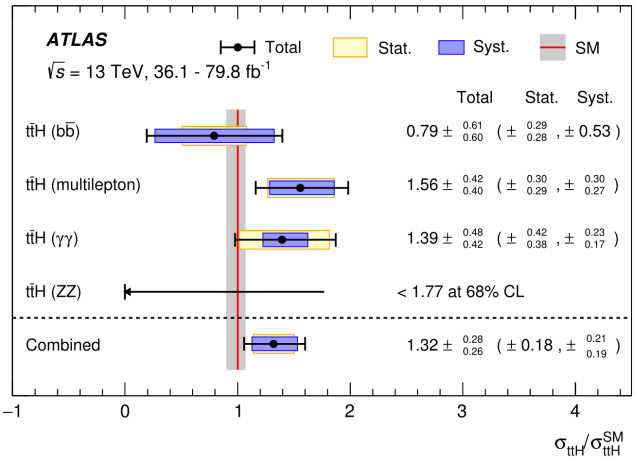


Figure 7.7: Observed $t\bar{t}H$ cross-sections in pp collisions for various individual analyses and combination divided by the Standard Model expectation [2]. Black uncertainty lines represent the total uncertainties, blue bands represent the statistical uncertainties and yellow bands represent the systematic uncertainties. The vertical red line at unity represents agreement with the Standard Model expectation and the vertical grey band represents PDF + α_s uncertainties and uncertainties due to missing higher-order corrections.

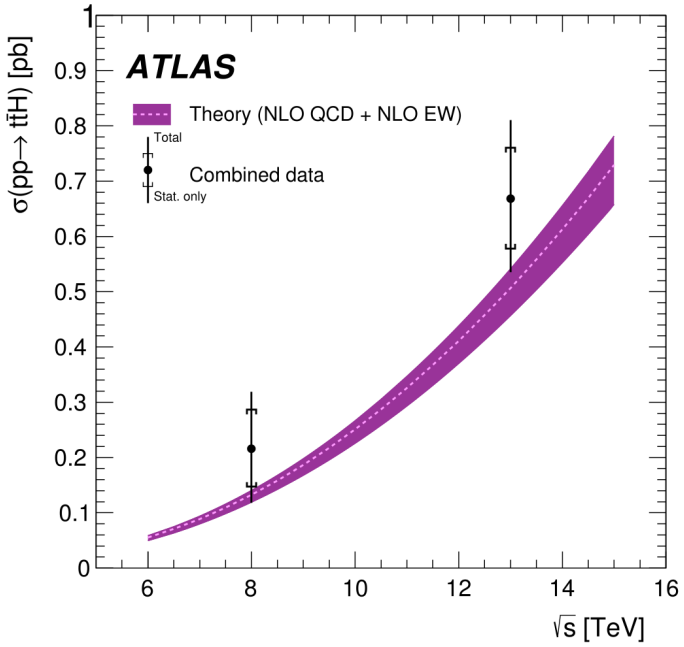


Figure 7.8: Measured $t\bar{t}H$ cross-sections in pp collisions at various centre-of-mass energies compared with theoretical expectations [2]. The measurements from data recorded are shown with both total and statistical-only uncertainties. The band around the Standard Model predictions represents PDF $+ \alpha_s$ uncertainties and uncertainties due to missing higher-order corrections beyond NLO.

7.4 Status of $t\bar{t}H\left(b\bar{b}\right)$ in 2020

Beyond the findings and the work on both the $t\bar{t}H\left(b\bar{b}\right)$ analysis and its infrastructure that are described in this thesis, the status of the $t\bar{t}H\left(b\bar{b}\right)$ analysis in 2020 corresponds to an integrated luminosity of $139\,\text{fb}^{-1}$ and is published in Reference [258]. This analysis is a full Run-2 measurement with $t\bar{t}H\left(b\bar{b}\right)$ events with one or two electrons or muons. It is a similar analysis to the $t\bar{t}H\left(b\bar{b}\right)$ analysis of 2018, though the 2020 analysis has an updated strategy targeting the simplified template cross-sections (STXS) formalism [92]. It features three channels with tightened selections which are dominated by $t\bar{t}+ \geq b$ with negligible $t\bar{t}+$ light, the semileptonic channel (corresponding to 5 STXS bins), the dileptonic channel (corresponding to 4 STXS bins) and the boosted channel (corresponding to 2 STXS bins). The analysis now features a deep neural network for classification of large-R jets as Higgs, t or QCD in $t\bar{t}H$, the output of which is an input to the classification BDT for the semileptonic channel [288, 289]. This 2020 status

for the $t\bar{t}H\left(b\bar{b}\right)$ analysis has the expected significance of $t\bar{t}H$ at 3.0σ and the observed significance at 1.3σ [258].

Chapter 8

Dimensionality reduction and multivariate techniques, and comparisons and interpretations of models for classification of $t\bar{t}H\left(b\bar{b}\right)$ signal and $t\bar{t}b\bar{b}$ background

You're standing on a bridge watching yourself go by.

Ram Dass

8.1 Dimensionality reduction and multivariate techniques

The first part of this chapter details a number of multivariate methods, some old and some new, and novel techniques for visualising and interpreting the modelling achieved by these methods. The second part of this chapter details application of these methods to the task of discrimination of the signal $t\bar{t}H\left(b\bar{b}\right)$ from the background $t\bar{t}b\bar{b}$ for the $t\bar{t}H\left(b\bar{b}\right)$ analysis. For a description of the hardware and software used in implementation of the multivariate methods and investigations and in the application of these approaches to the $t\bar{t}H\left(b\bar{b}\right)$ analysis, please refer to Section 3.4.

The review *A high-bias, low-variance introduction to Machine Learning for physicists* provides a detailed overview of machine learning with some theoretical foundations and Chollet's *Deep Learning with Python*

provides a practical guide to implementation of some contemporary deep learning methods [290, 291].

Given the detailed and successful theoretical efforts expended in the definition of physics event variables, a standard approach in high-energy physics to pursue discriminating power between classes of events is to apply different requirements on the variables used for classification, often referred to as cut-based strategies. Given the difficulty of event classification problems such as the signal/background discrimination associated with the SUSY dataset and with $t\bar{t}H\left(b\bar{b}\right)/t\bar{t}b\bar{b}$ event classification, and given that there does not currently exist a full understanding of precisely why some machine learning approaches work well, an experimental and speculative approach was taken to the testing of various approaches, with practical results preferred over conventional wisdom.

In the field of artificial intelligence, modelling using algorithms classed under such domains as Machine Learning (ML) and Deep Learning (DL) has become an area of intense research and activity, with applications showing success in a wide variety of fields, and applications suggesting new fields of investigation. Medical diagnoses, computer vision, natural language processing, and speech processing are examples of fields being advanced substantially by these algorithms. Inspired originally by simple models of networks of neurons observed in brains, these algorithms are simple approximations to the biological processes observed to be correlated with learning and memory encoding. It should be emphasised that much of the functioning of the neural networks of brains is not understood.

Measurements and interpretations of what is referred to as intelligence are difficult, as detailed by Chollet's paper *On the Measure of Intelligence* [292]. The paper notes that the lack of a satisfying answer to what is meant by intelligence is illustrative of the immaturity of the field of artificial intelligence. However, given the relative simplicity of the various multivariate models described in this thesis compared with much more complicated systems like brains, some interpretations of modelling achieved and internal representations are possible. Various novel approaches to interpretation of modelling were defined, explored and evaluated in order to attempt to reduce somewhat the 'black box' character that is often supposed of multivariate methods employed in high-energy physics.

8.1.1 Motivations for multivariate modelling methods, and novel techniques of interpretation in the context of experimental high-energy particle physics

Over the past decades, the use of machine learning methods has grown in high-energy physics for collection and analysis of data. An early introduction to machine learning techniques for the purposes of high-energy physics (including comments on the future LHC) was provided by the 14^{th} CERN School of Computing and is available in the corresponding CERN Yellow Report [293]. Pioneering studies in previous-generation experiments which made use of multivariate methods, such as neural networks and

boosted decision trees, led to machine learning being an essential component of LHC physics [294, 295, 296, 297, 298]. Higgs bosons are relatively rarely produced in the proton-proton collisions of the LHC and they decay in ways that are similar to the decays of much less rarely produced processes, and so machine learning methods were integral in the discovery of the Higgs boson [11, 12].

A core problem in experimental high-energy particle physics experiments is classification of particle interactions observed in detectors as signal and background. High-level objects are reconstructed, such as clusters, tracks and showers, associated with these particle interactions, and expressing the energies, velocities, shapes and other qualities of these objects with a variety of quantities. These quantities may be selected and presented to multivariate algorithms, such as Boosted Decision Trees (BDTs) and Multi-Layer Perceptrons (MLPs) in order to attempt to separate signal from background [299, 300, 301, 302, 303, 304, 229]. While these multivariate algorithms have been successful, flaws in the reconstruction of high-level features from raw detector data can lead to incorrect categorisation of physics events, the models can be limited in function and interpretability, and, to date, the features presented to the multivariate algorithms have consisted mostly of those only which have been previously imagined and implemented.

An essential component of physics searches in LHC experiments is event classification, such as distinguishing signal from background events. Discovering new particles in collisions of contemporary high-energy particle colliders can require solving difficult problems of classification of signal and background events, hence machine learning methods are often applied. More traditional models used have been 'shallow' and suited to computation performed on CPUs. These shallow models have a relatively limited capacity to learn complex non-linear functions of data (something significant for the task of distinguishing $t\bar{t}H$ $\left(b\bar{b}\right)$ signal from $t\bar{t}b\bar{b}$ background) and can require development of constructed non-linear features from more basic features in data. That is not so much to suggest that such shallow models cannot model extreme complexity, but it is more to suggest that the process of calibrating their variables takes more time than it would with models of other, more complicated or more parameterised structures. Advancements in the field of deep learning, both in terms of computational hardware and development of new algorithms, have made it possible to learn more complex functions to map inputs to outputs and so to better distinguish signal from background. With the constraint of limited quantities of data, every improvement in analytical tools is important. Increasing the efficiency of classification in turn increases particle discovery potential, and any additional modelling power from contemporary multivariate methods is valuable given the expense of high-energy physics apparatuses.

For discovery of a new phenomenon, selections of data are made in which the hypothesis of the new phenomenon is significantly different to the prediction of the null hypothesis, allowing for an effective statistical test. The ratio of the sample likelihood functions of the two hypotheses, called the relative likelihood, is the optimal discriminating quantity to be calculated in a search for new phenomena. In

contemporary high-energy physics analyses, Monte Carlo generated data is used for approximating the likelihood function. The relative likelihood function is a complicated function in high-dimensional space.

8.1.2 Motivations for investigation of new multivariate methods and novel techniques of interpretation in the context of the $t\bar{t}H\left(b\bar{b}\right)$ analysis

The stable search of the analysis to which this thesis work contributed was approved by the ATLAS Collaboration and was published in *Physical Review Letters D* in 2018 [1]. As mentioned previously, the ATLAS $t\bar{t}H\left(b\bar{b}\right)$ analysis is one of the most challenging of the analyses of the LHC physics programme. Vast datasets have been required to search for the $t\bar{t}H$ process, so any means by which the analyses concerned might be made more efficient is valuable. The $t\bar{t}H\left(b\bar{b}\right)$ analysis combines several methods to enhance the separation of signal and background processes, for the overall goal of maximising the significance of the result. Various selections are applied at many stages and multivariate analysis methods are used to combine in a mathematically compact way many observable variables of events in order to separate signal from background. Specifically, the $t\bar{t}H\left(b\bar{b}\right)$ analysis utilises Boosted Decision Trees (BDTs) in a staged way for reconstruction and classification purposes [304, 229].

This thesis gives detail on the classification abilities of various models implemented for the discrimination of the signal and main background: $t\bar{t}H\left(b\bar{b}\right)$ and $t\bar{t}b\bar{b}$, which have similar final state characteristics. These abilities were qualified using scores of accuracy and precision, together with Receiver Operating Characteristic (ROC) curves, areas under the ROC curve (AUC), Kolmogorov-Smirnov two-sample test statistics, separation plots and class prediction error plots [305].

In addition to this investigation, there was a variety of other aims. One aim was to define constraints on models such that the models were implementable in a reasonable amount of time. Another aim was to investigate the interpretability of the various models – not simply interpreting the capabilities of a classifier, but interpreting its internal modelling and internal representations of abstractions – in order to ensure that the 'black box' thinking prevalent in contemporary high-energy physics analysis is reduced somewhat. This involved interpretations of modelling such as the ranking of input features, the novel synthesis of idealised inputs and novel saliency techniques of interpretation of modelling.

Decision tree and multi-layer perceptron models are well-characterised standards already in use in high-energy physics, so the motivation for including these for comparisons and interpretations is clear. Newer, many-layered deep models were included in the investigation because they have some demonstrated modelling benefits, and in order to take advantage of contemporary GPU (Graphical Processing Unit) computing hardware. Convolutional network models, which are noted for efficiency in modelling of spatial data, were included in the spirit of exploration, as there is no clear mathematical guide currently

for selecting models appropriate to data of complex domains, but were included also with inspiration from observations of apparent cognitive map mental representations (which do not correspond solely to spatial data) in brains and other embedded representations of data, such as word vectors [306, 307, 308]. Inception convolutional network (CNN) models were included, as were ensemble models. Dimensionality reduction methods were assessed also as inputs to models, and an assessment was made of the impacts of including the outputs of some models, such as BDT models, as inputs to other models.

Benchmarks were established in 2014 for deep multi-layer perceptron neural networks trained on SUSY and charged Higgs Monte Carlo samples respectively [309, 20]. "The SUSY Data Set" from this research was formulated as part of a classification problem to distinguish between a signal process which produces supersymmetric particles and a background process which does not, where the final states both contain two leptons and large amounts of missing momentum due to particles invisible to detectors, and so it was used for the purposes of benchmark validation of the various models in preparation for application of equivalent models to the task of classification of signal $t\bar{t}H\left(b\bar{b}\right)$ and background $t\bar{t}b\bar{b}$.

A detailed report on application of the models to "The SUSY Data Set" was not included in this thesis for reasons of brevity, however a few brief comments on its relevant study are included here to give a sense of the background it provided. Following the example of the 2014 study, the features of the $t\bar{t}H\left(b\bar{b}\right)$ dataset were standardised over the entire training and testing datasets with mean zero and standard deviation unity, except for those values strictly greater than zero; these were scaled such that the mean was unity [309, 20]. In the 2014 study, a number of multivariate analysis methods were compared, including decision trees, shallow perceptron neural networks, with and without dropout, and deep multi-layer perceptron neural networks, also with and without dropout. It was observed that the decision trees and shallow perceptron neural networks trained using only low-level features had a performance significantly worse than those trained only on the high-level features, suggesting that the decision trees and shallow perceptrons were not succeeding in discovering independently the discriminating power of the high-level features. This is a well-known shortcoming of shallow learning methods and motivated the engineering of high-level features.

Methods trained in the 2014 study using only the high-level features were found to have a weaker performance than those trained using the ensemble of low-level and high-level features, which suggests that the engineered high-level features, despite the insight represented by them, do not capture all of the information contained in the low-level features. The deep learning methods showed nearly equivalent performance when trained using only the low-level features and when trained using the ensemble of low-level and high-level features, suggesting that the deep learning methods were discovering automatically some of the insights contained in the high-level features.

Finally, in the 2014 study the deep learning methods trained on the ensemble of low-level and high-level

features found additional separation power beyond what was contained in the high-level features, as demonstrated by the superior performance of the deep multi-layer perceptron trained only on low-level features to the performance of a shallow network trained on high-level features. These results demonstrate advantages to using deep learning methods for these types of problems.

8.1.3 Dimensionality reduction: t-SNE and UMAP

Dimensionality reduction approaches seek to produce a lower dimensionality representation of higher dimensionality data that preserves relevant structure, where relevance is application-dependent. Dimensionality reduction is a potential pre-processing step for machine learning, something which was tested in this thesis. Contemporary methods for dimensionality reduction can be classified broadly into two categories: those that involve matrix factorisation and that seek to preserve distance structure within data, such as Principal Component Analysis (PCA), and those that involve neighbour graphs and that seek to preserve local distance structure over global distance, such as T-distributed Stochastic Neighbour Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP).

PCA can be used to exclude dimensions in order to emphasise variation in data. However, t-SNE is a method which reduces the dimensionality of data to two or three dimensions without naïve exclusion of dimensions [310, 311]. It can create maps suggesting how data may be arranged in high-dimensional space. t-SNE converts distances between data in the original space to probabilities and thus can help to indicate whether classes are separable in some linear or non-linear representation. UMAP is another general purpose manifold learning and dimensional reduction method that can be used in a way similar to t-SNE, but is applicable also to general non-linear dimensional reduction [312, 313]. UMAP is considered to be more capable of preserving global structure in data while being more efficient computationally.

There are parameters that guide the functioning of the t-SNE and UMAP algorithms. Any t-SNE perplexity parameters used are quoted in the thesis. The UMAP minimum distance parameter was set exclusively to 0.1 in the thesis and the distance metric is simple Euclidean distance. As no obvious prescription exists for optimising the UMAP number of neighbours parameter, an attempt was made to optimise the UMAP projection of data by calculating and selecting the highest mean of the Kolmogorov-Smirnov statistic for two samples (the signal and background classes) in each axis of the UMAP projection in two dimensions.

8.1.4 Multivariate methods and machine learning algorithms

Machine learning methods use large datasets in order to reduce complexity and to find features in data. The current most frequently-used machine learning algorithms in high-energy physics are decision trees and neural networks. In analysis, usually variables relevant to a physics problem are selected and a

machine learning model is trained for classification or regression using signal and background events (or training instances). Training is computationally expensive while application, also called inference, of a trained model is relatively inexpensive. The decision trees and neural networks are used typically for classification of particles and physics events, and they are used also for regression modelling of a continuous function, for example to estimate the energy of a particle based on various detector measurements.

There is a wide variety of deep neural networks. Some are multi-layer perceptron-type fully-connected networks (FCN), convolutional neural networks and recurrent neural networks (RNN). Neural networks are used in the context of Generative Adversarial Networks (GAN), wherein a network is trained to model a multidimensional distribution of data – the training strategy is to define a competition between two networks; the generator network maps a source of noise to the input space while the discriminator receives as input either an instance of real data or an instance of generated data from the generator network [314]. Neural networks used commonly now for temporal problems, such as time series analysis and prediction are not currently used directly for high-energy physics analyses, but there is a growing interest in application of these methods to data quality monitoring and increasing efficiency in the use of computing resources.

Deep neural networks have been demonstrated to work well in searches for Standard Model and beyond the Standard Model Higgs bosons using 8 TeV LHC data [309]. Specifically, it has been shown that a deep neural network is capable of modelling high-level representations of data almost to the degree of human development of high-level discriminant variables. Other applications of deep neural networks of various forms have been applied in jet-tagging and neutrino event classification [315, 316, 317].

When the signal is small compared to the expected backgrounds, a cut-and-count approach can be used to reach sufficient discriminating power to separate signal from expected backgrounds [229]. However, when the signal count is small compared to the expected backgrounds, multivariate analysis methods such as decision trees and others in the domain of machine learning can be used. Multivariate analysis methods have a long-established, validated history of application in particle physics analyses and other fields. Multivariate methods are used heavily in the $t\bar{t}H$ process searches of ATLAS and CMS [2, 3]. In the context of searches for new physics, multivariate analysis methods adapt to correlations in particle final states in order to model relations between theoretical input parameters (arising from the Lagrangian) and the output, which is a physical final state observed as a particular radiation profile in a detector, as described in Chapter 3. Machine learning methods are of great use for studies involving minimal knowledge of the dynamics that associate the input with the output and in studies involving no clearly-defined physical model at all. However, as discussed, the underlying model, the Standard Model of particle physics, is well-established.

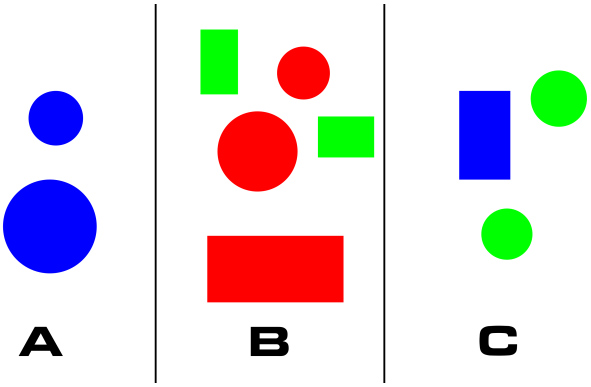
8.1.5 Decision trees

Decision trees are directed graphs beginning with one node and branching to many. They are a hierarchical data structure that represent data by implementing a form of divide-and-conquer strategy, and were likely first introduced as regression tree algorithms in 1963 [318]. There are two main types of decision tree; classification and regression, and both are used to make predictions based on data. Classification trees output a discrete category or class or target while regression trees output real values.

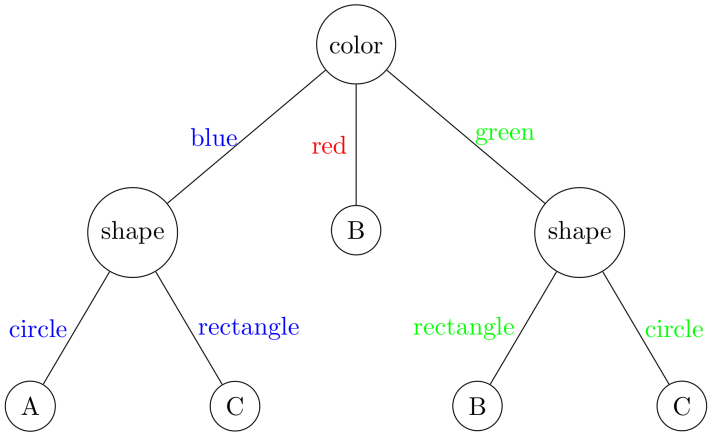
Moving through a decision tree, each node splits the input data based on what are usually simple if-then-else rules. Each node is a sort of cluster of cases that is to be split by further branches in the tree. Often trees are binary, wherein each node is split into two subsamples, but they do not have to be binary. The deeper the tree, the more complex the decision rules and the more fit the model can be. Decision trees can be seen as a compact way to represent a lot of data. A typical goal in defining a decision tree is to search for one that is as small as possible.

Decision trees have a number of useful features, including being usable for numerical and categorical data, being usable for multi-class classification problems. Learned models also are straightforward to interpret, so, for example, a classification of an input instance is explainable in a straightforward way using the logic of the fitted classification tree.

A simple example of a classification decision tree for coloured shapes is shown in Figure 8.1 while a more complex classification decision tree for the iris dataset is shown in Figure 8.2 [319].



(a) Coloured shapes input data.



(b) Classification decision tree.

Figure 8.1: An example of input data of coloured shapes that can be classified as A, B and C, and a classification decision tree that can classify the coloured shapes.

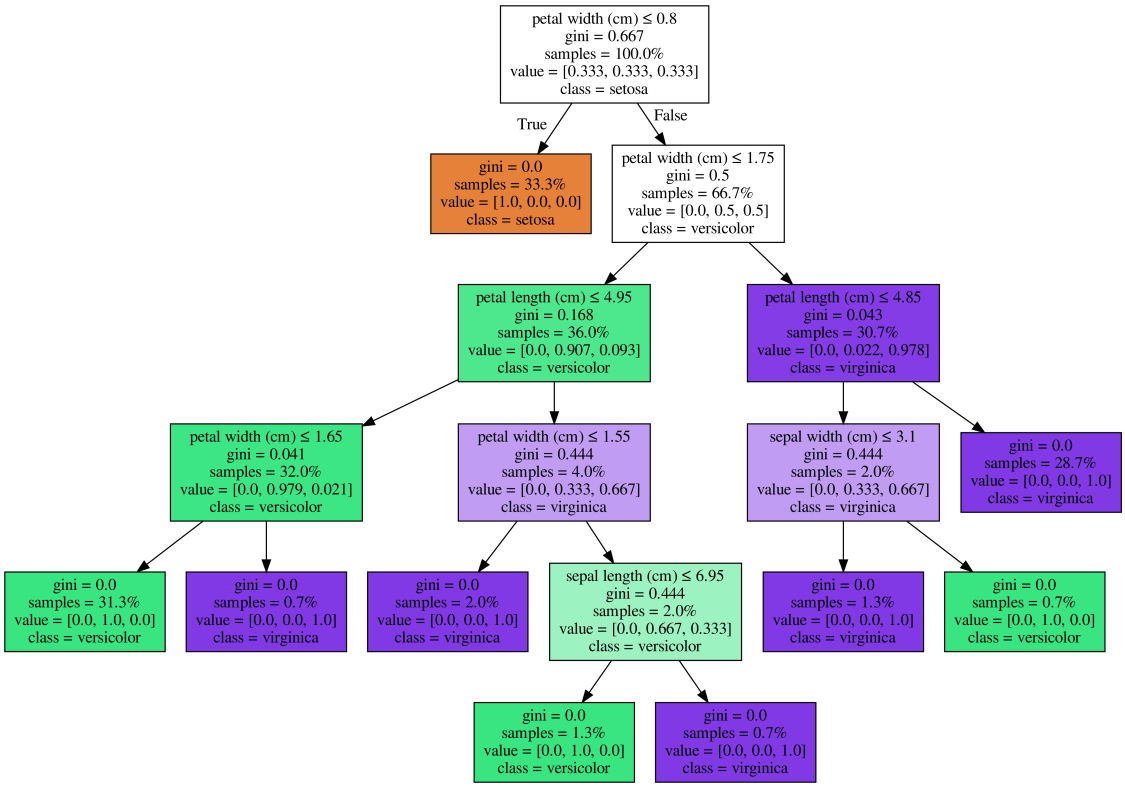


Figure 8.2: Decision tree for the iris dataset.

Gradient boosting classifiers are a machine learning algorithms that combine an ensemble of many weak learning models to form a stronger predictive model. So, a boosted decision tree (BDT) is a model arising from many individual decision trees.

Gradient tree boosting is a widely-used technique which is applied normally to decision trees and it has been shown to be effective on many classification benchmarks [320]. Boosting involves forming a number of hypotheses $h_t(x)$ for the input feature vector x and merging them to form a more accurate composite hypothesis, where the output of the boosted classifier is

$$H(x) = \sum_t \alpha_t h_t(x), \tag{8.1}$$

where α_t is the weighting of each hypothesis (or sub-model) $h_t(x)$. A representation of a boosted ensemble

of decision trees is shown in Figure 8.3.

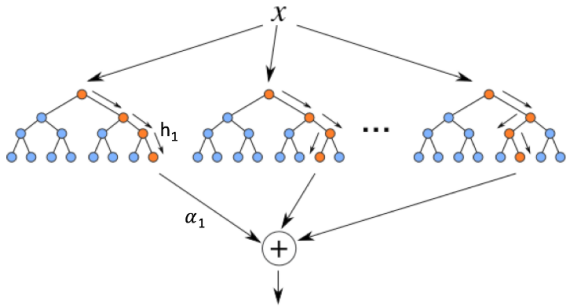


Figure 8.3: Boosted ensemble of decision trees, where x represents the input feature vector, h_1 represents one hypothesis arising from one decision tree and α_1 represents the weighting applied to that hypothesis [321].

There are sparsity-aware algorithms that have built on tree boosting and been successful in high-energy physics contexts, such as models implemented using the XGBoost (eXtreme Gradient Boosting) library [322]. For the case of the HiggsML Challenge, involving an event classification task to identify events featuring Higgs bosons, the winning method was implemented using the XGBoost library, a scalable end-to-end tree boosting system [322, 323]. XGBoost has demonstrated success, and also provides feature importance analysis, a kind of ranking of features in inputs.

8.1.6 Neural networks

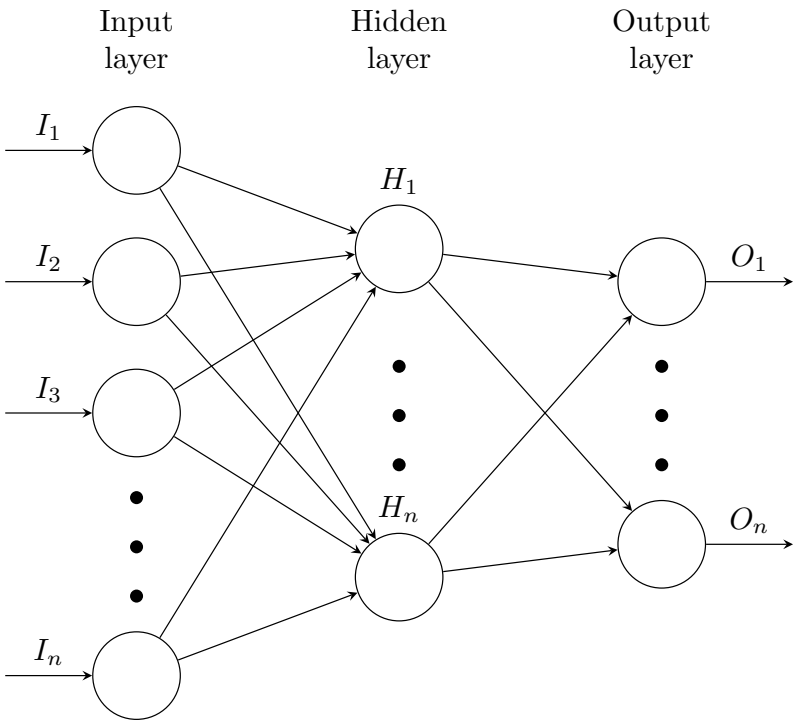
Many creatures feature biological neural networks. Human brains feature approximately ten billion neurons with each neuron connected to approximately ten thousand other neurons. Neurons propagate signals to other neurons. These signals are measured usually as firing pulses. The signals are electrical and mechanical and there are at least two models for neural activity, one being the established Hodgkin-Huxley model, which focuses particularly on the electrical aspects of neural membranes (involving ions moving through the membrane), and another being a soliton model, which is based on hydrodynamic and thermodynamic properties of the neural membranes (perhaps particularly the piezoelectric properties, with the threshold correlated with the force needed to change temporarily the membrane into a liquid crystal, and with anaesthetics possibly working by altering the physical properties of the membrane and, thus, increasing the threshold for propagating a signal) [324, 325, 326]. The neurons of animal brains can have responses that are firing rates (the number of action potentials generated by a neuron over time) and that are sub-threshold activity (such as depolarisations and hyperpolarisations in the membrane potential that do not generate action potentials). The concept of artificial neural networks

was inspired by the functionality of biological neural networks. William McCulloch and Walter Pitts outlined a computational model of a neural network in 1943 [299]. Donald O. Hebb published the book *The Organization of Behavior* in 1949, which outlined a model of biological neural networks in which the connections between neurons strengthened with use, a concept which proved to be fundamental to advancement of the understanding of human learning and to the training of artificial neural networks [300].

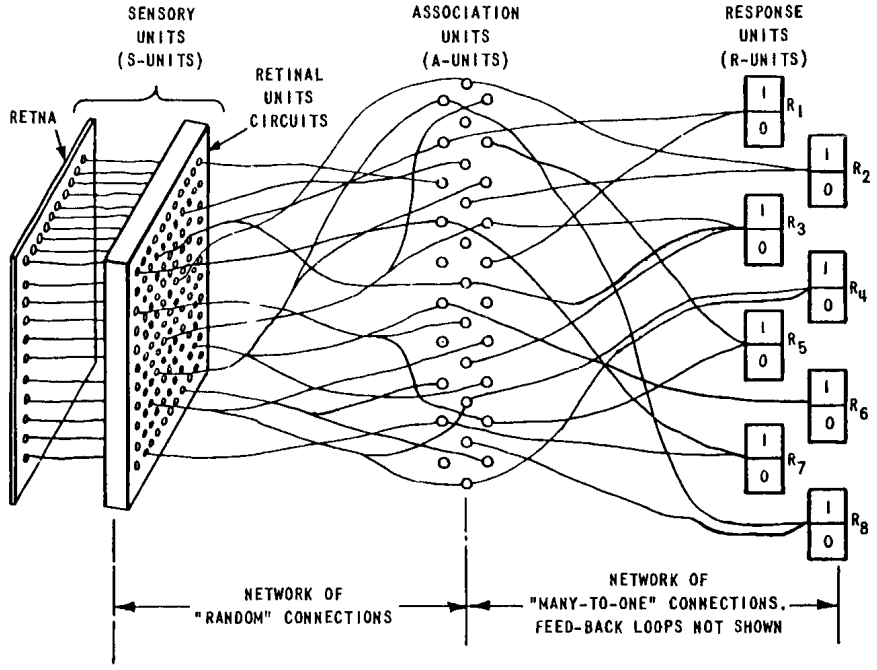
Artificial neural networks are models that attempt to achieve good performance via dense interconnection of simple computational elements. In this respect, artificial neural network structure is similar to a basic contemporary description of biological nervous systems. Neural network models explore many competing hypotheses simultaneously using massively parallel networks composed of many computational elements connected by links with variable weights. Computational elements or nodes used in neural network models are nonlinear and typically analog. A simple node sums N weighted inputs and passes the result through a nonlinearity.

By 1954, Belmont Farley and Wesley Clark, following from the research by McCullough and Pitts, ran the first recorded computer simulations of an artificial neural network (networks of up to 128 neurons and up to two layers, for recognition of simple patterns). By 1957, Frank Rosenblatt built on the work by Farley and Clark to apply their methods to deeper networks, which led to the conception of the perceptron, and to the development of the Mark I Perceptron device [302].

A perceptron (perceiving and recognising automaton) is a feed-forward neural network algorithm [301, 302, 303]. One is depicted in Figure 8.4a. This contemporary picture of the algorithm can be compared with a picture of the organisation of the Mark I Perceptron device in Figure 8.4b.



(a) Perceptron.



(b) Organisation of the Mark I Perceptron from the *Mark I Perceptron Operator's Manual* [302].

Figure 8.4: Comparison of a contemporary depiction of a perception with a 1960 depiction of a perceptron.

A relatively early example of a perceptron in high-energy physics analysis was used by the OPAL Collaboration, whereby Monte Carlo events were used to train a small multilayer perceptron neural network, implemented in the JETNET FORTRAN 77 subroutine package, to separate event classes for the study of $D^{*\pm}$ meson production from Z^0 decays [297, 327].

Such an algorithm as a perceptron features three types of node: input, hidden and output. Input nodes provide input data to the neural network and typically no computation is performed in any of these nodes. Hidden nodes receive information from input nodes and perform computations, and are not connected directly to the input data. Output nodes receive information from hidden nodes and perform computations.

Connections (or edges), between nodes have associated weights (or parameters), and given an input vector these weights express an output vector. Model parameters are parameters internal to the model that are estimated from training data while model hyperparameters are configurations external to the model and typically not estimated from training data. In a feed-forward network, information propagates in one direction from the input nodes through the hidden nodes (if there are any) to the output nodes and there are no cycles in the network. A single-layer perceptron does not contain hidden layers, and can learn only linear functions, but a multi-layer perceptron contains one or more hidden layers and can learn non-linear functions. Given a set of features and a target, a multi-layer perceptron can learn relationships between the features and the target for regression and classification. A goal of learning in this context is to assign appropriate weights to the edges between nodes, usually by broadcasting an error from output nodes back towards input nodes, in a calculation called back-propagation, which is one way by which an artificial neural network can be trained [328]. In this approach, edge weights are assigned pseudorandom values initially. For each instance in the training dataset, the neural network is activated and the output calculated. The output is compared with the target output and the error, or cost, is propagated back to previous layers by adjusting weights using a method such as gradient descent. This adjustment is repeated until the output error is below a certain threshold or until some other stopping criterion is met, resulting in a "learned" network which is then tested on new input data.

In updating the weights, because only the values of the output nodes are known directly, the weight adjustments for hidden nodes are made indirectly. In a multi-layer perceptron architecture, neurons between adjacent layers are fully-connected. For the i^{th} neuron in layer $N+1$, the output y_i is computed by $y_i = \sum_j f\left(w_{ij}x_j + b_i\right)$, where x_j is the output from the j^{th} neuron in layer N, w_{ij} is a weight associated with each edge and b_i is a number called the bias. In matrix form, information propagates from one layer to the next first by a linear transformation and then is processed by an activation function $f : Y = f\left(WX + B\right)$. So, at a simple level of description, a neural network computes a function that contains generally linear combinations of variables combined with functions that add non-linearities.

Broadly the mathematical principle underlying a typical form of neural network training is spectral decomposition, where the aim is to identify the principal axes of data, considered as a structure in a configuration space of some dimensionality. Reweighting connections between neurons changes the orientation of this structure. In back-propagation, an error function is propagated back from the outputs through the network. In one sense it is a feedback connection. A loss function quantifies the shortcomings of the network and, thus, an objective of training is a constrained minimisation of the loss function, using a method such as gradient descent, which involves calculating the gradient of the loss function and then varying its inputs in the direction of the negative gradient, hopefully towards a minimum for the loss function [295]. A typical error function might have a form like the following:

$$E_k = \sum_{k=1}^{N} \frac{1}{2} \left(y_k - \hat{y}_k \right)^2 , \tag{8.2}$$

where N is the total number of instances in the training dataset, y is the weighted neural network output and $\hat{y}$ is the correct response. Thus, in the gradient descent approach to minimising the error, on each iteration the weights are updated:

$$\vec{w}^{\rho+1} = \vec{w}^{\rho} - \eta \nabla_{\vec{w}} E, \tag{8.3}$$

where $\vec{w}$ represents the weights, η is the learning rate (a step size taken per training iteration), $\nabla_{\vec{w}}$ is the gradient with respect to the weights and ρ is the number of training cycles [295].

A neural network can result in a function of many thousands of parameters, such as weights and biases, with neurons considered as functions in layers of the network. The activation of each neuron in a layer is based on a weighted sum of activations in the previous layer (all the activations in the case of a typical multi-layer perceptron), with a bias, and this sum is composed with an activation function such as sigmoid, ReLU *etc*. Broadly, a motivation for different layers is for neurons in those layers to correspond to various learned features in data, at different levels of abstraction. In learning, labelled data is presented to the network and the weights and biases adjusted so as to improve the performance of the network. Hopefully what this layered structure learns generalises beyond the training data. This can be tested by ensuring that after training the network, it is presented with more labelled data with which it has not been presented previously.

The weights and biases are set pseudorandomly initially prior to training and learning. The term "learning" in this context tends to refer to the brute-force calculus exercise of finding a minimum of a function. The cost (an error measure) for each training case is calculated, usually as the sum of the squares of

the differences between the output activations and the desired output activations that correspond to the labels on the data.

There are many neural network architectures, some of which are shown in Figure 8.5.

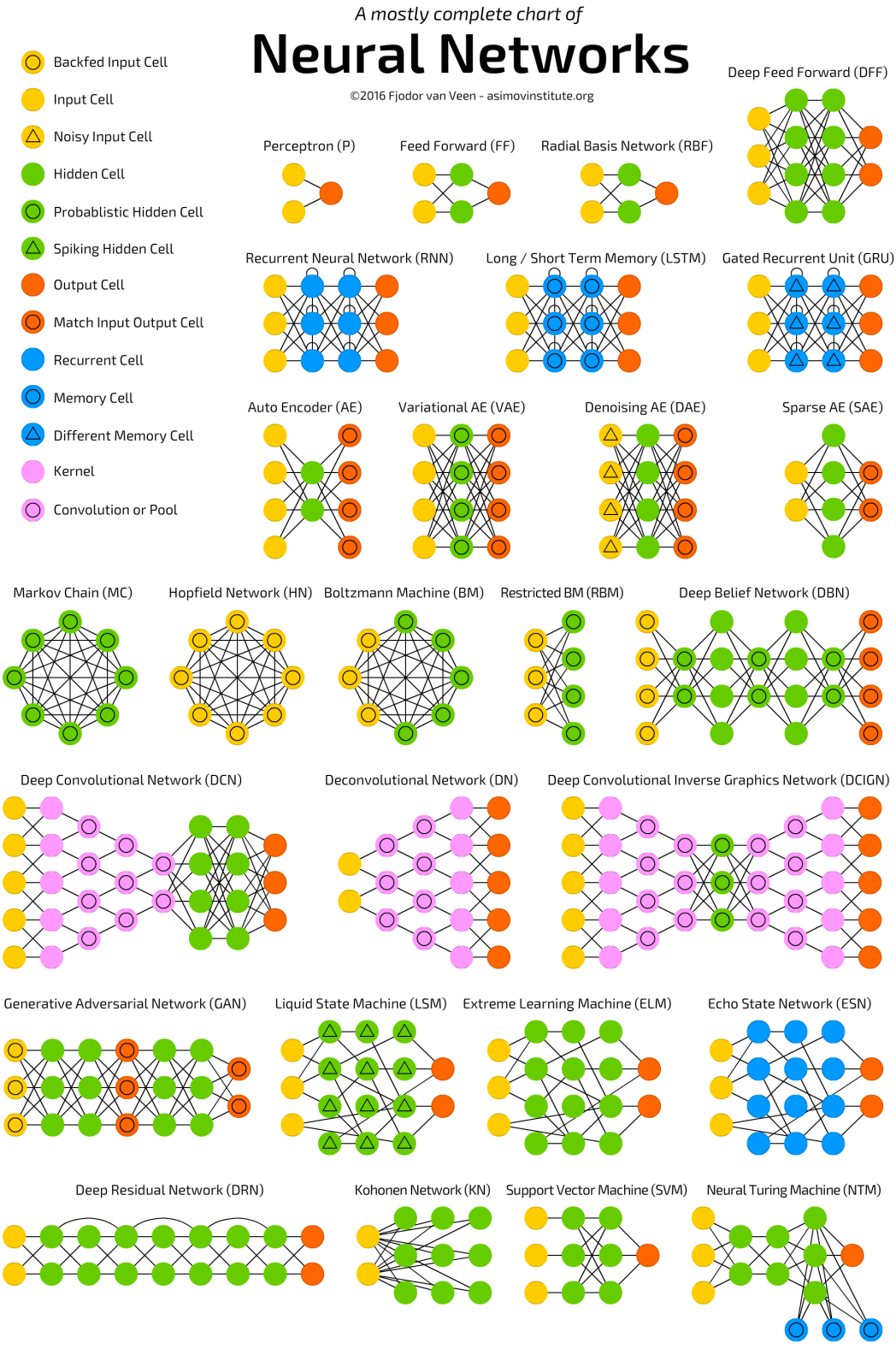


Figure 8.5: The Neural Network Zoo [329].

There are other characteristics to consider in the definition of a neural network model. A smaller batch size is generally preferable for the sole goal of ensuring generalisability of a neural network model [330, 331]. For the purposes of this thesis, batch sizes were selected that resulted in well-generalisable models within a few hours using the computing hardware available. The hyperbolic tangent activation function tends to perform better than the sigmoid and rectified linear unit activations (ReLU) and so was assessed and then used for models throughout this thesis [332].

8.1.6.1 Deep Learning

A deep multi-layer perceptron feed-forward neural network, with its many fully-connected hidden layers, is depicted in Figure 8.6.

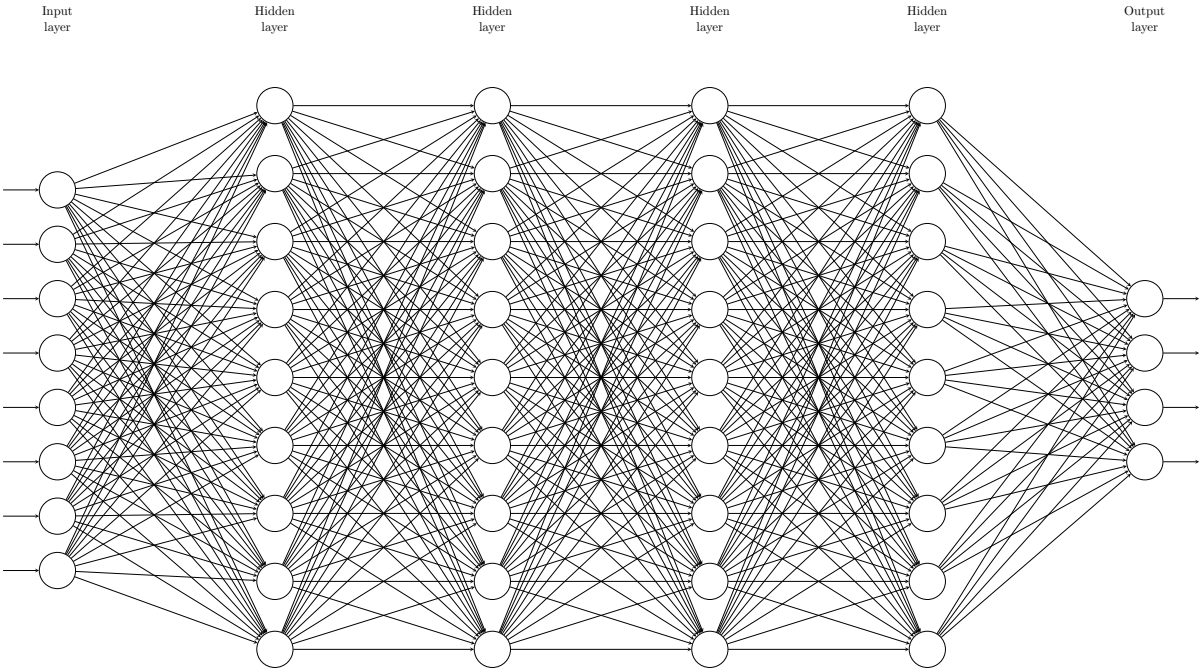


Figure 8.6: Deep Perceptron.

"Deep learning" is a contemporary term used to refer to modelling or learning complicated concepts of varying degrees of abstraction by building them from more basic ones in a hierarchies or multiple layers. Artificial neural networks of multiple layers are instances of such deep multi-layer hierarchies. Just as the mathematical field of chaos theory advanced with progression of computational power, so too has the development of large neural network and many layers and degrees of freedom been enabled by the progression of computational power, such as contemporary GPUs, and by the availability of large

datasets [333].

The success of deep neural networks is attributable to many of their characteristics. A principal one may be their ability to utilise the statistical properties of data such as stationarity and compositionality through localised statistics, such as are present in images and sounds observed in nature. These statistical properties are emergent physical phenomena.

Deep neural networks appear to have an advantage over shallow networks in learning complex functions, but this advantage is not entirely understood. One possibility is that the extra representational power provided by the deep network architecture is necessary to represent complex features relevant to classification. High-level engineered features are useful for classification, and attempts at deriving similarly complex features from low-level features alone are more easily facilitated by deep network architectures than shallow network architectures [334]. While deep learning is effective, it is not fully understood *why* it is effective [335]. The 2017 paper *Why does deep and cheap learning work so well?* considers how the efficiency of deep learning methods depends on both mathematics and physics [335][1].

8.1.6.2 Convolutional neural networks

Convolution is a mathematical operation whereby a tensor or a matrix or a vector is summarised or convolved into a smaller one. A one-dimensional input matrix is summarised along its dimensions; a multi-dimensional tensor can be summarised across all of its dimensions. Using the KERAS terminology, the layers *Conv1D* and *Conv2D* convolve along one or two dimensions, respectively.

For example, a vector could be convolved into a shorter vector in the following way. The vector a with n elements is convolved using the weight vector w with m elements into a smaller vector b with $n - m + 1$ elements:

$$b_i = \sum_{j=m-1}^{0} a_{i+j} * w_j, \tag{8.4}$$

where $i = [1, n - m + 1]$. So, for a vector of length n, the weight matrix also is n, $w_i = \frac{1}{n}$, then the convolution produces a scalar or a vector of length 1 equal to the average of the elements of the input matrix, a sort of degenerate convolution. If the same weight matrix is one less than the input matrix, then there is a moving average in the output of length 2 and so on.

[1] The paper also provides something of a dictionary for translation between contemporary mathematical physics terminology and contemporary machine learning terminology ("effective theory" → "nearly lossless data distillation", "translationally symmetric Hamiltonian" → "convolutional neural network" *etc.*).

$$\begin{bmatrix} a: & a_1 & a_2 & a_3 \\ w: & \frac{1}{2} & \frac{1}{2} & \\ w: & & \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} b: & \frac{a_1+a_2}{2} & \frac{a_2+a_3}{2} \end{bmatrix}. \tag{8.5}$$

A three-dimensional tensor or vector may be treated similarly;

$$b_{ikl} = \sum_{\substack{j_1=m_1-1 \\ j_2=m_2-1 \\ j_3=m_3-1}}^{0} a_{i+j_1,k+j_2,l+j_3} * w_{j_1 j_2 j_3}, \tag{8.6}$$

where $i = [1, n_1 - m_1 + 1]$, $k = [1, n_2 - m_2 + 1]$ and $l = [1, n_3 - m_3 + 1]$.

So, for a one-dimensional convolutional mechanism, information flows to the next layer by a convolutional operation and is then processed by an activation function

$$f : Y = f\left(W * X + B\right), \tag{8.7}$$

where $*$ denotes the discrete convolution operation between the input signal X and a convolution kernel W. A convolutional layer features typically several kernels all with the same size, it convolves the input signal with each kernel and then stacks the result as the output of the layer. During convolution, a sliding window of the size of the kernel moves through the input signal with a defined stride distance. During each hold of the window, the inner product of the examined section of the input signal with the kernel is calculated.

Fully-connected multi-layer perceptrons have each neuron in one layer connected to all neurons in the next layer. The fully-connectedness of these architectures makes them prone to over-fitting data. Convolutional neural networks, however, are designed to take advantage of hierarchical patterns in data and to assemble more complicated patterns using smaller and simpler patterns, which helps with regularisation to address over-fitting, and makes them less extreme in terms of connectedness and complexity [336]. Aspects of convolutional neural networks were inspired by processes of the visual cortex, specifically whereby individual brain neurons respond to features of the visual input in a restricted region of the visual field only, a region called the visual receptive field [337, 338]. Convolutional neural networks are a form of feed-forward neural networks. In a basic, standard convolutional classification neural network, there are two main groups of layers in the full stack of layers. The first layers are convolutional and max-pooling, which can act as feature extractors for inputs such as images. The second layers are fully-connected, multi-layer perceptron-style layers which perform non-linear transformations of extracted features and then output a classification, using a layer which could be softmax.

Convolution is the weighted sum between two functions. For a two-dimensional input, say an image, a section of the input of size $k \times k$, where k is referred to as the kernel size, is taken from location (x, y) of the input. Then the values in this kernel section are multiplied element-wise with a convolution filter (also of size $k \times k$) and then the results are summed to a single output. The convolution kernel is translated using a specific stride over the entire input matrix to generate an activation map, so basically if the trained convolution layer produces a high activation, it is assumed that that it has detected an internal representation or feature. A benefit of the convolutional approach is that it is translation invariant. After each convolution, the output reduces in size and it is often the convention to pad zeros to the boundary of the input layer such that the output is the same size as the input layer. Initial and lower convolutional layers are a search over small sections of the input, but the neurons of deeper, higher layers get information from larger sections of the input, so they can learn more complicated features and can enable derivation of more hierarchical features. A pooling layer is applied usually directly following a convolutional layer to reduce the spatial size (so reducing width, height and depth to just width and height, for example). It reduces the input by some fraction, often half. A flatten layer converts the output of a lower layer to a single array (for example, a three-dimensional array could be converted to a one-dimensional array).

8.1.6.2.1 Why try convolution in the context of classification of high-level high-energy physics event variables?

Part of the motivation for testing architectures with internal convolutional layers is in the spirit of the conventional haphazard approach of iterating and scanning through many architectures because we lack a clear theoretical guide on appropriate architectures for various data domains. Part of the motivation is in the spirit of recent successes in temporal convolutional neural networks, wherein convolutional layers are applied within recurrent layers. Part of the motivation is the finding that convolutional neural networks suited for object recognition can be trained to recognise words in the context of proposals for theoretical approximations to the visual word form area, which is a region of the human cortex involved in human reading acquisition [339]. Part of the motivation for convolutional neural network models also is the interpretability of the modelling of such networks [340].

A concept in convolutional neural networks is a sort of 'moving filter' (also referred to as a kernel), which passes over the input. This moving filter is a convolution that applies to a collection of nearby input values. In contrast with fully-connected neural networks, wherein every node of one layer is connected to every node in the next layer, in a convolutional neural network, connections are *sparse*, wherein not every input node is connected to the output nodes. In conventional, contemporary convolutional neural networks, each filter has constant parameters; as the filter passes over the input, the same weights are applied, and thus each filter performs a specific transformation across the whole input. Again, this is in

contrast with fully-connected neural networks which have a different weight value for every connection. Note that this is not to imply that each weight is constant *within* the filter; the weights within the filter are values derived from training. These features of the sparse connections and the constant filter parameters can reduce the numbers of parameters required in the network, in contrast with the numbers of parameters that might be required in fully-connected neural networks. The output of convolutional mappings is passed through some activation, in a step referred to as *feature mapping*.

In pooling, a sort of 'moving window' is used, with some appropriate statistical function applied to the values within this window. Often this is simply taking the maximum value in the window, which is referred to as "max pooling", but other variants are used, such as "mean pooling" or "L2-norm pooling". Pooling reduces the numbers of parameters in the network – it is referred to also as "down-sampling" for this reason. Also, particularly of use in the context of visual data, it makes feature detection more resistant to scale and rotational changes. In a sense, this makes pooling an operation that generalises over lower-level, more complicated information, and so enables a change from complicated *data* to lower 'resolution' *information*. In natural images, information is conveyed at various frequencies. Higher frequencies encode finer details and lower frequencies encode more global structures. Similarly, the output feature maps of a convolution layer can be considered a mixture of information at various frequencies [341].

So, broadly, conventional convolutional neural networks involve input data passing through convolution and pooling layers of various architectures, and, in a sense, multiple architectures at once, and then signals propagate from these layers to flattening, perceptron/fully-connected layers and then in, for example, the case of a classification problem, to a softmax layer for the full transformation of the input data to a statement of its classification. Convolutional neural networks are not limited to Euclidean manifolds [342]. Representations of convolutional neural network layers are shown in Figure 8.7.

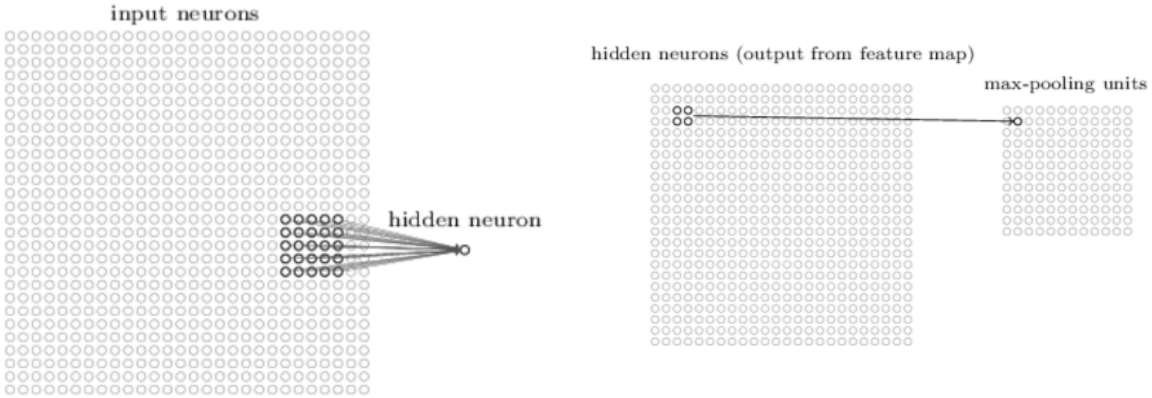


Figure 8.7: Representations of convolutional neural network layers [343].

8.1.6.2.2 Inception

Inception modules were designed and investigated in this thesis also. The concept of the inception module focuses on the kernel size. The inception module uses multiple types of kernel at once [344]. Dropout regularisation can be considered as an approach that trains a vast multiplicity of neural network architectures at once. The inception module could be considered in a similar way. A number of "towers" of convolutional layers, and possibly dimension reductions, of varying characteristics, such as kernel sizes, can be defined as an ensemble with the same input propagated through all towers and the outputs of all towers effectively concatenated and thence propagated to a typical multi-layer perceptron neural network infrastructure, followed by softmax for this case of classification. While a hyperparameter scan involving large variations in these forms of architecture could be worthwhile, only the briefest of examinations of these forms of neural network were considered with limited hyperparameter explorations.

8.1.6.3 Training and validation samples

Over-fitting, or over-training, put simply, is when a trained model works well on training data, but does not work well on validation or test data.

In order to ensure adequate model generalisability or flexibility, and to avoid over-training or over-fitting, a validation sample is used which contains instances much like the training sample. In training, model parameters – and hyperparameters, if used – generally increase in flexibility and modelling such that, for the case of a signal and background problem, separation between classifications improve as the decision boundary in the model improves. Generalisation performance of the model is tested on an independent

validation data sample. As training progresses, the usual observation is that performance on the validation sample increases to a certain point and then the model can become "overtrained". "Overtraining" is the observation of small-scale features in the decision boundary of the model being dominated by statistical (assumed irrelevant) fluctuations in the training sample rather than by realistic features and structures of the underlying data. The generalisation properties of a classification model are also considered in this context; an overtrained classification model does not generalise as it does not model the realistic, underlying features of the data.

For theoretical completeness, once a reasonable model generalisability has been trained and validated using the training and validation samples, a separate sample, called the testing sample, can be used for a more unbiased estimation of the generalisation performance of the classification model.

Equal numbers of training and validation instances (events) were used and, as is standard in experimental high-energy physics, a testing sample is considered as equivalent to a validation sample. This is considered acceptable if the number of training instances is sufficiently large such that statistical fluctuations are not significant (as can be ensured straightforwardly in the case of the Monte Carlo event samples of high-energy physics), and if there are a variety of other conditions [229]. For situations in which there are greatly differing numbers of instances of various classes (such as signal and background), there are methods such as "focal loss", which 'reshape' standard cross entropy loss such that the loss assigned to high-confidence instance classifications are under-weighted [345]. There are other ideas, such as adversarial training, which could be considered as equivalent to specifying complicated loss functions indirectly, and so they can be useful for generative purposes and for enforcing constraints and, as such, such approaches should be directly useful in scientific fields such as particle physics [346].

When evaluating the performance of the classification models presented in this thesis, 70 % of the sample was used for training and 30 % was used for validation/testing.

8.1.6.4 Performance metrics

The Gini coefficient is a measure of statistical dispersion which can be used as a single number that describes the income or wealth distribution of a population and is a commonly-used measurement of inequality. A Gini coefficient of close to zero corresponds to maximal equality, for example where a population has the same income for all members. A Gini coefficient of close to one corresponds to maximal inequality, for example where one member of a population has all the income and all other members have none. A Gini coefficient of greater than one might correspond to some members of a population having negative income, while a Gini coefficient of less than zero would correspond to a negative mean. Specifically, the Gini coefficient can be defined as half of the relative mean absolute difference. The mean absolute difference is the mean absolute difference of all pairs of items of the

population, and the relative mean absolute difference is the mean absolute difference by the mean in order to normalise for scale.

The Kolmogorov-Smirnov two-sample test statistic can be used to compare two samples. It quantifies a distance between the empirical distribution functions of two samples. It is considered one of the more useful and general nonparametric methods of comparison of two samples as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples [305]. The empirical distribution function F_n for n independent and identically distributed values X_i is

$$F_N(x) = \frac{1}{n} \sum_{i=1}^{n} I_{[-\infty,x]}\left(X_i\right), \tag{8.8}$$

where $I_{[\infty,x]}\left(X_i\right)$ is the indicator function, which is equal to 1 if $X_i \leq x$ and equal to 0 otherwise. Thus, the Kolmogorov-Smirnov test statistic for a cumulative distribution function $F(X)$ is

$$D_n = \sup_x \left|F_n(x) - F(x)\right|, \tag{8.9}$$

where $\sup_x$ is the supremum of the set of the distances.

The Kolmogorov-Smirnov two-sample test statistic is used to express the difference between two one-dimensional distributions. The statistic in this case is

$$D_{n,m} = \sup_x \left|F_{1,n}(x) - F_{2,m}(x)\right|, \tag{8.10}$$

where $F_{1,n}$ and $F_{2,m}$ are the empirical distribution functions corresponding to the first and second samples respectively.

For the purposes of this thesis, two classes, corresponding to signal and background, are considered, so the Kolmogorov-Smirnov two-sample test statistic may be applied. In 2021, the Kolmogorov-Smirnov test statistic was extended such that it is applicable to multivariate cases [347]. This development may be applied in machine learning, both because it enables the measurement of the statistical distance between a training dataset and a comparison dataset, and also because it may be used to confirm that some subsample of a dataset is representative, or largely statistically indistinguishable, from the dataset.

Classification models can be compared in terms of a simple score, expressing the fraction of cases classified correctly by a model. Classification models can be compared also by precision, which is given by

$$\frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fp}}, \tag{8.11}$$

where tp is the number of true positives and fp is the number of false positives. So, the precision is an expression of the ability of a classifier not to label as positive a sample that is negative.

An illustration of the performance of a signal/background binary classification model is provided by the Receiver Operating Characteristic (ROC) curve, which is a plot of the true positive rate against the false positive rate for various threshold values [348]. The true positive rate is plotted typically on the vertical axis and the false positive rate is plotted typically on the horizontal axis, so the top left of the plot corresponds to a false positive rate of zero and a true positive rate of one – the ideal performance. This implies that a larger Area under the Receiver Operating Characteristic Curve (AUC) usually corresponds to better performance at classification, with greater AUC values corresponding to a greater classification accuracy across a range of threshold values. Since it is ideal to minimise the false positive rate while maximising the true positive rate, the slope of ROC curve is illustrative of the performance. The AUC is a standard metric in machine learning and it is insightful because it is connected directly to the classification accuracy, which is the quantity training of the models attempts to optimise. Beyond this metric, other metrics could be of interest, such as signal efficiency at some fixed background rejection, or discovery significance calculated from the p-value in the null hypothesis. Small increases in the AUC can represent significant increase of discovery significance [309].

The class prediction error plot is here defined such that it illustrates the relative numbers of validation cases for each class (in this case, signal and background) in the classification model as a stacked bar chart. Each bar is segmented to show the proportion of predictions, including false negatives and false positives, for each class. This is a way to visualise those classes for which the model has a particular challenge in classification, and the fraction of incorrect classifications the model is making on a per-class basis. For two classes, it is somewhat limited, but for more than two classes, it can become useful for comparing the effectiveness of various classification models at predicting multiple classes.

8.1.6.5 Calibration

In a two-class classification network, the model involves some function to map the output to between 0 and 1. Each bin in the distribution of the output value can have associated with it both an accuracy and a confidence. Accuracy is the proportion of correct classifications over the number of outputs in a bin, while confidence is the mean of the outputs in a bin;

$$\text{accuracy} = \frac{1}{B_m} \sum_{i \in B_m} 1\left(\hat{y}_i = y_i\right) \tag{8.12}$$

$$\text{confidence} = \frac{1}{B_m} \sum_{i \in B_m} 1\left(\hat{p}_i\right). \tag{8.13}$$

8.1.6.6 Hyperparameter scanning and optimisation

In the context of multivariate models, such as multi-layer perceptrons, parameters are coefficients of the model which are usually determined by the model itself. Weights are an example of parameters, which are optimised during training. The contemporary term "hyperparameters" refers to the parameters or elements that define the architecture and other behaviours of the multivariate algorithms, such as the numbers of neurons in a layer or the neuron dropout probabilities or the numbers of convolutional towers in an inception module. Just as parameter training was performed in defining the models assessed, so too was hyperparameter scanning performed, in both automated and manual ways. Research is ongoing on automated approaches to prediction of model parameters by hypernetworks, which attempt to utilise the experience of training other networks to predict parameters of new networks, but the practical application remains limited [349].

The training optimisation algorithm Nesterov-accelerated Adaptive Moment Estimation (NADAM) was selected for cross-entropy loss due to its contemporary suggested superiority over the more typical adaptive moment estimation (ADAM) and other approaches [350, 351, 352]. Otherwise, generally standard approaches were taken for pre-training treatment of samples (standardisation, standardisation of imputation *etc.*) as were generally standard approaches used for consistency and reproducibility, such as ensuring consistent NUMPY seed numbers and so on [353].

8.1.6.7 Uncertainties

An introduction to systematic uncertainties is provided in Section 5.7, but broadly a measurement is a value with statistical uncertainties and systematic uncertainties stated, where statistical uncertainties are "unknown unknowns", basically noise, and where systematic uncertainties are "known unknowns" comprised of imperfections in experimental apparatus, imperfections in model simulations, limitations of theoretical knowledge and computational limitations. A nuisance parameter is a parameter of the measured data distribution that is not a parameter of interest for a given analysis. The major systematic uncertainties arising from calibrations and other experimental effects can usually be considered by studying the input observables, so can be detected in one-dimensional projections. If a multivariate selection explicitly attempts to disentangle correlations in such observables, it is more prone to systematic uncertainties arising from a possible mismodelling of the correlations. Study of underlying limitations in simulations can involve consideration of all possible input distributions in projections of one, two or more dimensions using subsets of the observables in the classification model.

While statistical uncertainties of machine learning methods can be estimated through relatively straightforward cross-validation, one of the challenges of applying machine learning methods to scientific investigations is the incorporation of systematic uncertainties, which affect both the robustness of inference

and the metrics used in the evaluation of a particular analysis strategy. Machine learning methods of the type used in this thesis are not considered theoretically to carry systematic uncertainties in themselves. Theoretically, following a typical training and deployment procedure, the free parameters of a resultant trained model are held constant for an analysis and, so, multivariate methods such as perceptron neural networks can be treated as functions mapping the space spanned by input variables to the space spanned by probabilistic predictions. Differences to more conventional functions, such as those which might be applied in the definition of some higher-level variable using the kinematic properties of particles, could be the dimensionality of the input space being very much higher, and hidden relations between the input and output spaces. For modern machine learning models, one could conceive of systematic uncertainties associated with various phenomena, everything from computational bits flipping in computing hardware to certain evolutionary algorithms learning imperfections in the silicon of Field Programmable Gate Arrays (FPGA), however these are considered inapplicable and negligible in the scope of analyses of the forms presented in this thesis [354].

In a typical contemporary experimental high-energy physics context, the training of machine learning models for classification does *not* involve introduction of systematic uncertainties [355]. The influence of systematic uncertainties is considered only on the classification model output distribution. This does not involve new training for each uncertainty, but simply an estimate of how the classification model output distribution changes for the already-trained model. This is a typical approach, but there are some advancements in uncertainty-aware machine learning. There are significant open aspects of research on the topic of systematics-aware machine learning and some suggestions that uncertainty-aware machine learning training could be subject to clear theoretical improvement, perhaps by making careful theoretical connections between concepts like dropout and the ensemble of various uncertainties on input data [356, 357].

Quantification of uncertainty in neural networks is potentially quite a complex subject theoretically, and one that may not be defined clearly, so the standard approach of propagating uncertainties through a trained model is used in the qualification of models in this thesis [358]. Specifically, for the case of high-energy physics data, models were trained on nominal simulated data corresponding to signal and background and then the data were perturbed by a collection of uncertainties identified in the analysis to gain a sense of the reliability of classification of models. A systematics-aware learning procedure optimises between both approaches, by attempting to learn an efficient representation, which is insensitive to perturbations induced by systematic effects. A number of approaches to this have been considered and continue to be under consideration [357]. Under investigation in 2019 was a boosting approach for boosted decision trees which was proposed as a way to include and reduce the effects of systematic uncertainties for the purposes of event classification in high energy physics [359]. Reduction of the correlation between

signal strength and sources of systematic uncertainty, resulting in a greater significance, was reported for the use of this QBDT approach (where the Q refers to the likelihood ratio) [359]. Under investigation in 2021 was the epistemic neural network approach that serves as an interface for uncertainty modelling in neural networks [360]. Also under investigation for the purposes of LHC analyses is Inference-Aware Neural Optimisation (INFERNO), which appears to provide a promising approach to addressing the problem of systematic uncertainties being neglected in the training of multivariate models [361]. In training, the loss is inversely proportional to the uncertainty on some parameter of interest, resulting in multivariate models being inclined to become sensitive to a point of interest and resilient to nuisances.

Whether the inclusion of systematic uncertainties in automated optimisation of a high-energy physics analysis impacts negatively the reliability of the statistical inference performed in the analysis could be considered an open question. One could conceive of a tuned automated mitigation of systematic uncertainties being over-sensitive to actual variations in data, a conceptually similar situation to that which might arise from overtraining a model, wherein a model becomes over-sensitive to statistical fluctuations in training data, impacting negatively its ability to generalise to other data. Multiple cross-check estimations of systematic uncertainties may be prudent in the pursuit of reliably generalising multivariate models. The open nature of questions on including systematic uncertainties in robust analysis strategies in modern machine learning models mandates a cautious approach in validation of proposed theory and deployed techniques.

8.1.6.8 Dropout

Combining predictions from many multivariate models is one way to reduce testing/validation errors, though this can be expensive and inefficient for models that can take days to train.

Perhaps hinted at by the ideas of memory decay in the Perceptron Mark I device (discussed in Section 8.1.6, and shown in Figure 8.4b), or by neural/sensory adaptation involving a gradual decrease of responsiveness of an organic sensory neuron over time, "dropout" concerns a probability of ignoring neural network nodes during the training phase, specifically meaning that a certain set of nodes are selected and then not considered during a particular forward and back-propagation pass in training [302, 362, 363]. At each training stage, individual nodes and their associated incoming and outgoing edges have a defined probability of being dropped from the neural network such that a reduced network remains. Dropout is an efficient form of model combination that increases the required training time only by a factor of about two for a dropout probability of 0.5 [364]. So, every time an input is presented in training, the neural network samples a different architecture, but all the architectures sampled share weights.

The 2012 research paper by Hinton *et al.* sets out its abstract as follows:

> When a large feed-forward neural network is trained on a small training set, it typically performs poorly on held-out test data. This "over-fitting" is greatly reduced by randomly omitting half of the feature detectors on each training case. This prevents complex co-adaptations in which a feature detector is only helpful in the context of several other specific feature detectors. Instead, each neuron learns to detect a feature that is generally helpful for producing the correct answer given the combinatorially large variety of internal contexts in which it must operate. Random "dropout" gives big improvements on many benchmark tasks and sets new records for speech and object recognition.
>
> *Improving neural networks by preventing co-adaptation of feature detectors* (2012) [362]

So, broadly, dropout is a model combination method used to prevent over-fitting, or is a regularisation strategy, and it is now a widely-used machine learning technique. In a neural network, a fully-connected layer of nodes occupies most of the neural network parameters and, hence, nodes can develop co-dependency amongst one another during training, which curbs the modelling power of each neuron, which leads to over-fitting of training data. So, dropout helps to reduce interdependent learning, or co-adaptation, amongst neurons. By avoiding training all nodes on all training data, dropout decreases over-fitting. Dropout is applied usually to hidden nodes; dropout of input nodes corresponds directly to loss of information. The strategy also can improve training speeds significantly.

Another way of viewing dropout is as an effective simulation of a large number of networks with very different network architectures while making neurons generally more robust.

In the 2012 research by Hinton *et al* which explicitly proposed dropout layers, a dropout probability of 0.5 was used on each of the fully connected (dense) layers before the output; it was not used in convolutional layers [362]. This became a commonly-used architecture. More recent research has shown that there is value in applying dropout to convolutional layers, although at dropout probabilities of much lower levels, perhaps between 0.1 and 0.2 [365]. In these more recent findings, dropout was applied after the activation function of each convolutional layer and regularisation of the convolutional neural networks was observed by adding this sort of noise to the output feature maps of convolutional layers, yielding robustness to variations in inputs. A mixture of the original approach and that more recent approach is one technique investigated as part of an ensemble of techniques presented in this thesis. The differences in the smoothness of any plots in this thesis corresponding to the training and and testing process, are accounted for by the dropout technique applied to the learning phase and not to the testing phase.

The application of dropout to fully-connected neural network layers is a different operation to the application of dropout to convolutional network layers. While dropout has been demonstrated to have benefits in

both cases, the benefits observed for fully-connected neural networks are different to the benefits observed for convolutional networks. This may be due to the fact that activation units in convolutional layers are spatially correlated so information can still flow through convolutional networks despite dropout [366].

An n-layer fully-connected neural network (ignoring bias) could be defined as

$$f\left(\vec{x}\left\{W_{i\in\{1,\ldots,n\}}\right\}\right) = \phi_n\left(W_n\phi_{n-1}\left(W_{n-1}\cdots\left(\phi_1\left(W_1\vec{x}\right)\right)\right)\right), \tag{8.14}$$

where ϕ_i is some non-linearity (*e.g.* ReLU), W_i for $i \in \{1, ..., n\}$ represents weight matrices and $\vec{x}$ is the input. Dropout applied here is equivalent to zeroing-out a column from the weight matrix associated with a full-connected layer. Qualitatively, this should promote redundancy in the weight matrix and sub-networks can robustly perform the desired operation.

An n-layer convolutional neural network (ignoring bias) could be defined as

$$g\left(\vec{x}\left\{K_{i\in\{1,\ldots,n\}}\right\}\right) = \phi_n\left(K_n * \phi_{n-1}\left(K_{n-1} * \cdots * \left(\phi_1\left(K_1 * \vec{x}\right)\right)\right)\right), \tag{8.15}$$

where $*$ is the convolution operator, ϕ_i is some non-linearity, K_i for $i \in \{1, ..., n\}$ represents weight matrices and $\vec{x}$ is the input. Dropout applied here again is equivalent to zeroing-out a column from the weight matrix associated with a convolutional layer, but this does not produce the same effect as that in the case of the fully-connected layer. Zeroing-out a column of the weight matrix corresponding to the convolutional kernel still allows the weights of that column to be trained. The effect, while measurable, is not interpretable in a straightforward way. While dropout applied to convolutional layers has shown some benefits, it is not interpretable in a theoretically straightforward way [367]. There are methods, such as dropblock, under investigation to attempt to develop a form of dropout for convolutional layers that is conceptually and theoretically consistent with dropout for fully-connected layers [366].

8.1.6.9 Ensembles: model averaging, or combining the predictions of multiple models

An ensemble is a collection of a number of models that are trained or optimised for the same task, such as classification of signal and background. A combination of many different classification models can often improve classifications [368].

Training deep neural networks is computationally expensive. Using contemporary computing equipment, training on millions of training cases can take days, weeks and months, and after such investment of time and resources, there is little guarantee that the final model will have low generalisation error, performing well on test cases not used during training. Neural network models are nonlinear methods, which means that they can learn complex, nonlinear relationships in data. A downside of this flexibility is that they are

chaotic and sensitive to initial conditions, both in terms of initialised weights and in terms of statistical fluctuations in training data. This stochastic nature means that each time a neural network model is trained, it may learn a different version of the function that maps inputs to outputs and that will have different performance on training and validation data (assuming a constant seed is not used to initialise weights). So, neural network models tend to have a low bias and a high variance. One approach to addressing the high variance of neural network models is to train multiple models and then to combine their predictions. This is usually the combination of predictions from multiple good but different models. One reason why this model averaging works appears to be that different models usually do not make all the same errors on the validation data [369].

Combining the predictions of multiple neural networks as an ensemble adds a bias that counters the variance of individual models. The results are predictions that are less sensitive to the specifics of the training data, the choice of training scheme and the serendipity of a single instance of training. The ensemble can result also in better predictions than any individual best model. Ensemble learning involves training mode than one model on the same data and then using each of the trained models to make a prediction before combining the predictions in some way to make a final overall prediction. The ImageNet research used model averaging of multiple well-performing convolutional neural network models in order to achieve state-of-the-art results for image classifications in 2012 [364]. The ensembling approach, referred to as a "committee" of models, involves calculating the overall prediction as the average of the predictions of the individual trained models [295]. There are more complicated approaches that involve weighting the predictions of the individual trained models, some of which were tested in the efforts of this thesis, sometimes called model blending, and this can be done by training a separate model to learn how best to combine the predictions from each ensemble member model [370].

It has been observed that some ensembles of models benefit from forms of over-fitting [368]. This possibility was tested in modelling for the cases of the SUSY dataset and the $t\bar{t}H\left(b\bar{b}\right)/t\bar{t}b\bar{b}$ dataset cases but no such benefit was observed reliably.

8.1.6.10 Interpretation of modelling: activation maximisation and saliency

The techniques of activation maximisation and saliency are used in this thesis to provide interpretations of modelling achieved [371, 372, 373, 374, 375].

Activation maximisation aims to maximise the activation of certain neurons [373]. Training a neural network involves changing the neural network weights to achieve minimal loss, so the input and desired output are kept constant while the weights are modified iteratively until some minimum is reached. Activation maximisation involves keeping the weights and the desired output constant while the input is modified such that it maximises the activation of certain neurons. The calculation of the activation

maximisation starts with the creation of a pseudorandom input. The trained neural network is applied to this input and the mean activation of a layer is calculated and from this are computed the gradients for the input feature values. The gradients are used to update the feature values in a way that maximises the mean activation. Activation maximisation elucidates what has been modelled by a neural network by synthesising data that corresponds not only to maximisations of classes such as signal and background, but also to maximisations of various degrees of confidence in classifications, for example, synthesising and visualising inputs the model might find maximally confusing. While a promising technique, there remains a lack of theoretical clarity on the degrees to which synthetic model-generated events may be compared fairly with qualified Monte Carlo data, and also a lack of clarity on fair comparisons in terms of modelling uncertainties.

Saliency maps calculate the effect of every feature of an input on the output of a model [374]. This involves calculating the gradient of the output with respect to every feature of the input. This elucidates how the output category value changes with respect to small changes in the input features. Positive gradient values imply that a small change to a feature shall increase the output value. Visualisation of these gradients can highlight those input features that most contribute to the output classification. A saliency map gives insight into how output classifications change with respect to small changes in the input features. It is possible to perform the associated calculation with guided back-propagation, which truncates all negative gradients in the calculation to zero in order to ensure that only features that have a positive influence on the classification probability are updated in the calculation [375]. It is this approach that was taken in this thesis. Visualising the saliency – the gradients of the output classification with respect to input features – can provide some insight into the attention given by a model to each of the features of the input. Saliency can be applied effectively in a visual context to convolutional layers to interpret filters modelled by these layers. Saliency is a tool which may be used to attempt to explain decisions made by a classification model. It is a score for each feature in data providing some indication of how relevant the feature is for explaining a classification. For saliency in this thesis, a statistical uncertainty has been quantified by measuring the variation in saliency results over a number of events. Propagation of systematic uncertainties associated with event variables through the classification process to the saliency interpretation of the model would be desirable but is outside the scope of this investigation.

8.2 Comparisons and interpretations of models for classification of $t\bar{t}H\left(b\bar{b}\right)$ signal and $t\bar{t}b\bar{b}$ background

This section presents interpretations of the problem of classification of signal $t\bar{t}H\left(b\bar{b}\right)$ and main background $t\bar{t}b\bar{b}$ events in the semileptonic $t\bar{t}H\left(b\bar{b}\right)$ resolved analysis. It presents a variety of solutions to the problem, in the form of various multivariate methods, and a number of interpretations and visualisations of the multivariate modelling solutions. The various multivariate methods and techniques of interpretation and visualisation of modelling are described in Section 8.1. Results were derived from analysis of Monte Carlo samples corresponding to an electron + jets events selection with ≥ 6 jets and ≥ 4 b-tagged jets.

In this section, first the features of the data are described, with visualisations of inherent correlations and with both t-SNE and UMAP dimensionality reduction interpretations. Then, the performances of various multivariate methods are described and compared, together with the architectures of the methods. Interpretations of the modelling of some of the multivariate methods are presented and compared.

Given that the standard multivariate approach based on decision trees used in the analysis is effective and computable in a practical time, the constraint of no more than a few hours of hyperparameter scanning and training was applied to the computation of the various models for reasons of ensuring comparable practicality.

8.2.1 The $t\bar{t}H\left(b\bar{b}\right)$ dataset and its features

Various features of inputs to the typical analysis classification BDT (the features can in practice vary, as can their precise definitions) are listed in Tables 8.1, 8.2 and 8.3. A selection of these variables was used for the purposes of comparisons of the various classification models described in this chapter and for interpretations of the modelling. The initial treatment of the data was similar to that applied to the SUSY dataset, which was used as a benchmark and validation for the various multivariate methods and novel interpretation techniques [309, 20]. Approximately the same number of signal events as background events was used for such expressions of classification efficiency as separation plots. This is distinct from plots derived from recorded data, in which it could be usual to see signal events normalised by some factor in order to be visible in plots, such as is the case in Figure 7.1.

A great deal of effort, from theoretical definition to analytical testing, goes into these physics features. Some of the details of those efforts and the resulting variables are described in References [256] and [1].

Variable	Description
$\Delta R_{bb}^{\text{avg}}$	Average ΔR for all b-tagged jet pairs
$\Delta R_{bb}^{\max p_{\text{T}}}$	ΔR between the two b-tagged jets with the largest vector sum p_{T}
$\Delta\eta_{jj}^{\max \Delta\eta}$	Maximum $\Delta\eta$ between any two jets
$m_{bb}^{\min \Delta R}$	Mass of the combination of the two b-tagged jets with the smallest ΔR
$m_{jj}^{\min \Delta R}$	Mass of the combination of any two jets with the smallest ΔR
$p_T^{\text{jet 5}}$	p_{T} of the fifth leading jet
N_{30}^{H}	Number of b-jet pairs with invariant mass within 30 GeV of the Higgs boson mass
$H_{\text{T}}^{\text{had}}$	Scalar sum of jet p_{T}
$\Delta R_{\text{lep}-bb}^{\min \Delta R}$	ΔR between the lepton and the combination of the two b-tagged jets with the smallest ΔR
aplanarity	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor built with all jets
centrality	the scalar sum of the p_{T} divided by the sum of the E for all jets and the lepton
$H1$	Second Fox-Wolfram moment computed using all jets and the lepton

Table 8.1: General kinematic input variables to the classification BDT in the semileptonic $t\bar{t}H$ $(b\bar{b})$ channel [256].

Variable	Description
BDT	BDT output
m_{H}	Higgs boson mass
$m_{H,b_{\text{lep } t}}$	Mass of Higgs boson and b-jet from leptonic t-quark
$\Delta R_{H\ bb}$	ΔR between b-jets from the Higgs boson
$\Delta R_{H,t\bar{t}}$	ΔR between Higgs boson and $t\bar{t}$ system
$\Delta R_{H,\text{lep } t}$	ΔR between Higgs boson and leptonic t-quark
$\Delta R_{H,b_{\text{had } t}}$	ΔR between Higgs boson and b-jet from hadronic t-quark

Table 8.2: Input variables to the classification BDT from the reconstruction BDT output in the semileptonic $t\bar{t}H\left(b\bar{b}\right)$ channel channel [256].

Variable	Description
D	Likelihood discriminant

Table 8.3: Input variables to the classification BDT from likelihood calculation in the semileptonic $t\bar{t}H\left(b\bar{b}\right)$ channel channel [256].

Variable	Description
w_{b}^{H}	Sum of binned b-tagging weights of jets
B_{j^3}	Third jet binned b-tagging weight (sorted by weight)
B_{j^4}	Fourth jet binned b-tagging weight (sorted by weight)
B_{j^5}	Fifth jet binned b-tagging weight (sorted by weight)

Table 8.4: Input variables to the classification BDT from b-tagging in the semileptonic channel [256].

A hierarchically-clustered correlations map can be made of the data. Hierarchical clustering in this context is an agglomerative ("bottom-up") clustering method that starts with each element viewed as a separate cluster, and then merges elements successively into larger clusters [376, 377]. Specifically, in each step of hierarchical clustering, the closest pair of clusters is identified and these are merged into a new

parent cluster. This is repeated until only one cluster is formed after $N-1$ iterations, where N is the number of objects. The result of hierarchical clustering in this way is described with a tree structure plot referred to as a dendrogram. A hierarchically-clustered correlations map, with associated dendrogram representations, of the various input features is shown in Figure 8.8.

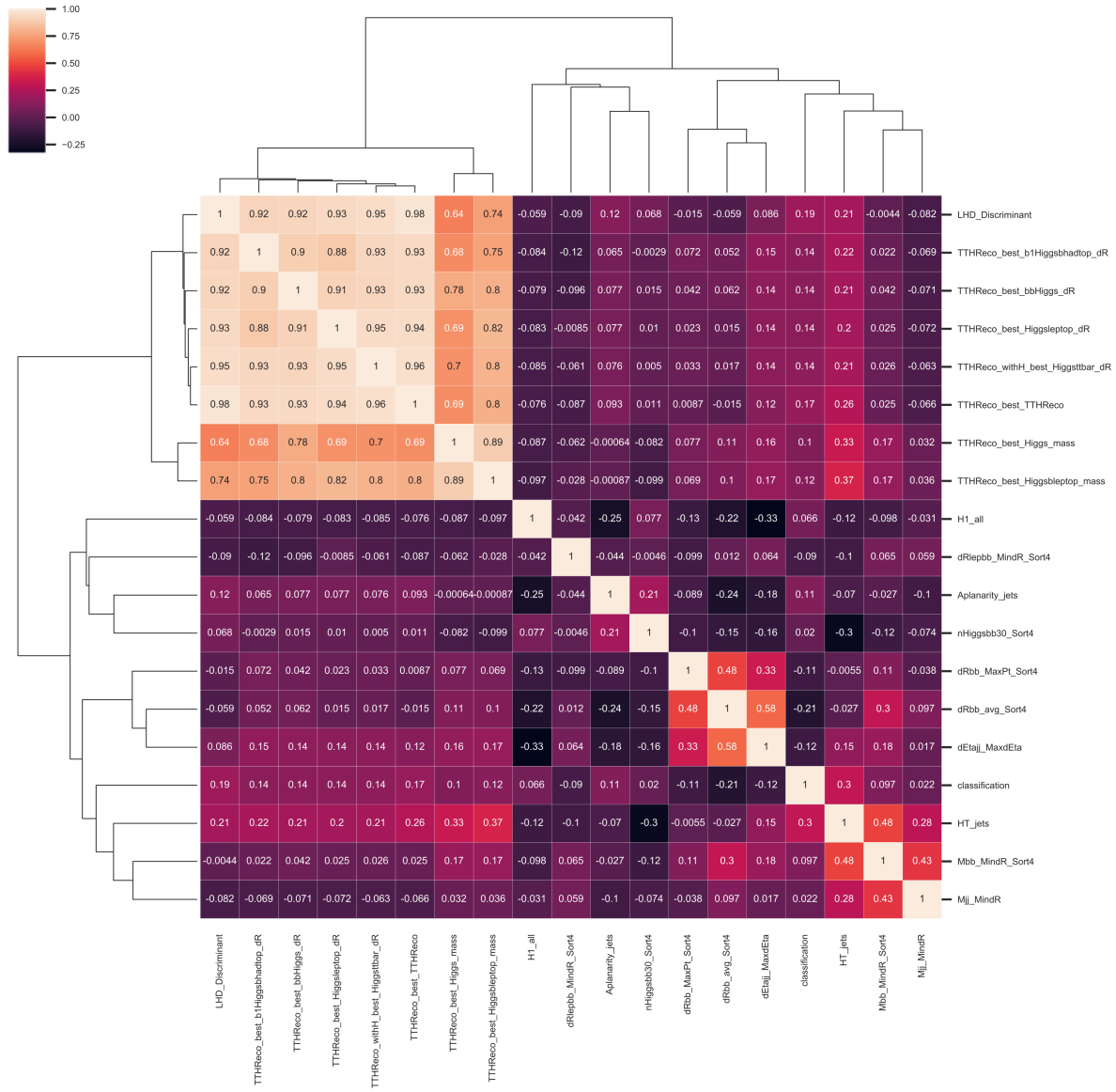


Figure 8.8: Model training variables shown in a hierarchically-clustered correlations map. The colours show the correlations of the variables while the marginal lines are a dendrogram representation of a hierarchical clustering analysis of variable correlations.

8.2.2 Dimensionality reduction: t-SNE and UMAP

A subset of the data was projected down to two dimensions using the t-SNE algorithm for the purposes of Figure 8.9, with signal coloured red and background coloured blue. It can be seen that the t-SNE analysis is identifying statistical lobes in the data (five), and some regions of each lobe can be seen to be more signal-rich and more background-rich. One could imagine a hyperplane calculated to separate these regions for the purposes of selecting analysis regions that are rich in signal or rich in background.

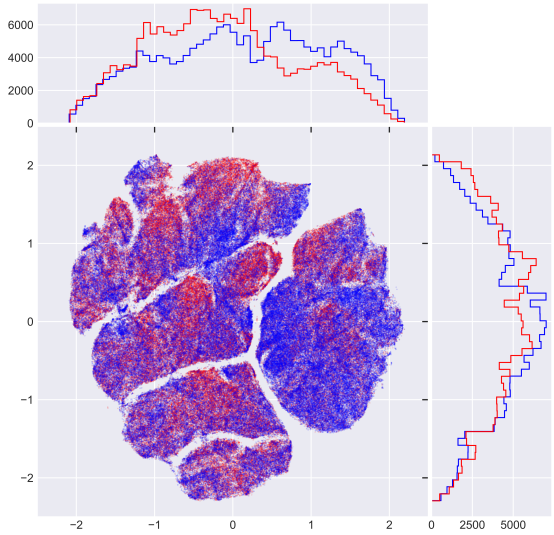


Figure 8.9: t-SNE (perplexity 30) dimensionality reduction applied to data used for $t\bar{t}H\left(b\bar{b}\right)/t\bar{t}b\bar{b}$ classification. Red represents signal labels and blue represents background labels.

Some features (selected with reference to mean saliency maps for signal and background derived from deep multi-layer perceptron modelling, described later) are shown on t-SNE dimensionality reduction projections in Figure 8.10, where the intensity of each data point corresponds to the magnitude of the feature for that data point. The same statistical lobes can be seen in the data as shown in Figure 8.9 and it can be seen that certain regions of each lobe appear to correspond to stronger or weaker values of the features shown. For example, high-magnitude values of the second Fox-Wolfram moment feature $H1$ can be seen to correspond to a vertically-oriented band in every separate statistical lobe of the s-SNE projection.

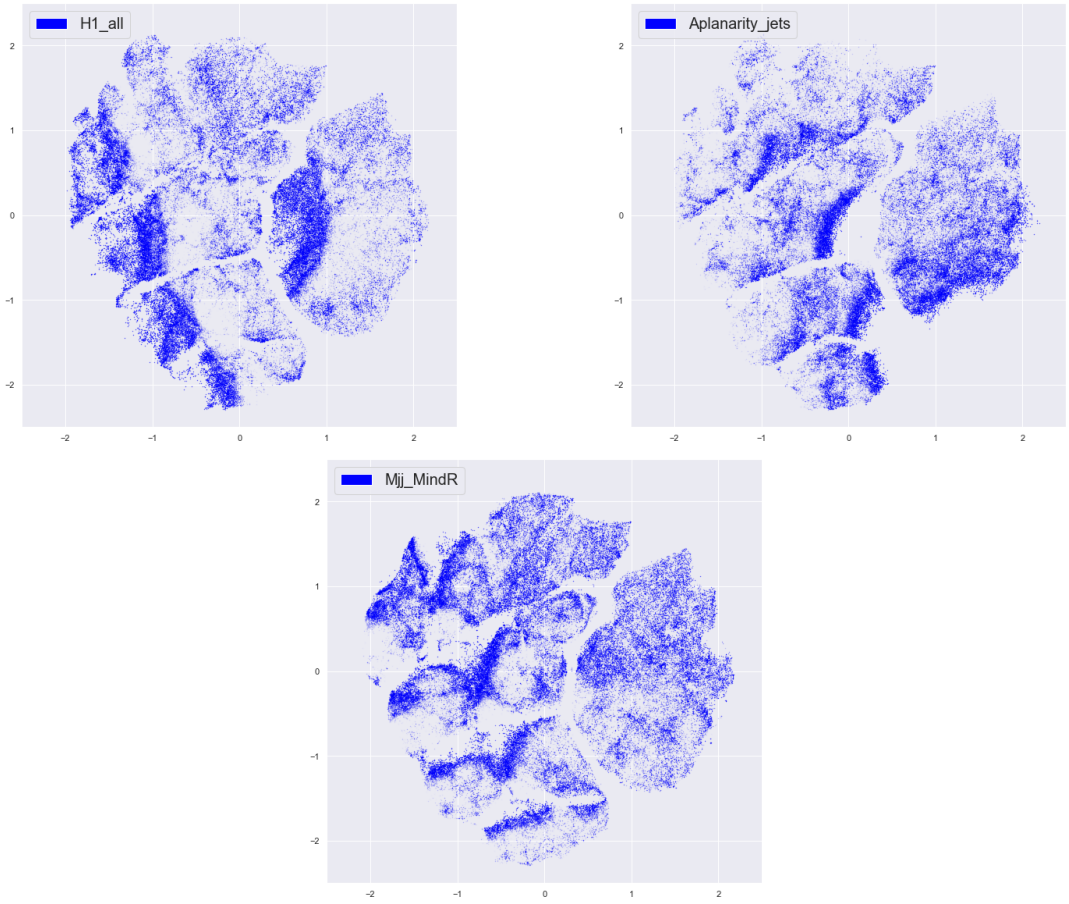


Figure 8.10: t-SNE (perplexity 30) dimensionality reduction applied to data used for $t\bar{t}H\left(b\bar{b}\right)/t\bar{t}b\bar{b}$ classification with some data features shown such that colour intensity represents the value of the feature.

An unguided UMAP projection was made of the data using the same hyperparameters and the optimisation approach described in Section 8.1.3. The resulting optimised UMAP projection is shown in Figure 8.11. This projection also identifies separate statistical lobes in the data (five, similar to the t-SNE projection).

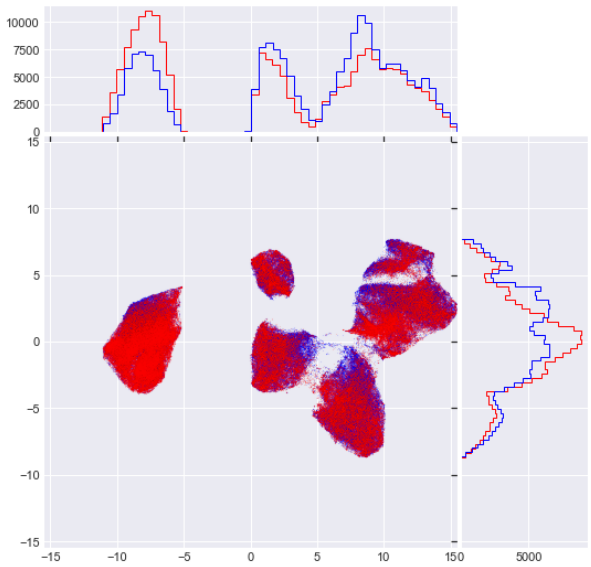


Figure 8.11: The data used for $t\bar{t}H\left(b\bar{b}\right)/t\bar{t}b\bar{b}$ classification transformed using UMAP (nearest neighbours 6), corresponding to a projection of 18 dimensions down to two. Red represents signal labels and blue represents background labels. The Kolmogorov-Smirnov statistic for two samples (signal and background) is 0.15564 for the horizontal projection and 0.10238 for the vertical projection.

In a plot similar to that shown in Figure 8.8, the unguided UMAP projection is shown together with scatter plots of the various training variables hierarchically clustered by their correlations in Figure 8.12.

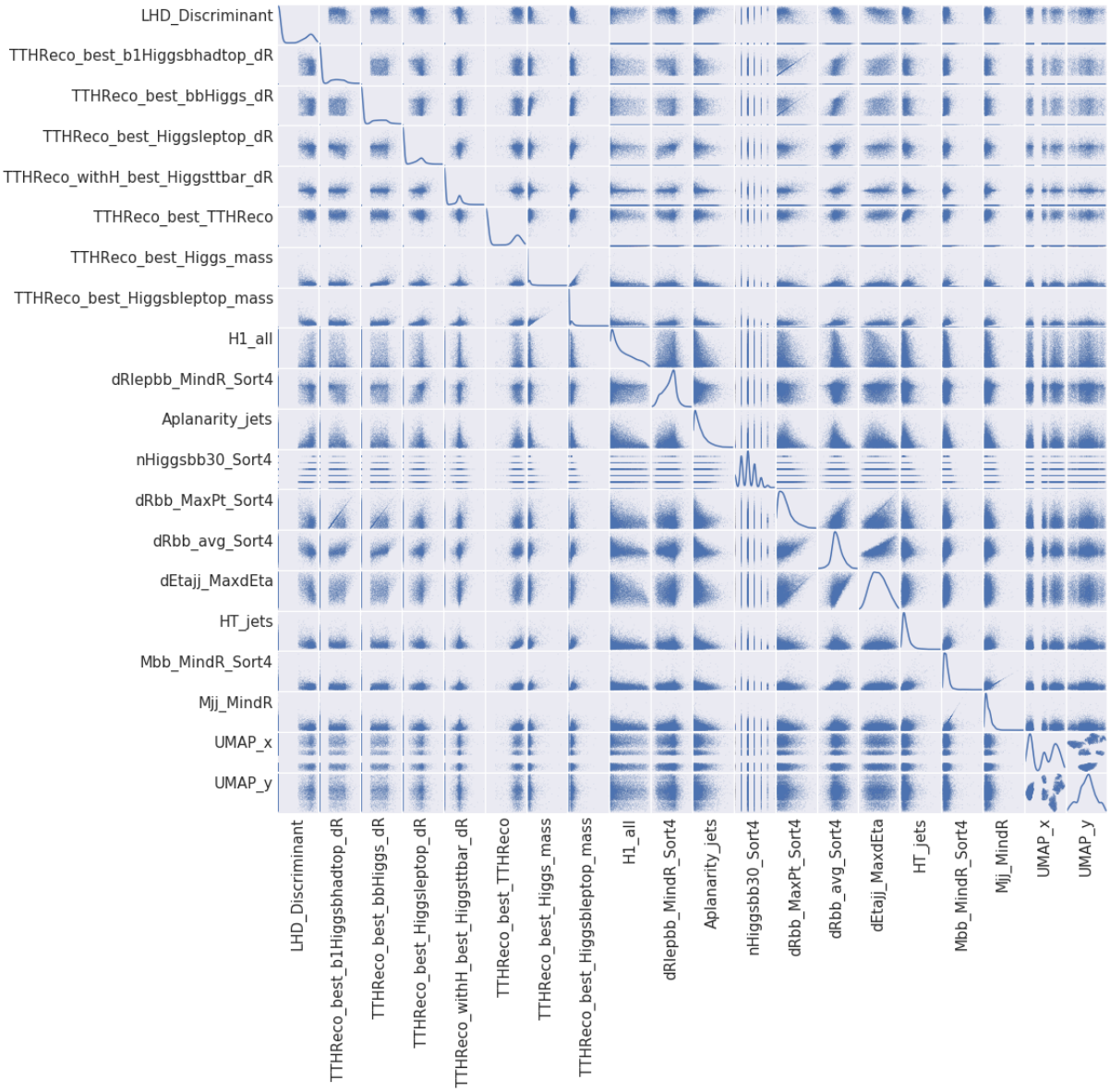


Figure 8.12: Model training variables and unguided UMAP projection variables shown in a hierarchically-clustered correlations map of scatter plots.

A guided UMAP projection was made of the data again using the same hyperparameters and the optimisation approach described in Section 8.1.3, with the algorithm supplied with class label information and optimisation by the Kolmogorov-Smirnov approach halted before visible structure within clusters

was obviously decreased. The resulting optimised guided UMAP projection is shown in Figure 8.13. The internal structure of the individual signal and background classes has been broadly retained, as has the global structure, while the individual classes have been cleanly separated from one another, with the inter-relationships between the classes preserved. It can be seen that the number of statistical lobes visible has approximately doubled (to ten), corresponding to the inclusion of information on the two classes.

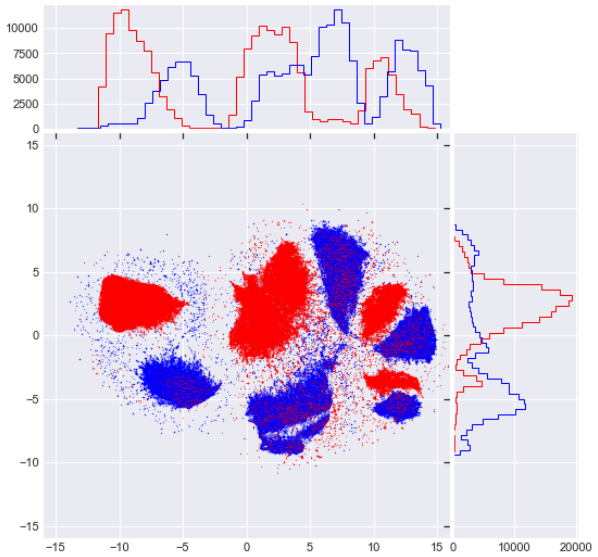


Figure 8.13: The data used for $t\bar{t}H\left(b\bar{b}\right)/t\bar{t}b\bar{b}$ classification transformed using guided UMAP (nearest neighbours 6), corresponding to a projection of 18 dimensions down to two. Red represents signal labels and blue represents background labels. The Kolmogorov-Smirnov statistic for two samples (signal and background) is 0.35794 for the horizontal projection and 0.518 for the vertical projection.

An unguided UMAP projection was made of the data again using the same hyperparameters and the same optimisation approach, but this time the UMAP algorithm was used to project from 18 dimensions down to one, resulting in a histogram. The resulting optimised unguided UMAP projection is shown in Figure 8.14.

Figure 8.14: The data used for $t\bar{t}H\left(b\bar{b}\right)/t\bar{t}b\bar{b}$ classification transformed using UMAP (nearest neighbours 145), corresponding to a projection of 18 dimensions down to one. Red represents signal labels and blue represents background labels. The Kolmogorov-Smirnov statistic for two samples (signal and background) is 0.15653.

UMAP was used to make a projection from 18 dimensions down to three, resulting in a three-dimensional scatter plot, an unguided example of which is shown in Figure 8.15. While viewed as a rotatable animation, the plot is more interpretable, but on printed paper it is not obviously more interpretable than the two-dimensional plot of Figure 8.11.

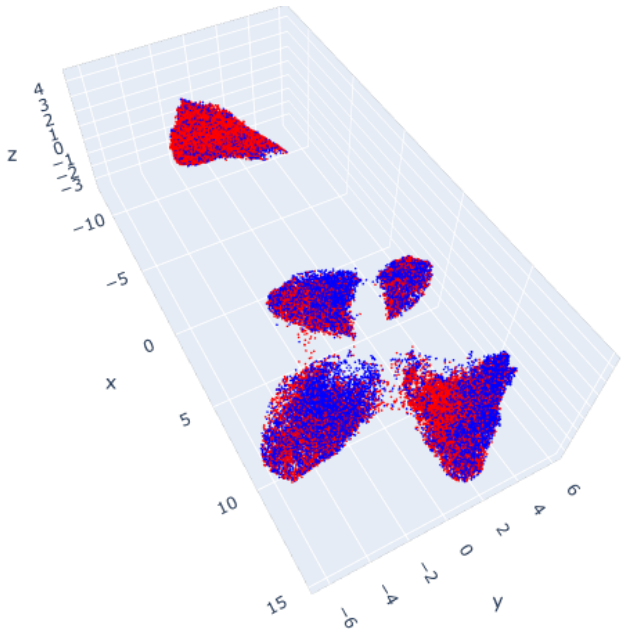
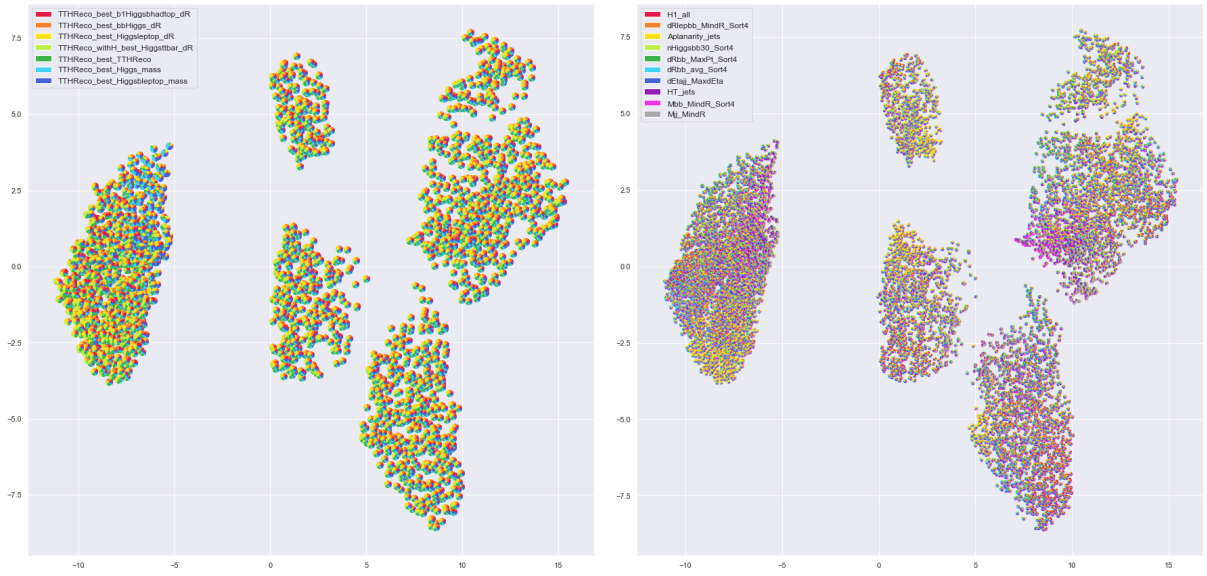


Figure 8.15: The data used for $t\bar{t}H\left(b\bar{b}\right)/t\bar{t}b\bar{b}$ classification transformed using guided UMAP (nearest neighbours 6), corresponding to a projection of 18 dimensions down to three. Red represents signal labels and blue represents background labels.

Following the technique outlined in Section 8.1.3, UMAP projections were made in which each data point is plotted as a pie chart of the relative magnitudes of the features of that data point, as shown in Figure 8.16.



(a) Features from reconstruction BDT.

(b) Features not from reconstruction BDT.

Figure 8.16: The data used for $t\bar{t}H\left(b\bar{b}\right)/t\bar{t}b\bar{b}$ classification transformed using UMAP (nearest neighbours 6), corresponding to a projection of 18 dimensions down to two. Pie charts at each point show relative magnitudes of features in the data.

These various presentations of the various dimensionality reduction techniques enable identification of groups of statistically similar subsets within the data and identification of regions within those subsets that correspond to certain characteristics of data features.

8.2.3 Classification models: architectures, results, comparisons, interpretations

Various multivariate modelling techniques for the purposes of classification of $t\bar{t}H\left(b\bar{b}\right)$ and $t\bar{t}b\bar{b}$ events were implemented, tested and compared, and some techniques for interpretation of the modelling were applied.

Following from independent testing on the SUSY dataset involving the inclusion of features from a UMAP projection to two dimensions and optimised with reference to the mean Kolmogorov-Smirnov two-sample test statistic, models were tested both with and without this preprocessing treatment. While the addition of the two features did in some cases result in a slightly more brisk identification of an effective network, after some training the benefits of the addition were not clear. It may be that as training progresses, sufficiently complex models learn what UMAP in isolation learns.

8.2.3.1 Architectures

The architecture of the single decision tree classification model is shown in Figure 8.17 and is representative of the trees of the boosted decision tree (BDT). The architectures of the various neural network classification models – deep multi-layer perceptron (DMLP), convolutional neural network (CNN) and inception convolutional neural network (ICNN) – are shown in Figure 8.18. An ensemble model was constructed that consisted of taking the unweighted mean predictions of the BDT, BDT with UMAP, DMLP, CNN with UMAP, ICNN and ICNN with UMAP models.

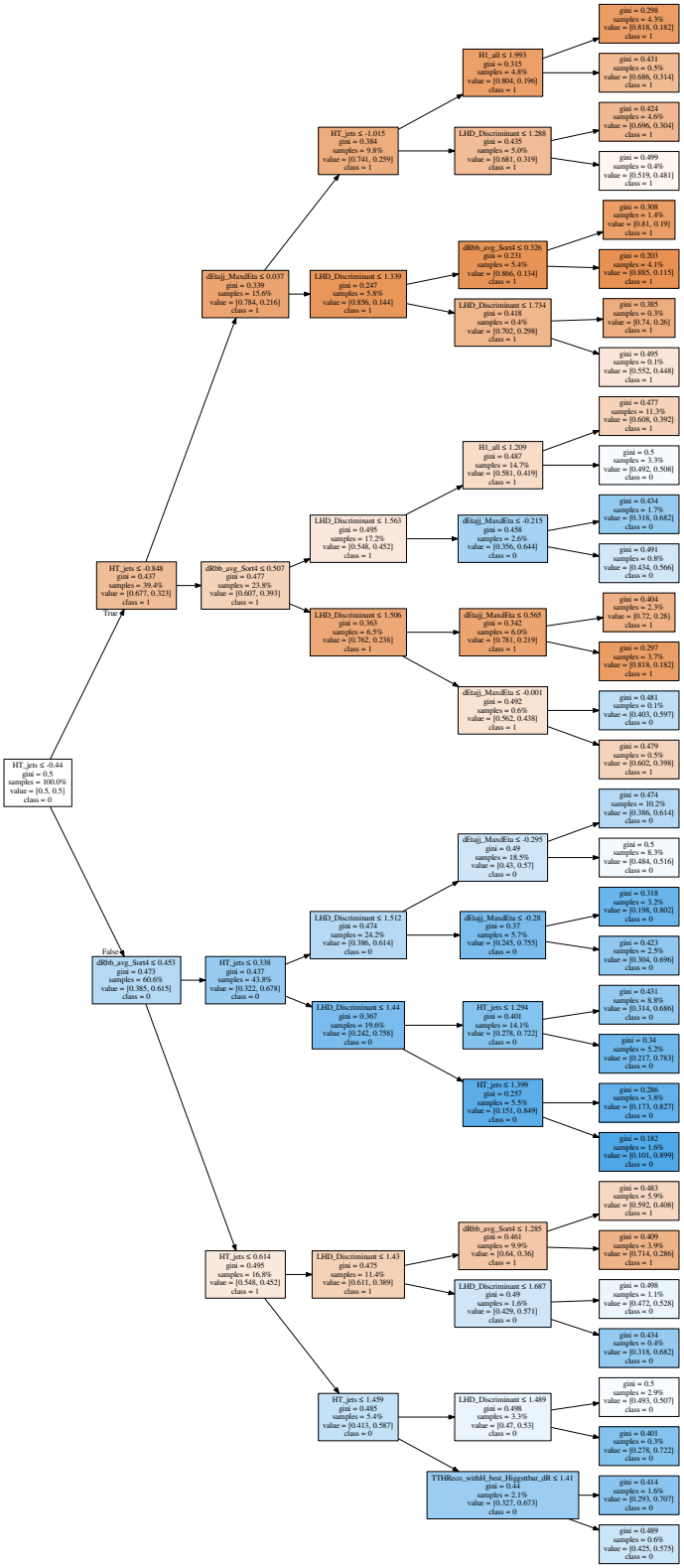
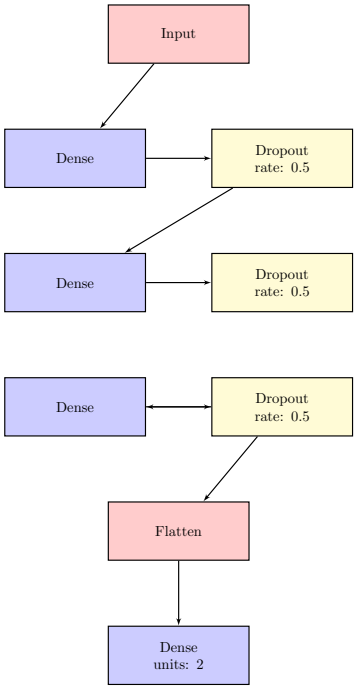
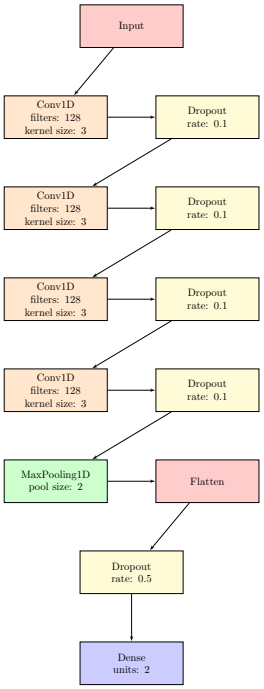


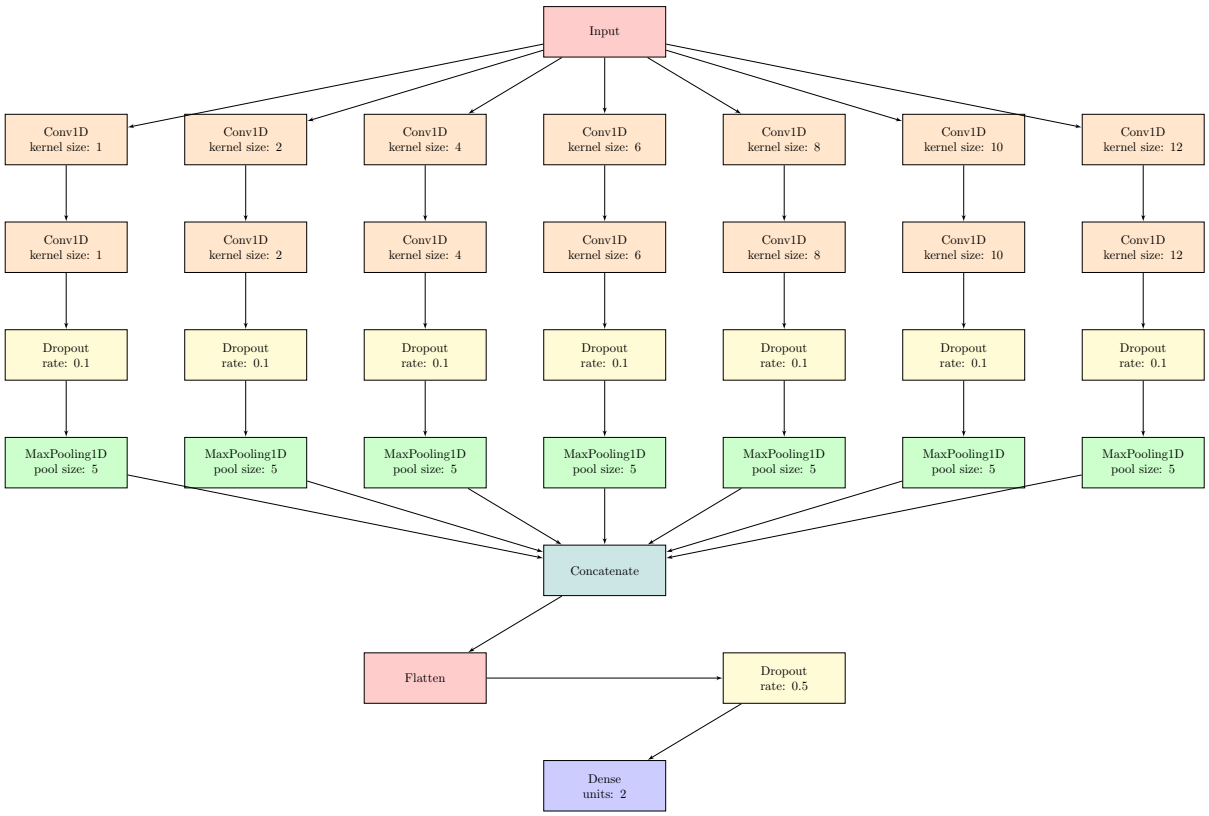
Figure 8.17: Architecture of decision tree classification model.



(a) DMLP.



(b) CNN.

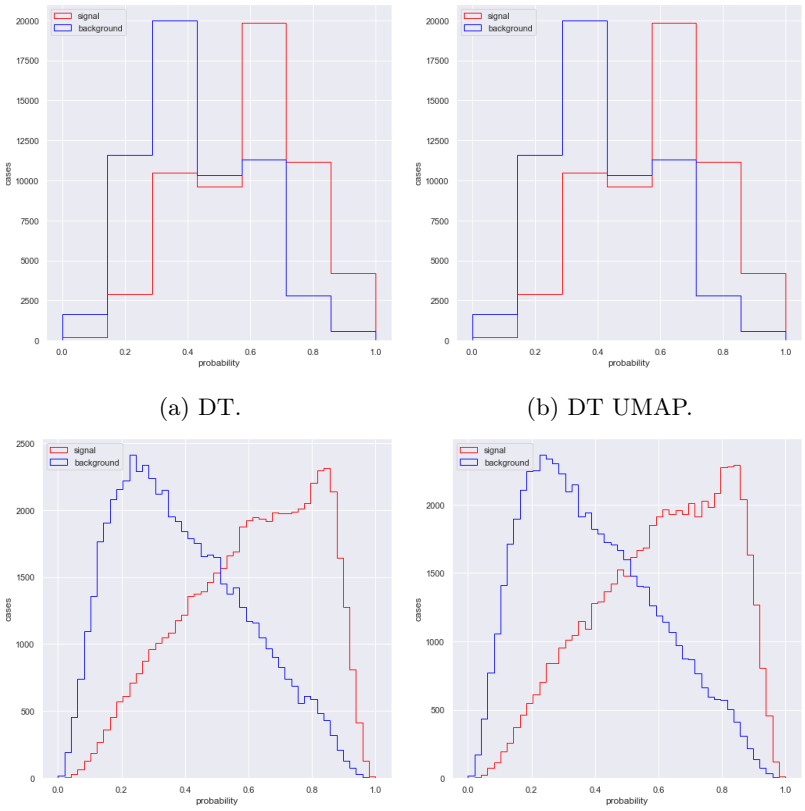


(c) ICNN.

Figure 8.18: Architectures of the various neural network models – deep multi-layer perceptron (DMLP), convolutional neural network (CNN) and inception convolutional neural network (ICNN).

8.2.3.2 Results and comparisons

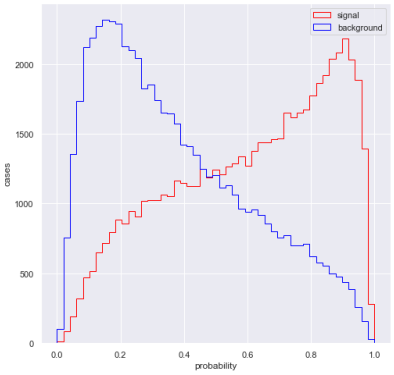
The results obtained for the various models, including the ensemble model, are shown as separation plots in Figures 8.19 and 8.20 and as class prediction error plots, shown with class efficiency ratios, in Figures 8.25 and 8.26. The models are compared using ROC curves in Figure 8.27 and using a variety of scoring metrics (described in Section 8.1.6.4) in Table 8.5.



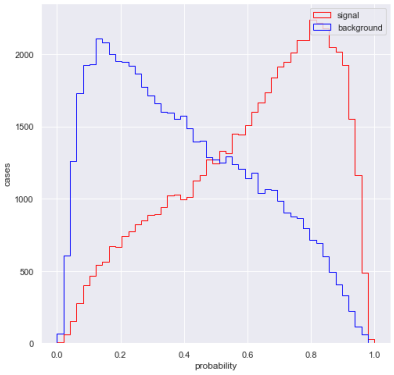
(a) DT. (b) DT UMAP.

(c) BDT. (d) BDT UMAP.

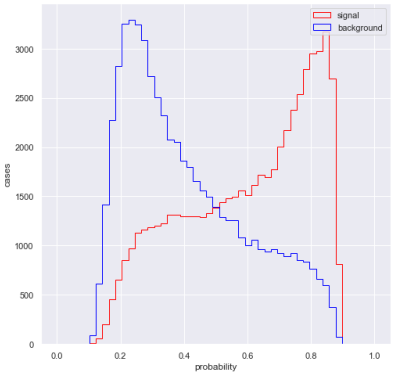
Figure 8.19: Separation plots for signal $t\bar{t}H\left(b\bar{b}\right)$ (red) and background $t\bar{t}b\bar{b}$ (blue) for the various multivariate methods. Note that it is not possible to calculate exact probabilities for various classifications of a decision tree, so instead the predicted class 'probability' is the fraction of samples of the same class in a leaf.



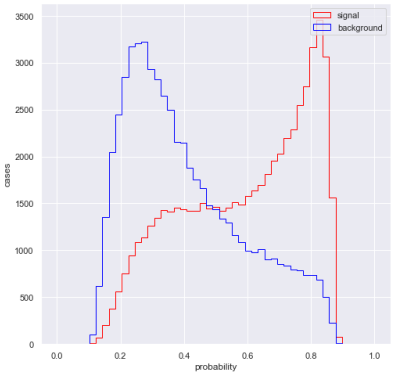
(a) DMLP.



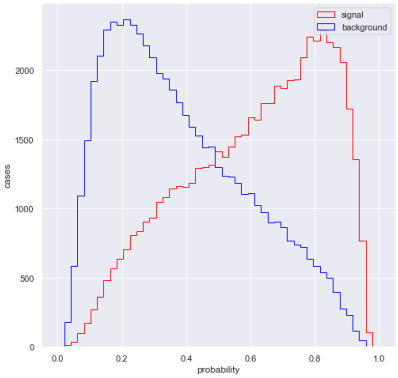
(b) DMLP UMAP.



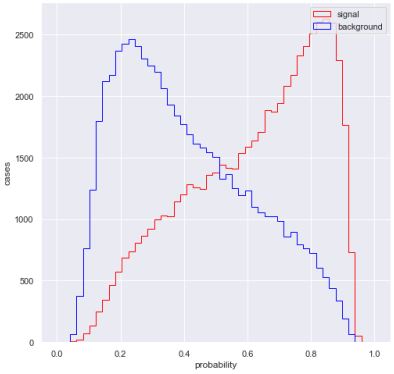
(c) CNN.



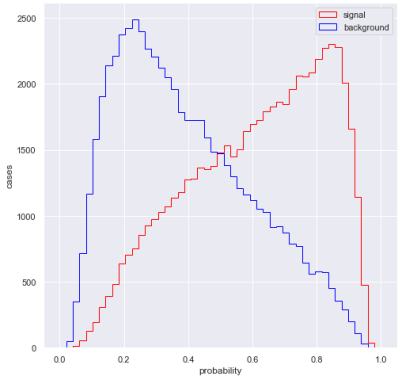
(d) CNN UMAP.



(e) ICNN.



(f) ICNN UMAP.



(g) Ensemble.

Figure 8.20: Separation plots for signal $t\bar{t}H\left(b\bar{b}\right)$ (red) and background $t\bar{t}b\bar{b}$ (blue) for the various multivariate methods. Note that it is not possible to calculate exact probabilities for various classifications of a decision tree, so instead the predicted class 'probability' is the fraction of samples of the same class in a leaf.

The classification efficiency of the BDT with a progressively increasing number of estimators was considered. The increasing classification efficiency can be seen in the form of the scores of accuracy and precision increasing with the number of estimators, as shown in Figure 8.21, and in the form of ROC curves in Figure 8.22. Separation plots for a progressively increasing number of estimators are shown in two dimensions in Figure 8.23 and in three dimensions in Figure 8.24.

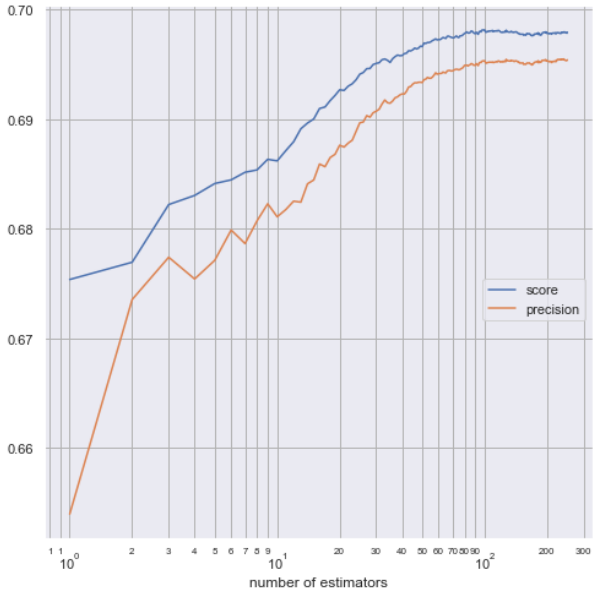
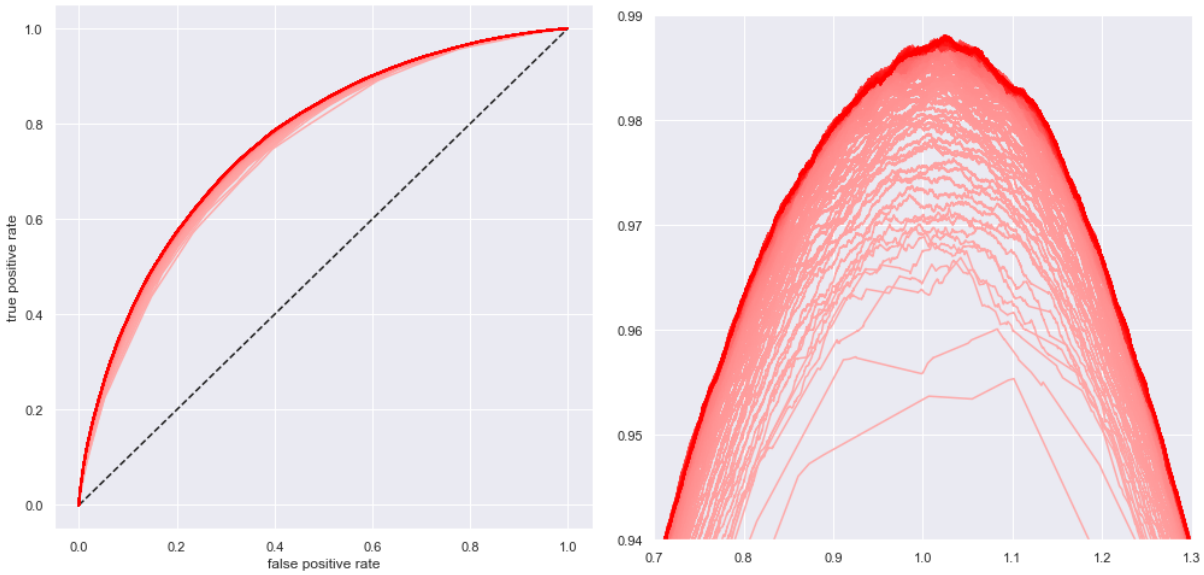


Figure 8.21: Scores of accuracy and precision for a progressively increasing number of estimators in the BDT.



(a) ROC curves.

(b) ROC curves rotated clockwise and cropped.

Figure 8.22: ROC curves for a progressively increasing number of estimators in the BDT, as indicated by increasing intensity of the colour red. Figure 8.22b is ROC curves rotated -45° about the point $(1, 0)$ to aid in visualisation of distinctions in the behaviour of models.

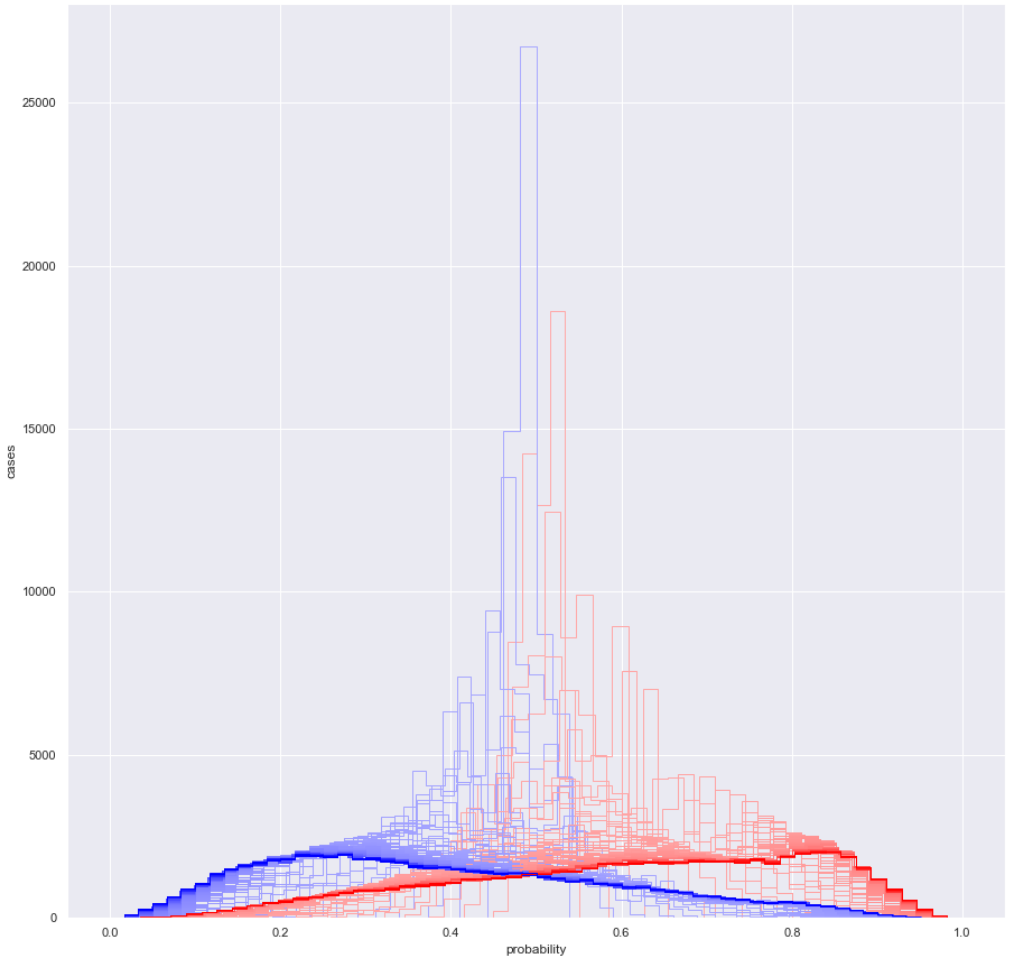


Figure 8.23: Separation plots for signal $t\bar{t}H\left(b\bar{b}\right)$ (red) and background $t\bar{t}b\bar{b}$ (blue) for a progressively increasing number of estimators in the BDT. The increasing number of estimators in the BDT is indicated by increasing intensity of the colours red and blue. So, a pale red colour and a pale blue colour indicates a separation plot corresponding to a smaller number of estimators (and generally lesser separation) while a dark red colour and a dark blue colour indicates a separation plot corresponding to a larger number of estimators (and generally greater separation). Binning used for histograms here was automated using the Freedman-Diaconis rule [378]. Note that it is not possible to calculate exact probabilities for various classifications of a decision tree, so instead the predicted class 'probability' is the fraction of samples of the same class in a leaf.

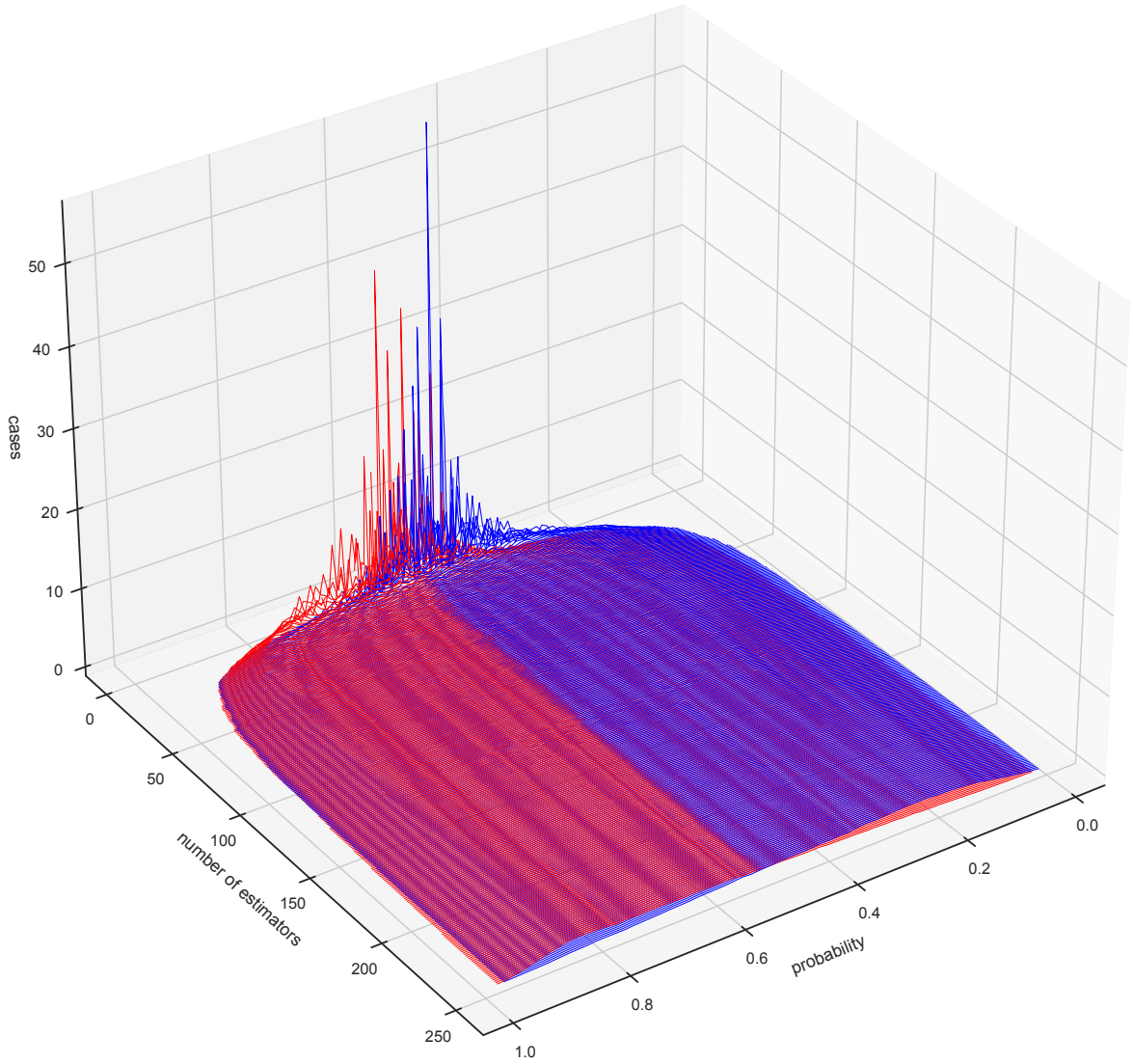
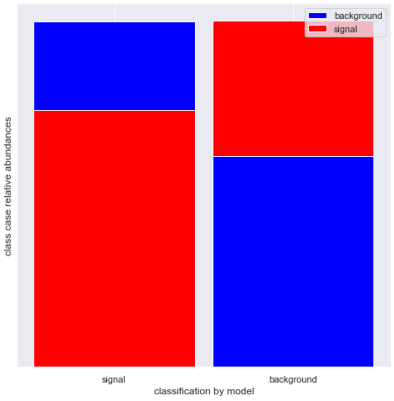
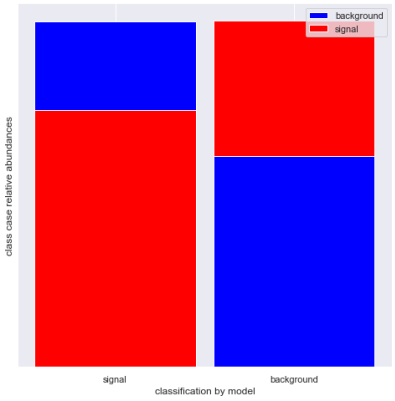


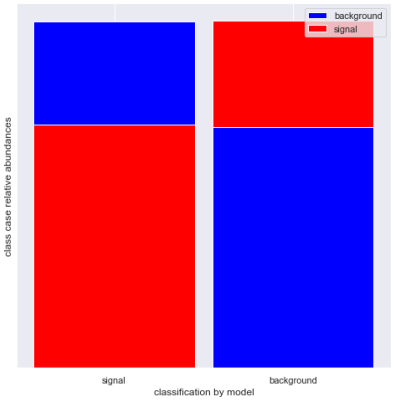
Figure 8.24: Separation plots for signal $t\bar{t}H\left(b\bar{b}\right)$ (red) and background $t\bar{t}b\bar{b}$ (blue) for a progressively increasing number of estimators in the BDT. In contrast with Figure 8.23, the increasing number of estimators is indicated not by increasing colour intensity, but by individual separation plots placed along a separate axis indicating the number of estimators. Note that it is not possible to calculate exact probabilities for various classifications of a decision tree, so instead the predicted class 'probability' is the fraction of samples of the same class in a leaf.



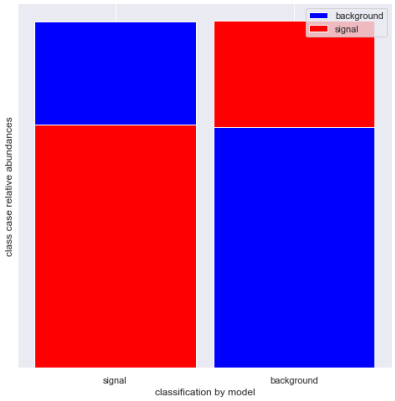
(a) DT. 0.743:0.608.



(b) DT UMAP. 0.743:0.608.

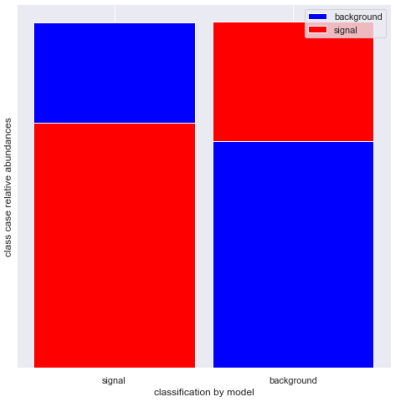


(c) BDT. 0.703:0.693.

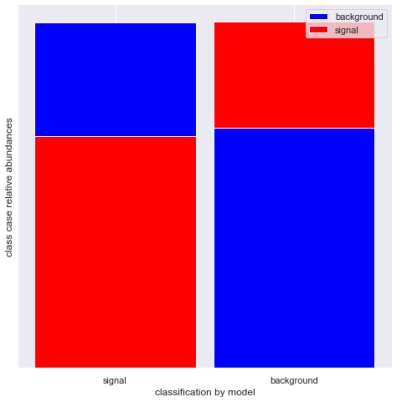


(d) BDT UMAP. 0.702:0.692.

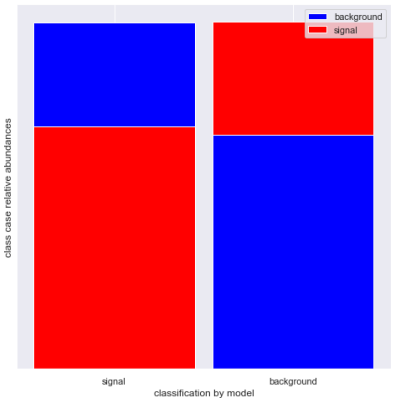
Figure 8.25: Class prediction error plots for signal $t\bar{t}H\left(b\bar{b}\right)$ (red) and background $t\bar{t}b\bar{b}$ (blue) for the various multivariate methods, with class efficiency ratios.



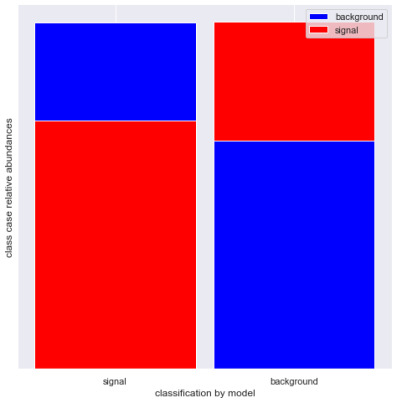
(a) DMLP. 0.709:0.655.



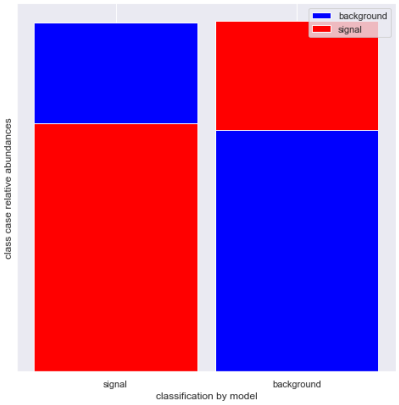
(b) DMLP UMAP. 0.671:0.693.



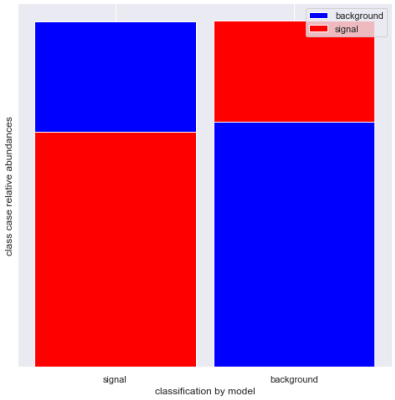
(c) CNN. 0.701:0.675.



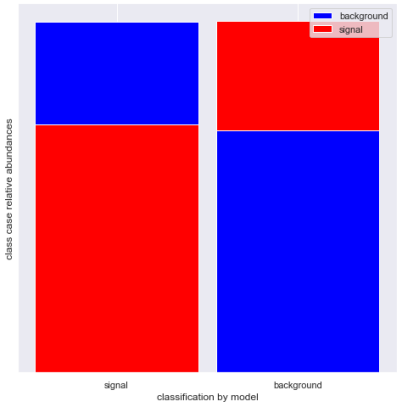
(d) CNN UMAP. 0.716:0.658.



(e) ICNN. 0.689:0.712

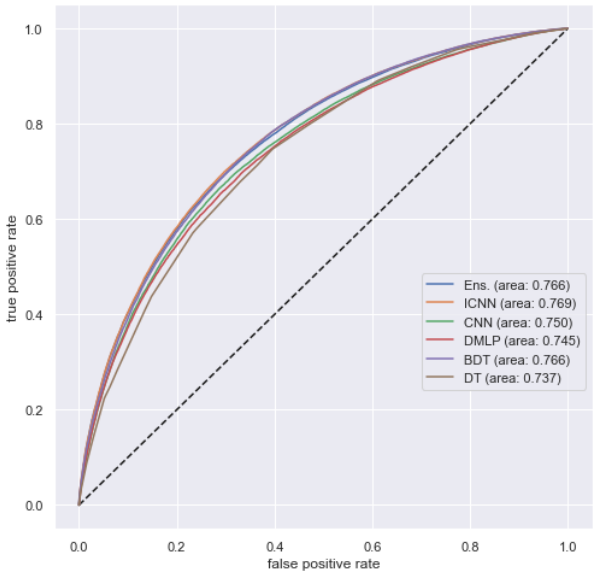


(f) ICNN UMAP. 0.682:0.708.

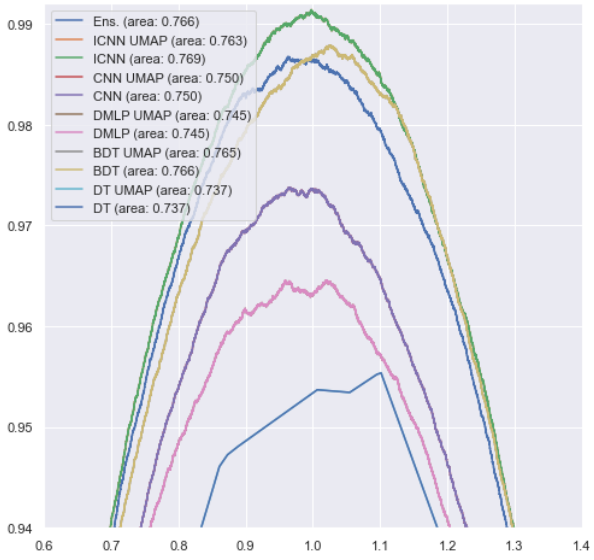


(g) Ensemble. 0.707:0.687.

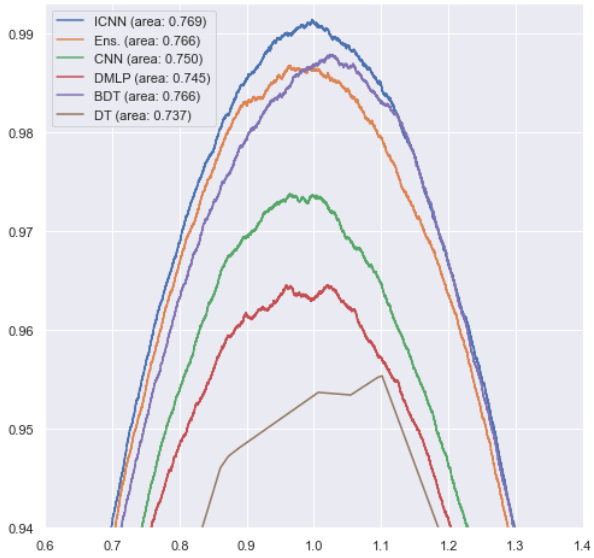
Figure 8.26: Class prediction error plots for signal $t\bar{t}H\left(b\bar{b}\right)$ (red) and background $t\bar{t}b\bar{b}$ (blue) for the various multivariate methods, with class efficiency ratios.



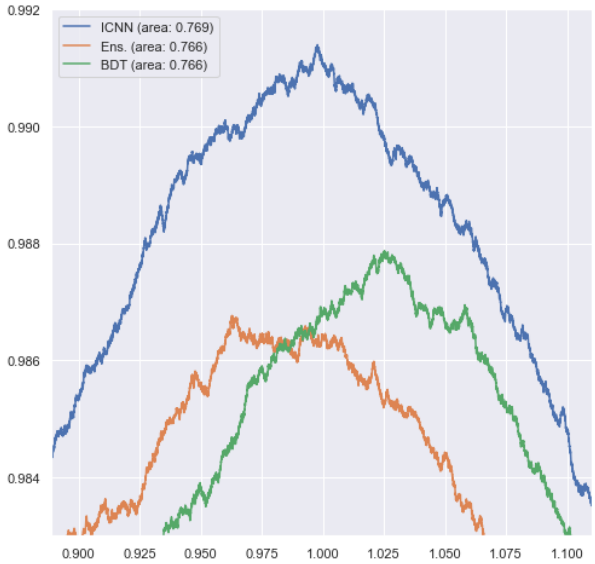
(a) ROC curves.



(b) ROC curves rotated clockwise and cropped.



(c) ROC curves rotated clockwise and cropped.



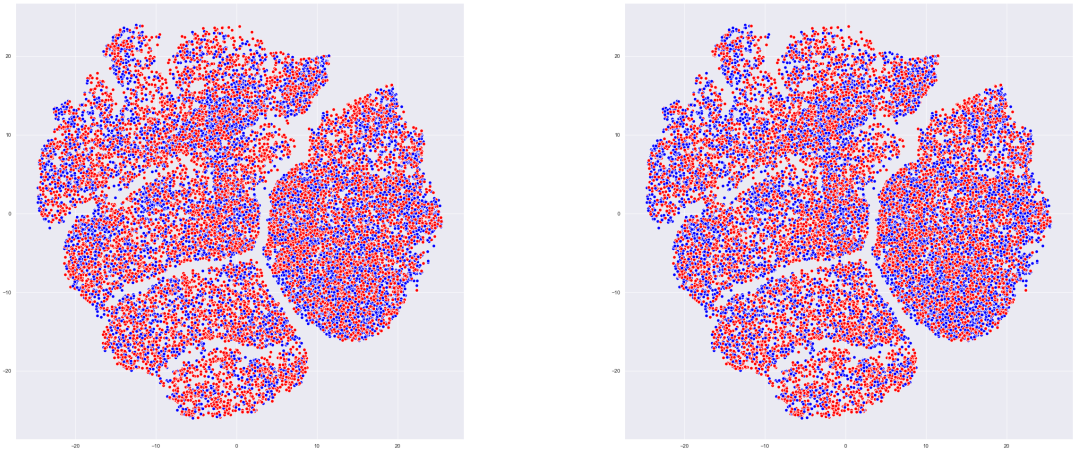
(d) ROC curves rotated and cropped further.

Figure 8.27: Receiver Operating Characteristic (ROC) curves for the various classification models and the ensemble model. Figures 8.27b, 8.27c and 8.27d are ROC curves rotated -45° about the point $(1, 0)$ to aid in visualisation of distinctions in the behaviour of models.

Model	Score	Precision	KS2	AUC
ICNN	0.7006	0.7063	0.4021	0.7689
BDT	0.6979	0.6954	0.3971	0.7655
Ensemble	0.6974	0.7018	0.3955	0.7658
BDT UMAP	0.6973	0.6949	0.3965	0.7655
ICNN UMAP	0.6946	0.6902	0.3904	0.7628
CNN	0.6883	0.6938	0.3772	0.7501
CNN UMAP	0.6870	0.6992	0.3743	0.7495
DMLP	0.6818	0.6928	0.3641	0.7450
DMLP UMAP	0.6816	0.6783	0.3639	0.7449
DT	0.6754	0.6540	0.3511	0.7371
DT UMAP	0.6754	0.6540	0.3511	0.7371

Table 8.5: Comparison of the score, precision, Kolmogorov-Smirnov two-sample test statistic and Area Under the ROC Curve (AUC) of the various models for classification of $t\bar{t}H\left(b\bar{b}\right)$ signal and $t\bar{t}b\bar{b}$ background, with the first, second and third best respective metrics indicated by gold, silver and bronze. Note for the cases of the Kolmogorov-Smirnov two-sample test statistic and the AUC that it is not possible to calculate exact probabilities for various classifications of a decision tree, so instead the predicted class 'probability' is the fraction of samples of the same class in a leaf.

In order to attempt to visualise the consistency of classification decisions made by some models, a subset of the data was projected down to two dimensions using t-SNE for the purposes of Figure 8.28. Figures 8.28a and 8.28b are the same t-SNE projection of data, but have the points classified as signal (red) or background (blue) by the CNN and ICNN models respectively, with incorrectly classified points excluded. It can be observed that the models appear to be identifying largely similar regions of the projected data.



(a) t-SNE of events with CNN classifications.

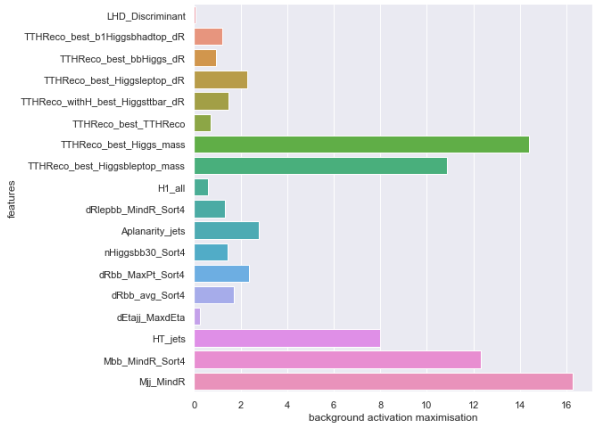
(b) t-SNE of events with ICNN classifications.

Figure 8.28: t-SNE (perplexity 30) dimensionality reduction applied to data used for $t\bar{t}H\left(b\bar{b}\right)/t\bar{t}b\bar{b}$ classification. Red represents signal labels or classifications and blue represents background labels or classifications.

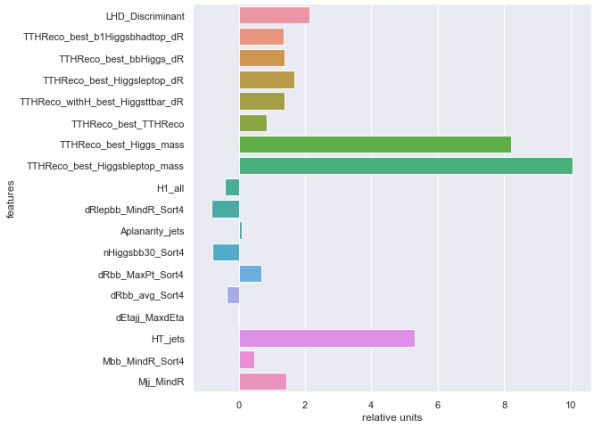
8.2.3.3 Interpretations

8.2.3.3.1 Activation maximisation

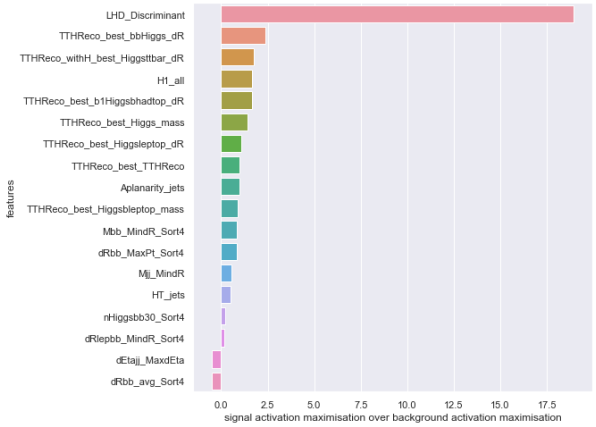
For the case of a deep multi-layer perceptron trained for classification of events of the dataset, a "synthetic event" for maximised classification as background, generated solely from the trained deep multi-layer perceptron model, is shown in Figure 8.29a. This is shown alongside a Monte Carlo background event that was classified as background with high confidence by the trained model in Figure 8.29b. Accepting that there are uncertainties associated with modelling and fluctuations in data, some consistencies can be seen between the model-generated synthetic event and the Monte Carlo event. A synthesised signal event divided by the synthesised background event is shown in Figure 8.29c.



(a) Synthetic event generated from model to maximise classification as background.



(b) Monte Carlo event classified with high confidence by model as background.



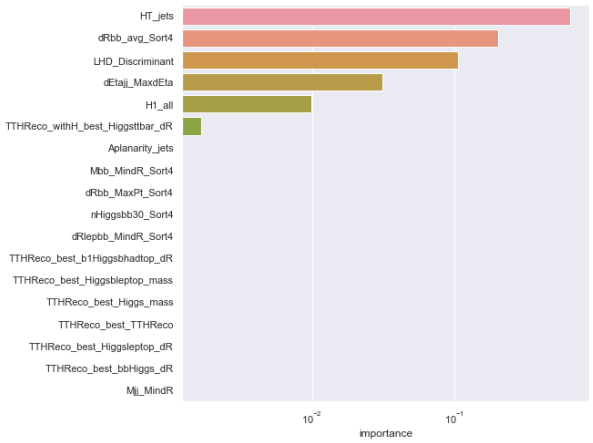
(c) Synthetic event generated from model to maximise classification as signal divided by synthetic event generated from model to maximise classification as background.

Figure 8.29: Comparison of synthetic background event generated from trained network with Monte Carlo background event classified with high confidence as background, and synthetic signal event divided by synthetic background event. Units are standardised.

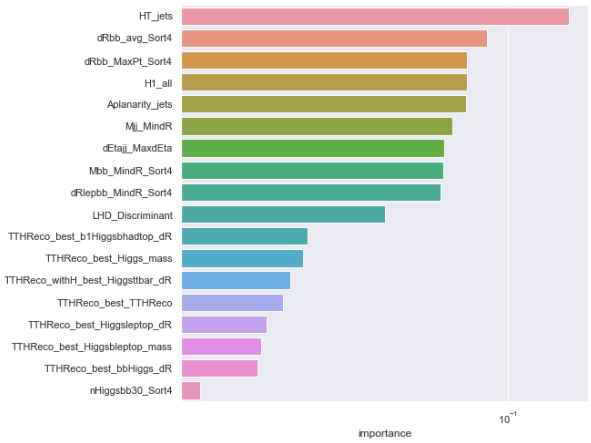
8.2.3.3.2 Feature importance rankings and mean saliency maps

The feature importance rankings extracted from the decision tree modelling are shown in Figure 8.30a and those from the BDT modelling are shown in Figure 8.30b. The mean saliency maps extracted from the

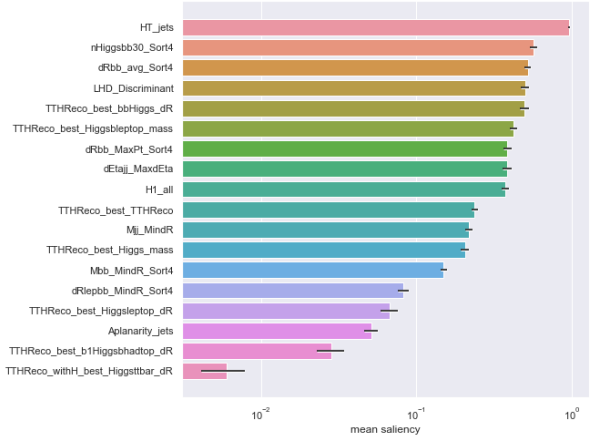
deep multi-layer perceptron modelling, which are derived from taking the mean of 100 saliency maps for each separate class of signal or background, are shown in Figure 8.30c for the signal and in Figure 8.30d for the background, and the signal mean saliency map is divided by the background mean saliency map in Figure 8.30e in order to rank features by how different their saliency is depending on their classification as either signal or background. That ranking of the mean saliencies ratio is shown on a log scale to make differences between feature saliencies more apparent.



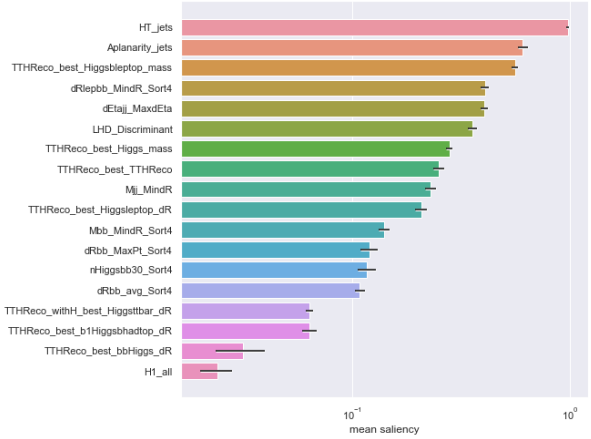
(a) Feature importances of DT classification modelling.



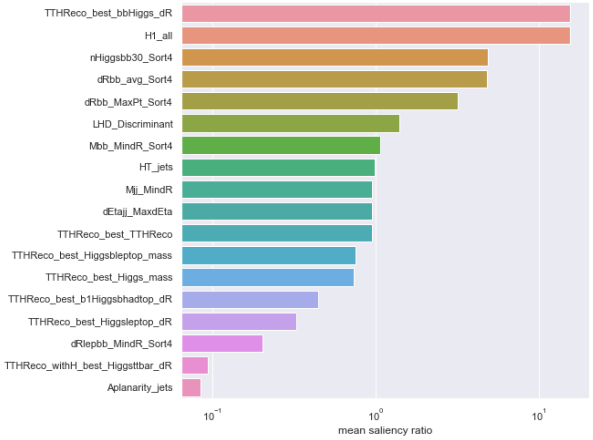
(b) Feature importances of BDT classification modelling.



(c) Mean saliency map of signal from DMLP classification modelling.



(d) Mean saliency map of background from DMLP classification modelling.



(e) Mean saliency maps of signal over background from DMLP classification model.

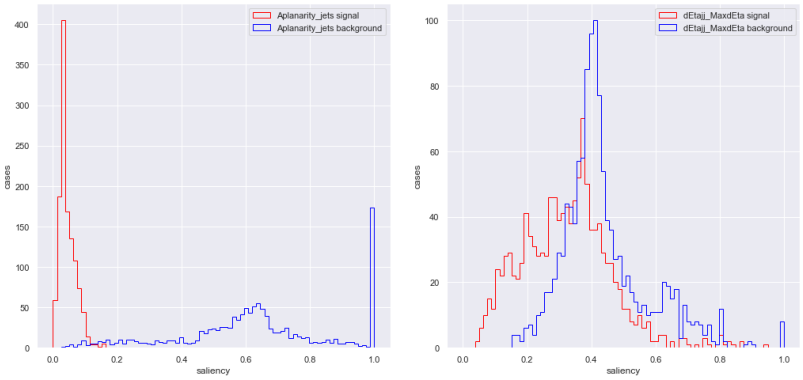
Figure 8.30: Feature importance rankings extracted from the decision tree and BDT modelling and mean saliency maps for signal and background classifications extracted from the deep multi-layer perceptron modelling.

In the case of the feature importance rankings, the importance of a feature corresponds to its contribution to Gini impurity. It is a general ranking of features regardless of classification and gives a sense of what the models based on decision trees find to be important in the data. The BDT approach gives more information than the single decision tree approach, so it is of more interest. The BDT finds that H_{T}, the scalar sum of jet p_{T}, is of particular importance, with ΔR_{bb}^{avg}, the average ΔR for all b-tagged jet pairs, found to be the second most important feature.

In the case of the mean saliency maps, the saliency of a feature corresponds to how important small changes in that feature are to having large changes in classification, and this measure can be guided such that only changes positive to a particular classification are considered, resulting in a per-classification measurement of feature importance. The mean saliency maps of the deep multi-layer perceptron identify that H_{T} is salient for classification of data as either signal or background (which is consistent with the findings of the decision tree models), while N_{30}^{Higgs}, the number of b-jet pairs with invariant mass within 30 GeV of the Higgs boson mass, is found to be salient for classification as signal and the aplanarity of the jets is found to be salient for classification as background. The likelihood discriminant is salient for classification both as signal and as background. As detailed in the main analysis paper, H_{T}, the scalar sum of jet p_{T}, is an important discriminant used in some of the analysis regions and the mean saliency maps are correctly identifying this variable as important [1]. Identification of the other second most salient features for the signal and background classifications is useful because the identification of which features in data are most important for classification may not have a unique answer, particularly when the features have some degrees of correlation.

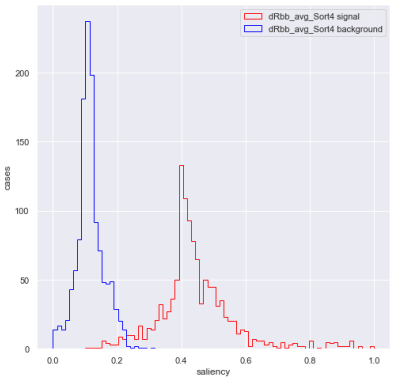
The saliencies of a variety of features were calculated for a subset of events from the test dataset, and the resulting histograms are shown in Figures 8.31 and 8.32. These histograms correspond to the findings of the mean saliency map feature rankings shown in Figures 8.30c and 8.30d in that the number of b-jet pairs with invariant mass within 30 GeV of the Higgs boson mass (Figure 8.32d) tends to be more salient for signal classifications, the aplanarity of the jets (Figure 8.31a) tends to be more salient for background classifications, and the likelihood discriminant (Figure 8.32a) tends to be salient for both signal and background classifications. Some saliencies have someway comparable shapes, such as the mass of the combination of the two b-tagged jets with the smallest ΔR (Figure 8.32b) and the mass of the combination of any two jets with the smallest ΔR (Figure 8.32c), which capture conceptually similar information.

The saliencies calculated for a subset of events from the test dataset can be further split into events that were correctly and incorrectly classified by the model, as shown in Figure 8.33.

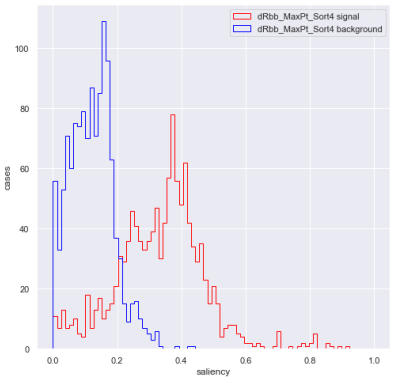


(a) Aplanarity

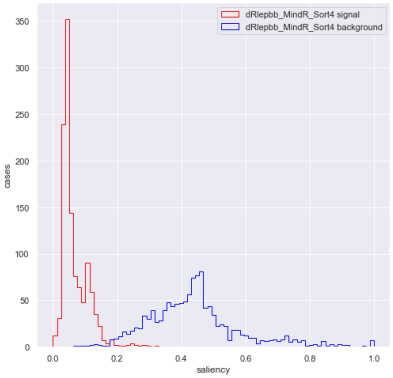
(b) $\Delta\eta_{jj}^{\max \Delta\eta}$



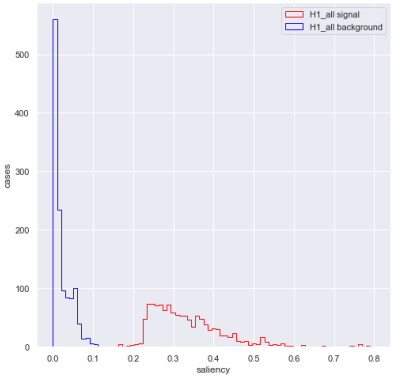
(c) $\Delta R_{bb}^{\text{avg}}$



(d) $\Delta R_{bb}^{\max p_{\text{T}}}$

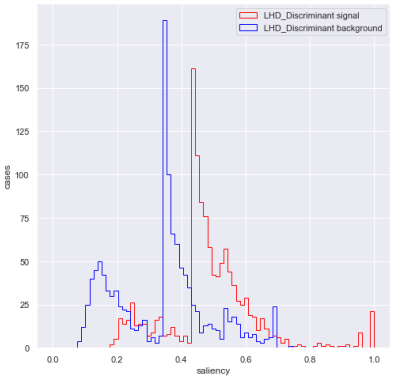


(e) $\Delta R_{\text{lep}-bb}^{\min \Delta R}$



(f) $H1$

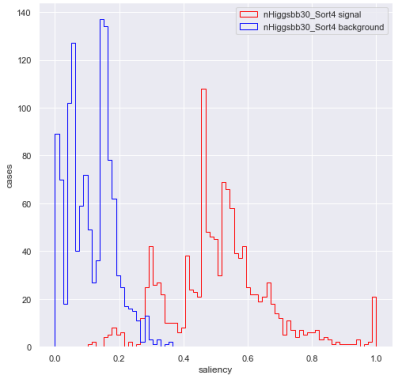
Figure 8.31: Saliencies of various features for both signal (red) and background (blue) cases.



(a) LHD

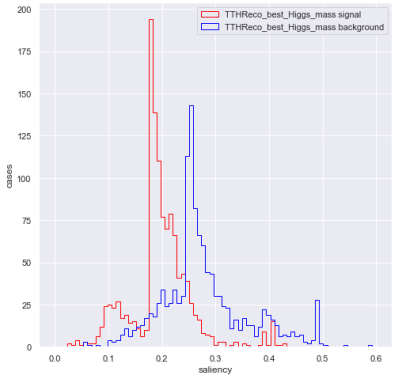
(b) $m_{bb}^{\min \Delta R}$

(c) $m_{jj}^{\min \Delta R}$

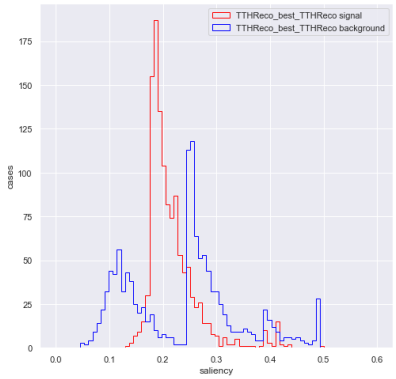


(d) N_{30}^{Higgs}

(e) $m_{H,b_{\text{lep. }t}}$

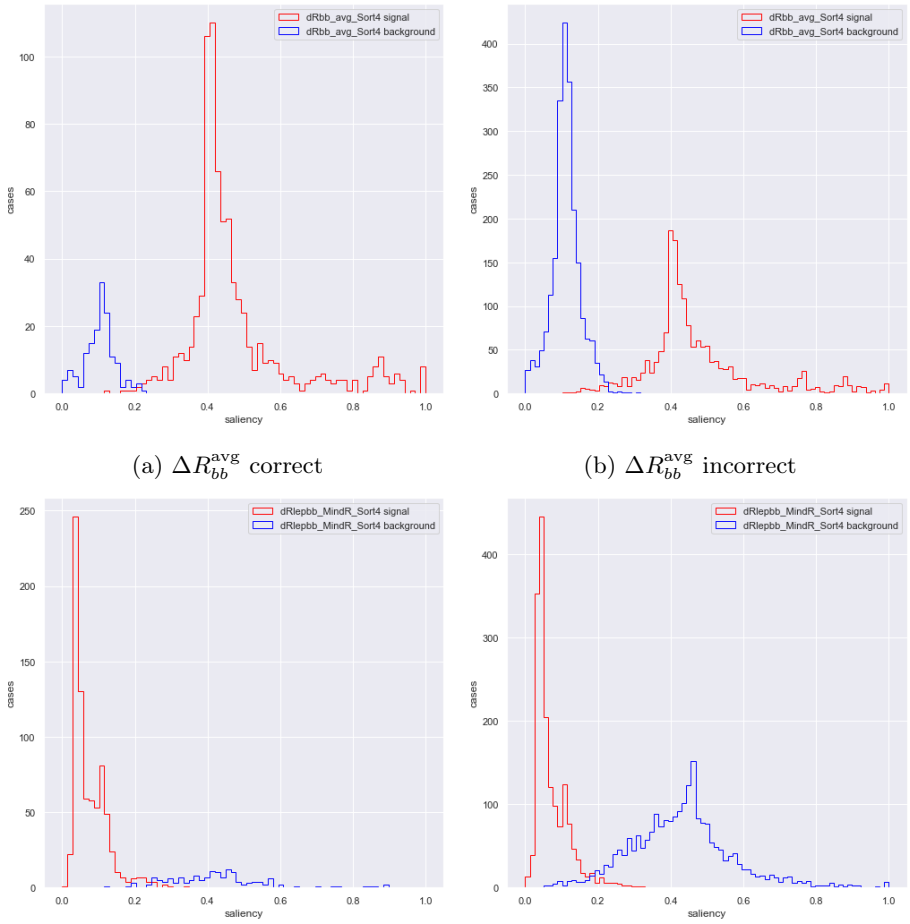


(f) m_H



(g) BDT

Figure 8.32: Saliencies of various features for both signal (red) and background (blue) cases.



(a) $\Delta R_{bb}^{\text{avg}}$ correct

(b) $\Delta R_{bb}^{\text{avg}}$ incorrect

(c) $\Delta R_{\text{lep}-bb}^{\min \Delta R}$ correct

(d) $\Delta R_{\text{lep}-bb}^{\min \Delta R}$ incorrect

Figure 8.33: Saliencies of the average ΔR for all b-tagged jet pairs and the ΔR between the lepton and the combination of the two b-tagged jets with the smallest ΔR for both signal (red) and background (blue) cases, with distributions in Figures 8.33a and 8.33c corresponding to events that were classified correctly by the model and Figures 8.33b and 8.33d corresponding to events that were classified incorrectly by the model.

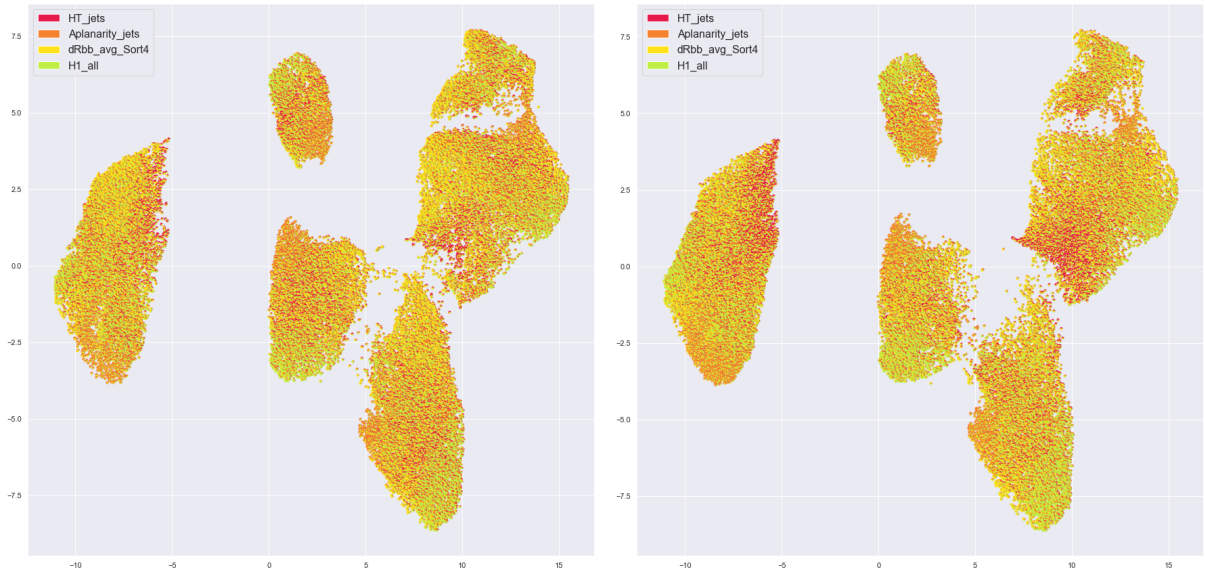
The relative scale factors for the correctly-classified and incorrectly-classified saliencies for signal and background events are characteristic to the individual features. Such comparisons might serve to elucidate the feature modelling achieved and also might be used as validation metrics for new models.

In order to view the features identified as particularly salient in a larger context, UMAP projections were made in which each data point is plotted as a pie chart of the relative magnitudes of the salient features

of that data point, as shown in Figure 8.34.

Figure 8.34: The data used for $t\bar{t}H\left(b\bar{b}\right)/t\bar{t}b\bar{b}$ classification transformed using UMAP (nearest neighbours 6), corresponding to a projection of 18 dimensions down to two. A pie chart at each point shows the relative magnitudes of features in the data.

The same approach was taken for the data separated into signal and background, this time with a much greater number of data points and a reduced pie chart radius (roughly corresponding to a higher resolution). This is shown in Figure 8.35. It should be emphasised that while differences between signal and background data are visible in this representation, a more intensive multivariate approach is needed to highlight these differences clearly. This representation is intended mostly to give some sense of particularly salient features of data in the larger context made possible by UMAP projection.



(a) Signal.

(b) Background.

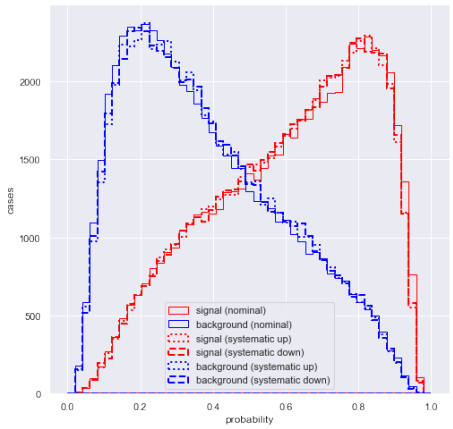
Figure 8.35: The data used for $t\bar{t}H\left(b\bar{b}\right)/t\bar{t}b\bar{b}$ classification transformed using UMAP (nearest neighbours 6), corresponding to a projection of 18 dimensions down to two. A pie chart at each point shows the relative magnitudes of features in the data.

8.2.3.3.3 Comparison of the reliability of the inception convolutional neural network and the reliability of the boosted decision tree in the context of systematic uncertainties

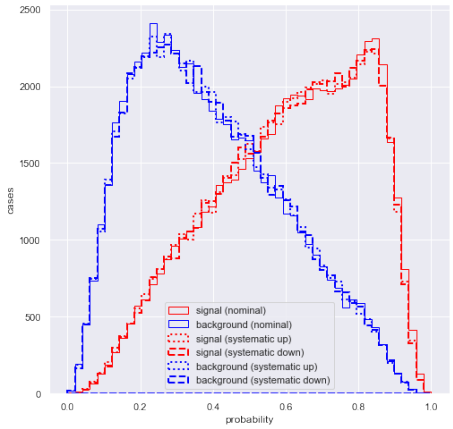
The highly-parameterised model of the ICNN architecture was identified as effective, as described in Section 8.2.3.2. It was amenable to novel techniques of interpretation of its modelling (beyond those available for BDT models), as described in Section 8.2.3.3. Following the identification of the ICNN and BDT models as effective, what was considered next was the robustness to systematic uncertainties of the ICNN model as compared with the BDT model. While statistical uncertainties of machine learning methods can be estimated through relatively straightforward cross-validation, as mentioned in Section 8.1.6.7, the topic of making neural networks and classification models generally aware of systematic uncertainties is under active investigation, as described in Section 8.1.6.7, and here a typical approach used in contemporary high-energy physics is taken, whereby the influence of systematic uncertainties is considered by application of uncertainties to data which is filtered through the model pre-trained on nominal data [359].

As previously, the relative numbers of signal and background events are not realistic and, for the purposes of this comparison, a subset of systematic uncertainties associated with the analysis is propagated through the models in order to give a sense of the stability of the models under systematic changes; a

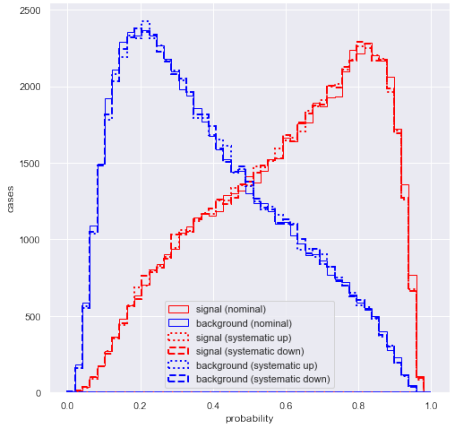
full treatment may involve all required samples and uncertainties and fitting to data. Table 7.2 provides a list of contributing sources to uncertainties. In this comparison, beyond straightforward statistical uncertainties, included are representative systematic uncertainties corresponding to $t\bar{t} \geq 1b$ (referred to here as systematic 1) and $t\bar{t}H$ modelling (referred to here as systematic 5), background model statistical uncertainty (referred to here as systematic 2), b-tagging efficiency and mis-tagging (referred to here as systematic 3), and JES and JER (referred to here as systematic 4). These are propagated through both the ICNN and BDT models in the manner outlined in Section 8.1.6.7. It should be noted also that uncertainties for background normalisation factors are not included, as these would be free-floating parameters in fitting. Separation plots for signal $t\bar{t}H\left(b\bar{b}\right)$ and background $t\bar{t}b\bar{b}$ for the ICNN and BDT models with each of the systematic uncertainties applied are shown together with the nominal distributions in Figures 8.36 and 8.37.



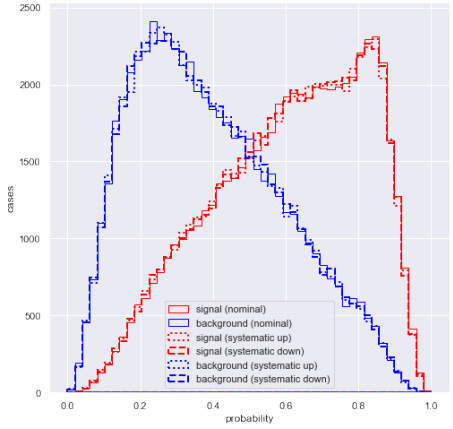
(a) ICNN systematic 1 (score up 0.6960, score down 0.6969)



(b) BDT systematic 1 (score up 0.6939, score down 0.6932)

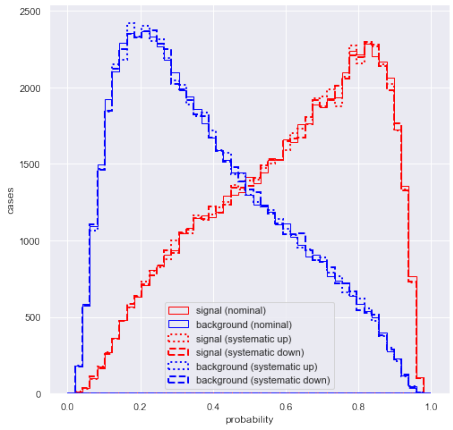


(c) ICNN systematic 2 (score up 0.6991, score down 0.6990)

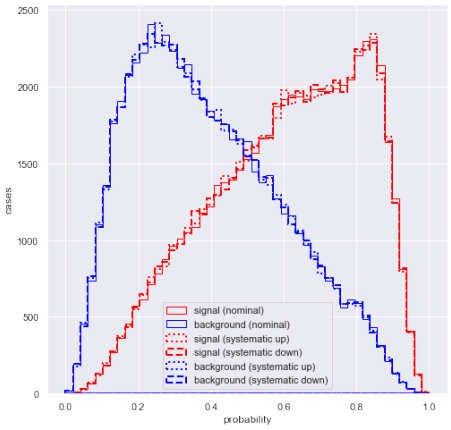


(d) BDT systematic 2 (score up 0.6963, score down 0.6961)

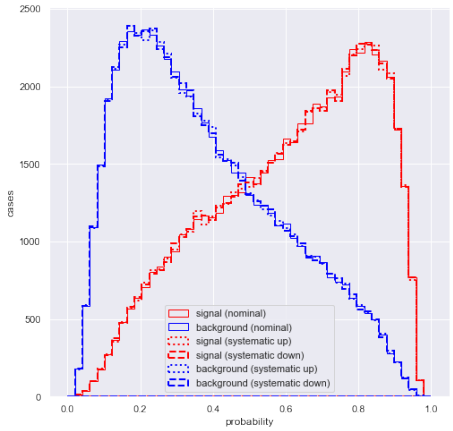
Figure 8.36: Separation plots for signal $t\bar{t}H\left(b\bar{b}\right)$ (red) and background $t\bar{t}b\bar{b}$ (blue) for various multivariate methods with scores shown. The solid lines correspond to the nominal distributions, the dotted lines correspond to systematic variation up and the dashed lines correspond to systematic variation down. Note that it is not possible to calculate exact probabilities for various classifications of a decision tree, so instead the predicted class 'probability' is the fraction of samples of the same class in a leaf.



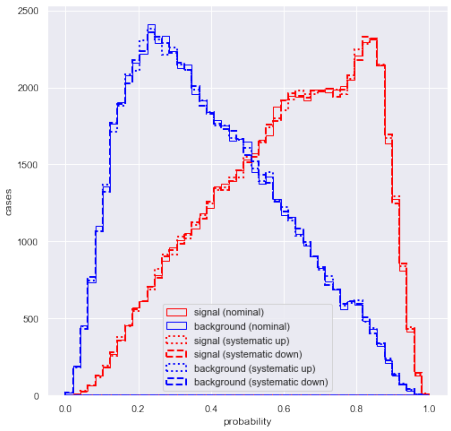
(a) ICNN systematic 3 (score up 0.6998, score down 0.6996)



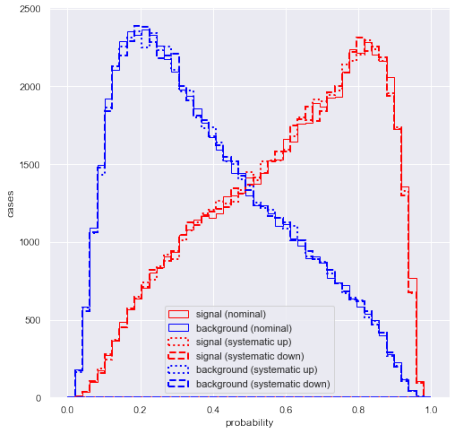
(b) BDT systematic 3 (score up 0.6965, score down 0.6971)



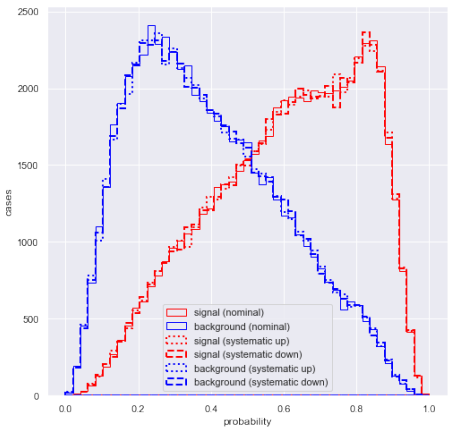
(c) ICNN systematic 4 (score up 0.7000, score down 0.7006)



(d) BDT systematic 4 (score up 0.6971, score down 0.6968)



(e) ICNN systematic 5 (score up 0.6990, score down 0.6997)



(f) BDT systematic 5 (score up 0.6948, score down 0.6960)

Figure 8.37: Separation plots for signal $t\bar{t}H\left(b\bar{b}\right)$ (red) and background $t\bar{t}b\bar{b}$ (blue) for various multivariate methods with scores shown. The solid lines correspond to the nominal distributions, the dotted lines correspond to systematic variation up and the dashed lines correspond to systematic variation down. Note that it is not possible to calculate exact probabilities for various classifications of a decision tree, so instead the predicted class 'probability' is the fraction of samples of the same class in a leaf.

In a manner similar to that of Section 8.2.3.2, the models with statistical uncertainty and the five systematic uncertainties included are compared using a variety of scoring metrics (described in Section 8.1.6.4) in Table 8.6. In Table 8.6, it can be seen that a consistent conclusion may be drawn that the ICNN model is observed to be more efficient than the BDT for all metrics calculated. Both models are observed to be robust to both statistical and systematic uncertainties, with the statistical uncertainties dominating for the validation sample used for calculations. A more extensive treatment (beyond the scope of this thesis) involving greater numbers of events for training, testing and validation may correspond to a reduction in statistical uncertainty, and also a more complete inclusion of further estimations of systematic uncertainties may correspond to an increase in total systematic uncertainty.

Metric	ICNN	BDT
Score	$0.7006 \pm 0.0097\,(\text{stat.})^{+0.0033}_{-0.0028}\,(\text{syst.})$	$0.6979 \pm 0.0136\,(\text{stat.})^{+0.0027}_{-0.0031}\,(\text{syst.})$
Precision	$0.7063 \pm 0.0175\,(\text{stat.})^{+0.0040}_{-0.0031}\,(\text{syst.})$	$0.6954 \pm 0.0105\,(\text{stat.})^{+0.0036}_{-0.0036}\,(\text{syst.})$
KS2	$0.4025 \pm 0.0165\,(\text{stat.})^{+0.0059}_{-0.0046}\,(\text{syst.})$	$0.3971 \pm 0.0250\,(\text{stat.})^{+0.0050}_{-0.0063}\,(\text{syst.})$
AUC	$0.7689 \pm 0.0084\,(\text{stat.})^{+0.0036}_{-0.0030}\,(\text{syst.})$	$0.7655 \pm 0.0098\,(\text{stat.})^{+0.0035}_{-0.0032}\,(\text{syst.})$

Table 8.6: Comparison of the score, precision, Kolmogorov-Smirnov two-sample test statistic and Area Under the ROC Curve (AUC) of the ICNN and BDT models for classification of $t\bar{t}H\,(b\bar{b})$ signal and $t\bar{t}b\bar{b}$ background, with the first and second best respective metrics indicated by gold and silver, with statistical uncertainties and systematic uncertainties derived from the application of five systematic uncertainties to a validation sample.

8.2.3.4 Conclusions

Beyond the relatively simple single decision tree, all of the models developed and investigated were found to be effective for classification of events as signal $t\bar{t}H$ and background $t\bar{t}b\bar{b}$.

The BDT is a capable classification model for the analysis and it has the benefit of being relatively easily interpretable, such as by ranking features in the data which are found to be important for classification decisions. Other, newer models were developed and tested and these models easily achieved comparable performance to the BDT and also marginally surpassed it.

It was observed that inclusion of UMAP as an input feature is of questionable benefit, perhaps given the advent of contemporary computing capabilities. It may be that the models are able to learn much the same information as UMAP is capable of extracting. Similarly, it was observed that inclusion of the classification BDT discriminant as an input feature in the newer models did not result in a noticeable improvement in classification efficiency. It may be that the newer models can learn much the same

information that the BDT can learn.

One of the most apparently effective of the newer models, the ICNN, was compared with the BDT in terms of classification efficiency following the inclusion of statistical uncertainty and several systematic uncertainties. Both the ICNN and BDT models remained effective for classification under the systematic variations, with the ICNN surpassing marginally the BDT in classification efficiency. Improvements to the traditional approach taken for inclusion of the systematic uncertainties are likely possible, as the topic of uncertainties-aware machine learning is under current active research, as described in Section 8.1.6.7.

Importantly, while newer models may have marginally surpassed the performance of the BDT, they also made accessible alternative and additional insights and interpretations of their respectively achieved classification modelling. These included the techniques of mean saliency mapping and activation maximisation. Saliency mapping provided not only a means of ranking features in data on a per-class basis, but also a means of providing saliency measures for individual features in data on a per-event basis, something not obviously possible using conventional models such as decision trees. As detailed in Chapter 10, while existing generators for simulated high-energy physics data are reliable and based on robust underlying theoretical models, machine learning alternatives are perhaps worthy of consideration in terms of reducing usage of computational resources and in terms of reducing time needed for generation, perhaps providing far larger simulated data samples for future experiments [379, 380, 381]. Using the technique of activation maximisation, synthetic events were generated and investigated. While there were some promising results, detailed comparisons between synthetic events generated from successful modelling and simulated Monte Carlo events were not pursued due to the lack of a clear theoretical foundation for such comparisons. However, such synthesis methods may be of benefit in new generators.

Chapter 9

Conclusions

> I always like to end on a positive note, so here is a rousing, uplifting song, which is *guaranteed* to cheer you up.
>
> Tom Lehrer, speaking before performing his 1959 song *We Will All Go Together When We Go* on the topic of the benefits of nuclear apocalypse [382].

9.1 Conclusions

Discovery of the Standard Model Higgs boson in 2012 by the ATLAS and CMS Collaborations at the LHC was an extraordinary moment in the advancement of science. The work presented in this thesis contributed to the search for Higgs bosons produced in association with t-quarks, with particular focus on those Higgs bosons that decay to b-quarks, which helped lead to evidence for $t\bar{t}H$ and subsequent observation.

Identification of jets likely containing b-quarks was key in the search for the $t\bar{t}H\left(b\bar{b}\right)$ decay and this was achieved with the validated and calibrated ATLAS b-tagging algorithms used in Run-2 of the LHC. The installation of the IBL introduced an additional pixel layer to the Inner Detector of ATLAS, which, together with improvements to the b-tagging algorithms, has resulted in extensive advancement of the efficiency of b-tagging of jets, with improvements to light-flavour jet rejection, c-jet rejection *etc.*, and these improvements all have impacts on analyses that use b-tagging, significantly enhancing the testing capabilities of ATLAS for many theories of the physics programme of the LHC.

The analysis software and computing infrastructure and the mechanisms of distributing the efforts in-

volved in their creation, development, and validation were and are crucial to the analysis process. Everything from encoding physics objects together with their calibration recommendations, to the adding and testing of event selection mechanisms, to the ways of smoothly including new analysis channels and collaboratively sharing tools, involves enormous efforts in learning, coordination and validation. The TTHbbLeptonic analysis framework and the efforts involved in its development were successful in the search for $t\bar{t}H$.

New multivariate methods were developed and investigated for classification of events as signal $t\bar{t}H\left(b\bar{b}\right)$ and background $t\bar{t}b\bar{b}$ in order to test classification efficiency, and these methods proved to be effective. New techniques for interpretation of modelling achieved by multivariate methods, such as saliency and activation maximisation, were developed and investigated, and these techniques proved to be effective also.

The highly-parameterised inception convolutional neural network (ICNN) model was compared with the standard analysis boosted decision tree (BDT) model. The ICNN was observed to classify a fraction of $0.7006 \pm 0.0097\,(\text{stat.})^{+0.0033}_{-0.0028}\,(\text{syst.})$ events correctly while the BDT was observed to classify a fraction of $0.6979 \pm 0.0136\,(\text{stat.})^{+0.0027}_{-0.0031}\,(\text{syst.})$ events correctly. Further, the ICNN and other newer models also made accessible alternative and additional insights and interpretations of their respectively achieved classification modelling. These included the techniques of mean saliency mapping and activation maximisation. Saliency mapping provided not only a means of ranking features in data on a per-class basis, but also a means of providing saliency measures for individual features in data on a per-event basis, something not obviously possible using conventional models such as decision trees.

In the main analysis, measurements of Standard Model Higgs bosons decaying to a pair of b-quarks, produced in association with a pair of t-quarks were performed. The results were based on the Run-2 dataset of pp collision data corresponding to an integrated luminosity of $36.1 \pm 0.8\,\text{fb}^{-1}$ collected at $\sqrt{s} = 13\,\text{TeV}$ in the ATLAS detector at the LHC. Events were selected to contain one electron or one muon from the t-quark decays and the events were then categorised according to the number of jets counted in the event and the number of those jets which were identified as likely b-jets. Multivariate methods were used to discriminate between signal and background events. Many backgrounds were considered and all are dominated by $t\bar{t}+\text{jets}$ production.

The $t\bar{t}H\left(b\bar{b}\right)$ analysis is notable within ATLAS for being particularly complex in its use of event regions and multivariate methods, which are applicable to other analyses. The $t\bar{t}H$ search involved development of many facets, detailed and referenced throughout this thesis, such as improvements to the modelling of background systematic uncertainties. Beyond this, the search has provided a direct way to probe the t-H Yukawa coupling and has served as a significant test of the Standard Model and has led to much more detailed qualification of the character of the Higgs boson.

Chapter 10

Outlook

If you think about it... isn't particle physics just the EDM of nature?

Fionn Ball

10.1 Analysis of $t\bar{t}H\left(b\bar{b}\right)$, other Higgs searches and multivariate methods

Following the observation of the $t\bar{t}H$ process, the $t\bar{t}H\left(b\bar{b}\right)$ analysis continues, with gradual improvements and validations of modelling of processes to improve signal sensitivity. Improved modelling of uncertainties particularly from such processes as $t\bar{t}$ and the t-quark, which all are similar in contribution overall to systematic uncertainties, are promising routes to improving the sensitivity of the analysis.

While there has been great success in the $t\bar{t}H$ analyses, the uncertainties associated with the Higgs boson processes are significant and so possibilities such as couplings of the Higgs boson beyond those predicted by the Standard Model, or other physics beyond the Standard Model, cannot be ruled out. Future runs of the LHC are planned to increase the data available, and so are promising for reducing uncertainties associated with the Higgs boson.

The multivariate methods described and tested in this thesis and the techniques for interpretation of modelling, such as saliency and activation maximisation, were shown to be effective. However, in the $t\bar{t}H\left(b\bar{b}\right)$ analysis, a great deal of effort has been put into the definition of effective discriminants and into their optimisation. The degree to which these efforts can be effectively reproduced by multivariate methods has not been assessed. It could be worthwhile to attempt to apply the multivariate methods described, and more advanced methods, at lower levels in the analysis (such as reconstruction) and perhaps higher levels (such as optimisation based on multiple assessments of systematic uncertainties, as discussed in

Section 8.1.6.7). One step could be application at the level of the reconstruction BDT, perhaps leading to a method that could merge the definition of the multivariate variables into the classification BDT. Even further, there are efforts at object reconstruction using tracking and calorimeter data input to a multivariate system. Could a single multivariate system complete a large part of object reconstruction, the definition of high-level variables, and then act in place of the classification BDT across the various analysis regions? As mentioned in Chapter 8, investigations are underway into ways to ensure systematic uncertainties are included in the training of multivariate models. Approaches such as Inference-Aware Neural Optimisation (INFERNO) appear to be promising for LHC analyses [361].

This thesis described a variety of tentative techniques for interpretation of modelling achieved by various deep neural network models. Since the research for this thesis was started, there have appeared the beginnings of other approaches to interpretation of modelling achieved by neural network models, with the broad intent of mapping a neural network modelling solution into a space of human-interpretable observables. One approach from 2021 began with a minimal set of basic jet observables and resulted in condensation of the behaviour of a convolutional neural network model to a small set of energy flow polynomials (EFP) observables (introduced in 2018, a complete set of jet substructure observables which form a discrete linear basis for all infrared-safe and collinear-safe observables) which can be used to approximate closely the decisions of a model [383, 384].

The p_{T} spectrum of the Higgs boson is highly sensitive to new physics, and much larger datasets will enable reduction of uncertainties and increases in precision measurements of differential distributions in all possible channels, which present possible routes to new physics and possible ways to further test theories of reality [385]. For example, investigation of dark matter is possible using searches for the Higgs boson decaying to dark matter particle candidates, wherein all the decays are inferred indirectly through $\not{E}_{\mathrm{T}}$, $H \rightarrow$ invisible [386]. Inclusion of jet-substructure algorithms, boosted analyses and multivariate techniques all hold promise for future exploration.

It should be mentioned that development and application of multivariate techniques should be completed with all due ethical considerations and constraints. Yann LeCun, the inventor of convolutional neural networks, has mentioned that their negative applications can include weapons, biased profiling systems and mechanisms intrusive to privacy [387].

A broad contemporary roadmap for machine learning in high-energy physics has been developed in the form of a white paper [388].

10.1.1 Multivariate methods applied to the study of jets

As studies find that deep neural networks match or outperform traditional jet variables, new jet algorithms involving new multivariate techniques are a promising approach, particularly for analyses involving jets, and the all-hadronic channel of the $t\bar{t}H$ process involves a very high multiplicity of jets. Applying new multivariate techniques could lead to improved definitions of other physics objects. Deep learning methods enable the use not only of a small set of high-level, human-defined observable features, but also of information on jet substructure that is not captured generally by the high-level features [309, 389, 390].

Classification of jets as initiated by quark or by gluon is important in analysis of data from high-energy particle collisions and in the search for physics beyond the Standard Model. Deep neural networks have been integrated in the processing of low-level detector information and has resulted in improvements to the classification efficiency of models to tag jets as quark or gluon [289].

Data with inherent underlying structure, such as trees and graphs, are prevalent with the LHC experiments. As the planned collision rates of the LHC increase over the next decade by approximately an order of magnitude, the application of powerful machine learning methods will likely be essential for data collection and analysis. For example, the particles of a jet can be matched to a tree structure, and recurrent neural networks have been used to match jets to b-quarks [391]. Recurrent neural networks and graph convolutional networks are promising methods for the future study of jets [392].

10.1.2 Focal loss to address limited numbers of data samples

The numbers of examples of different data classes for training networks can be unequal, as can be seen in Monte Carlo generation of data samples, which can make a network inconsistent in its reliability of identification of the different classes. Increasing statistics is one strategy to address this situation. Another strategy is to modify the loss function for training a network to apply a greater focus on scarce, hard examples. This approach is called focal loss and could be beneficial to the large samples used for modelling in high-energy physics analyses [345].

10.1.3 Generative adversarial networks

Fast simulation of particle interactions with detector materials has a number of shortcomings in accuracy for high-precision physics searches and measurements. Recent developments in machine learning methods, such as generative adversarial networks and variational auto-encoders, make sampling high dimensional feature distributions from learned existing data samples a promising alternative for simulation.

Generative adversarial networks can generate simulated data that approximates experimental data. While existing generators for simulated high-energy physics data are reliable and based on robust underlying

theoretical models, the machine learning alternatives are perhaps worthy of consideration in terms of reduced usage of computational resources and in terms of reduced time needed for generation, perhaps providing far larger simulated data samples for future experiments [379, 380, 381].

The generative adversarial network methods could lead from fast simulated physics to realistic data [346]. Indeed, first steps have been taken to develop generative models to enable fast, high-fidelity simulation of particle showers in multilayer electromagnetic calorimeters using generative adversarial networks, achieving calculation speedup factors of up to five orders of magnitude [380]. This is potentially important not only for improvement of physics investigations but also for efficient use of the LHC Computing Grid, approximately 60 % of which is used currently for simulation of particle showering [379]. The generative adversarial network method is a powerful tool for efficient physics simulation and the demonstrated ability to model physics domain knowledge in the neural network models suggests both the flexibility and extensibility of the method for many applications, including mitigation of explicit simulation mismodelling [380]. Investigations have started into fast simulation of the ATLAS calorimeter system using generative adversarial networks to model the response of the ATLAS calorimeter to different types of particle [393].

Adversarial training as been applied successfully in the context of tagging jet substructure, resulting in modelling strategies that enhance signal purity and that reduce the impact of systematic uncertainties in background modelling, results that could be applicable to future iterations of the $t\bar{t}H\left(b\bar{b}\right)$ analysis and others [315].

Adversarial neural networks have been used in considering effective field theory extensions of Higgs boson production in association with jets in order to include known sources of systematic and theoretical uncertainties during training. This prompts a route to more reliable event classification on an event-by-event basis, and perhaps new ideas on how to perform parameter fits for particle physics data [394].

10.1.4 Other multivariate methods under development

The briskly-changing fields of multivariate modelling that involve ideas such as convolutional neural networks, recurrent neural networks, diffusion models and more explorative ideas such as deep belief networks, recursive neural networks, neural Turing machines and memory neural networks hold promise as tools for testing models and hypotheses and perhaps yielding insights into how we think, or better fundamental ideas on which to base science. Such tools will influence computing hardware of the future [395, 396]. New architectural concepts such as dynamic networks, associative-memory structures and sparse activations may require changes to computing hardware architectures to ensure that the calculations implicit are tractable, while such fields as geometric deep learning, in which multivariate techniques are suited for study of data with non-Euclidean underlying structure, may inspire new ways

of formulating physical theories [397].

The approach of probabilistic programming has had recent advancements relevant to LHC analysis and there are plans to develop the demonstrated ETALUMIS framework in tandem with its application to more detailed detector simulation with GEANT4 and more complex particle decays [398]. While probabilistic programming languages remain a challenging topic, they do present a possible means of using existing high-energy physics generators to perform rapid Bayesian inference in complex physics analyses.

As mentioned in Chapter 8, research is ongoing on automated approaches to prediction of model parameters by hypernetworks, which attempt to utilise the experience of training other networks to predict parameters of new networks. While the practical application is limited today, such investigations could be beneficial in the future [349].

10.2 Software and computing

The outlook for the specifics of the analysis framework of the $t\bar{t}H\left(b\bar{b}\right)$ analysis of this thesis is described in Chapter 6.

More generally, with regards to outlooks and plans for future research and development of high-energy physics software and computing, it is vital that there will be support for computing hardware infrastructure capable of acquiring, managing, processing and analysing the much larger amounts of data to be recorded in future experiments. Support for infrastructure necessary for implementation of new multivariate methods is needed. It is vital also that there will be support for research and development of software for these future experiments at every level of abstraction, from efficient readout and trigger software to software that makes physics analysis easier. *A Roadmap for HEP Software and Computing R&D for the 2020s* provides some guidance on this [283].

CERN has long supported open access science, open data and free/libre open source software and hardware [399, 400, 401, 402, 403]. In continued pursuit of free/libre open source software and hardware, CERN has been engaged in a next step with the Microsoft Alternatives project (MALT) which aims to reduce reliances CERN has on corporate dominance and coercion [404, 405, 406].

10.3 ATLAS upgrades in Long Shutdown 2

Extensive upgrades to the CERN accelerator complex and experiments have been underway in preparations for continued LHC runs. The accelerator chain is being renovated as part of the LHC Injectors Upgrade (LIU) project, new LHC equipment is being installed, and experiments are being upgraded and are having numerous components, including entire sub-detectors, upgraded in order to prepare for

higher luminosities. In order to extend sensitivity to new physics, to engage in precision measurements of the Higgs boson, and to investigate further the Standard Model during the next years of the LHC physics programme, ATLAS is undergoing upgrades during the current LHC Long Shutdown 2 (LS2) in preparations for an extended LHC Run-3.

Some upgrades of ATLAS during LS2 are aimed mainly at increasing trigger efficiency for leptonic and hadronic signatures with a p_{T} of at least 20 GeV. The selectivity of the electron trigger will be improved by increasing the level of segmentation available at the trigger level by upgrading components of the front-end electronics of the Electromagnetic Calorimeter, which promises to improve jet rejection, which in turn should better preserve electrons and photons. New electronics suited to the trigger information of increased granularity is to be introduced to the ATLAS trigger and data-acquisition systems. The forward-muon end-cap systems are being upgraded to New Small Wheel (NSW) detectors in order to introduce newer detector technologies which will be capable of operating under the much harsher background conditions of future LHC operations, specifically being able to withstand the anticipated much higher flux of neutrons and photons. This promises to reduce fake muon triggers in the forward region and to improve the sharpness of the trigger threshold, enabling much the same selection capability of the present high-level trigger [407].

The activities of the ATLAS LS2 upgrades are many and varied and will also enable tests in preparation for installation of a new all-silicon inner tracker which is scheduled for Long Shutdown 3 (LS3) [408, 409, 410, 411, 412].

10.4 High-Luminosity Large Hadron Collider and ATLAS High-Luminosity upgrades

The High-Luminosity Large Hadron Collider (HL-LHC) is an upgrade to the LHC started in 2018 that is expected to deliver a collision rate of about an order of magnitude greater than that of the LHC. Specifically, the upgrade aims to increase the luminosity of the machine to 10^{35} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [413, 414]. In order to withstand the much higher instantaneous luminosity of the HL-LHC, ATLAS must undergo many upgrades. This next phase of the LHC is scheduled a number of years in the future. The data resulting from the High-Luminosity LHC should make it possible to detect extremely rare phenomena and to far improve the precision of measurements.

The HL-LHC is scheduled to operate from approximately 2028. Reports have been published attempting to quantify the potential of the physics programmes planned for the HL-LHC, taking into account hundreds of extrapolations for physics measurements, expected changes in relevant physics theory, im-

provements to detectors, and educated guesses at possible new analysis techniques [415, 16].

A sense of the uncertainties that are expected currently for Higgs coupling modifier parameters for the combination of ATLAS and CMS extrapolations to about the end of the HL-LHC physics programme in perhaps approximately 20 years is summarised in Figure 10.1.

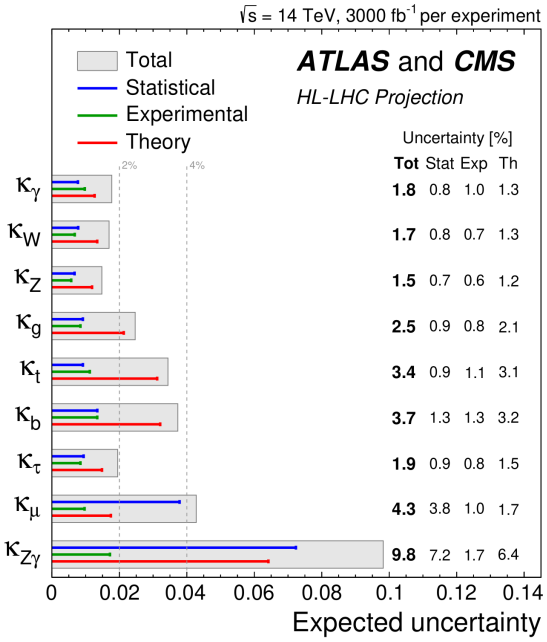


Figure 10.1: Summary plot showing the total expected $\pm 1\sigma$ uncertainties on Higgs coupling modifier parameters for the combination of ATLAS and CMS extrapolations, with coupling modifier parameters defined within the κ-framework, wherein a set of coupling modifiers are introduced in order to parameterise potential deviations from Standard Model predictions of Higgs boson couplings to Standard Model bosons and fermions, *i.e.* for a production process or decay mode j, a coupling modifier κ_j can be defined such that $\kappa^2 = \frac{\sigma_j}{\sigma_j^{\mathrm{SM}}}$, with all κ_j values in the Standard model positive and equal to unity [16, 84]. Total uncertainty is shown as a grey box while statistical, experimental and theory uncertainties are shown as blue, green and red lines respectively.

The main Higgs boson couplings are estimated to be measurable at the HL-LHC with a precision at the percent level [16]. Larger statistics are expected to be of particular help in the study of complex final states, such as those arising from $t\bar{t}H$ production. The constraining power of contemporary $t\bar{t}H$ analyses has been limited to plausible improvements to theory predictions, in particular in the $t\bar{t}H\left(b\bar{b}\right)$ channel.

The 3.4 % precision on the κ_t coupling modifier for the t-quark thus obtained is due mostly to the other direct $t\bar{t}H$ measurement channels [16].

10.5 Future Circular Colliders

The Future Circular Colliders (FCC) exploratory study programme was launched in February 2014 as a five-year international design study on the possibility of a future next-generation circular collider capable of centre-of-mass collision energies of 100 TeV and an integrated luminosity of at least a factor of five greater than that of the planned HL-LHC [416, 417]. Where the LHC has a circumference of about 27 km, this future circular collider may have a circumference of up to 100 km, as shown in Figure 10.2. It could extend the current energy frontier by almost an order of magnitude.

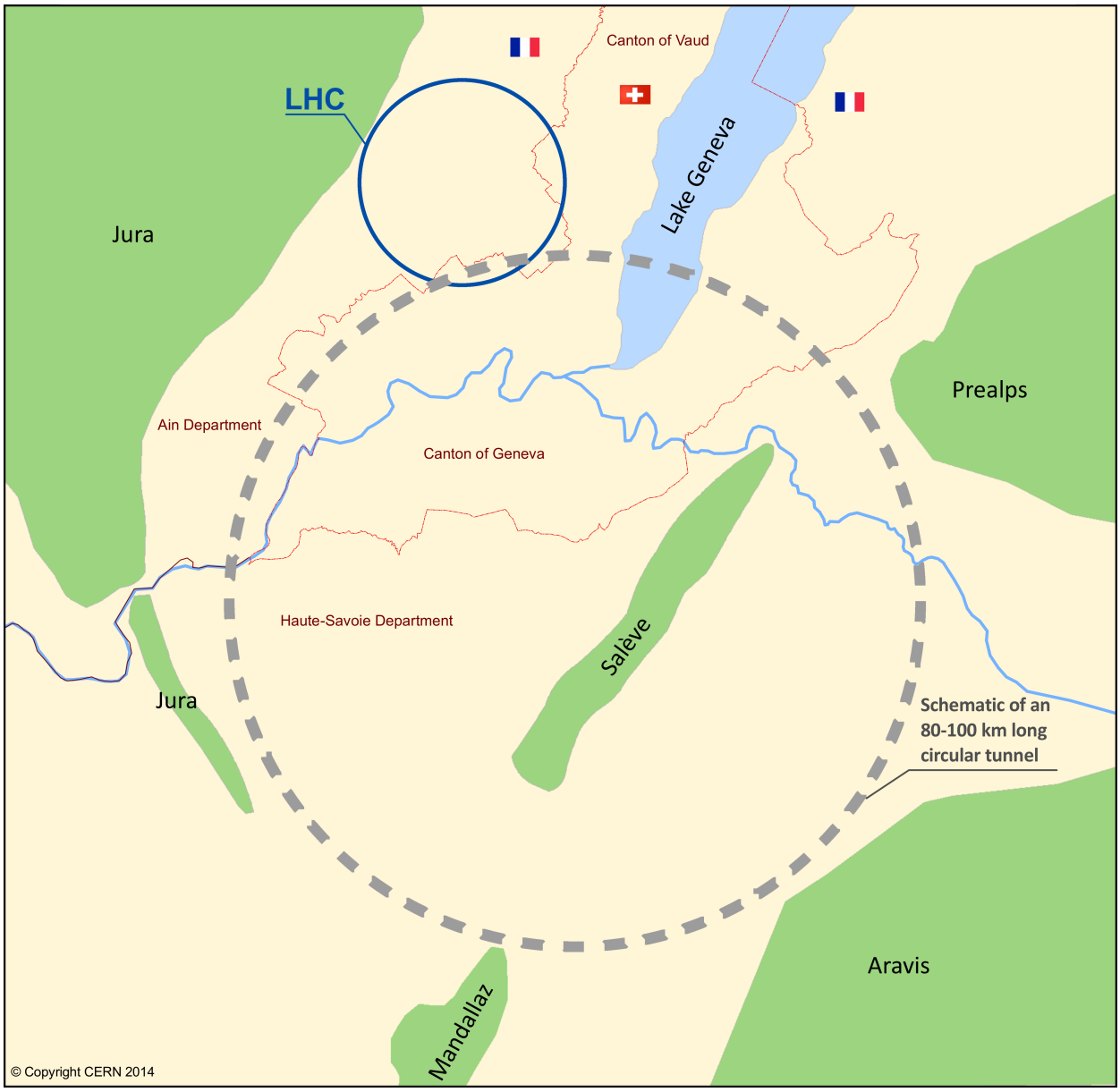


Figure 10.2: Schematic of an 80–100 km long circular tunnel at CERN [418].

The two main types of large particle collider to date – lepton colliders and hadron colliders – tend to complement each other, with hadron colliders used more in discovery of new particles and new physics and lepton colliders used more for refining of these discoveries by making precision measurements. A pre-feasibility study on the FCC concluded that the most feasible project would be an 80 km tunnel

under Lakeside Geneva, as opposed to a tunnel under the Jura mountains, at a depth of up to four times that of the current LHC tunnel [416].

The FCC study is conducted in parallel with research on the proposed Compact Linear Collider (CLIC), a lepton collider capable of collision energies of up to 3 TeV, far greater than the collision energies of any previous lepton collider [419]. Muon colliders are also a topic of current research and, while they pose some technical challenges (some of which were explored by research on the International Muon Ionization Cooling Experiment (MICE)), they present a promising approach for future experimental physics, including the remarkable possibility of a pulsed 14 TeV muon collider in the LHC tunnel [420, 421, 422].

The FCC study includes considerations of an intermediate 90 to 400 GeV lepton collider and a lepton-hadron collider. The proposed electron-positron collider, called Triple LEP (TLEP), could be an intermediate successor to the LHC while the proposed ultimate successor to the Large Hadron Collider is called... the *Very High Energy Large Hadron Collider* (VHE-LHC)[1].

The mass reach for discoveries with the FCC could reach several tens of TeV, opening possibilities such as the production of new particles, precision measurements of self-coupling of the Higgs boson and the dynamics of electroweak symmetry breaking.

10.6 Replicability of experimental particle physics results

Recently there has been a focus on replicability of scientific results, prompting reviews of methodologies of research, publication *etc.* Experimental particle physics involves particularly large amounts of time and effort to reach results and there are many stringent checks and safeguards in the field which do appear to mitigate this form of problem [424, 425]. Where applicable, these approaches to maximise the reliability of particle physics results should be communicated clearly to other fields while also the field of experimental particle physics should be reviewed as far as reasonable to detect any replicability problems.

10.7 Global challenges: from COVID-19 to global warming to wellbeing and participation

There are a number of serious challenges impacting the outlook of particle physics research.

One significant challenge facing the world is the COVID-19 pandemic. Extensive commentary on the impacts of this pandemic on people participating in particle physics is well beyond the scope of this doc-

[1] This follows the naming convention described by xkcd 1294 [423].

ument. CERN has, in many ways, been at the forefront of behaviours and technologies for decentralised collaboration, in everything from the web to videoconferencing (which features significant benefits, such as inclusion of people who are unable to meet in person, and which features significant detriments, such as exclusion of people who need to meet in person) [426]. Beyond this, however, there are negative impacts associated with the pandemic on wellbeing, mental health, focus, collaboration and productivity [427, 428]. The LHCb Collaboration has attempted to quantify some of these impacts and to offer some suggestions to ameliorate the situation [429]. This and further similar investigations could be beneficial, as could implementation of their recommendations.

Beyond COVID-19, there are other challenges associated with health, in everything from ability to participate in science, to wellbeing in academia. As codified in Article 27.1 of the Universal Declaration of Human Rights, "Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits." [430]. This right may be better protected through methods such as cancellation of debt, unconditional guaranteed income, opposition to those lethal economic programmes of structural adjustment and austerity, and by addressing wealth and other forms of inequality [431, 432, 433, 434, 435]. Goals would be improving the wellbeing of those currently participating in physics, and enabling more people around the world to participate in physics.

While CERN is arguably a benefit for local biodiversity and for local regions that include woodlands and wetlands, CERN contributes to global warming [436]. CERN has identified a number of environmental indicators as being significant at CERN and external associates and has published a number of environmental goals, such as a commitment to a significant reduction in direct emissions [437]. These goals and commitments should be implemented and further efforts should be made to respond to global warming. The energy usage of contemporary machine learning implementations is not insignificant and ways to address this should be devised also [438, 439].

10.8 Expanding the frontiers of high-energy physics

The machines of the LHC and their associated collaborations have had extraordinary successes. The discovery of the Higgs boson, and in more recent times its observation in production associated with t-quarks, completes the matrix of particles and interactions that constitutes the Standard Model of particle physics. Yet while this model has been consistent and predictive, indeed to the degree that it has been successful at describing apparently all physical phenomena accessible to collider experiments, several experimental observations require the extension or modification of the model, such as the non-zero masses of the neutrinos, the abundance of matter over antimatter and the evidence for dark matter. Further, theoretical issues such as the hierarchy problem and the dynamical origin of the Englert-Brout-

Higgs-Guralnik-Hagen-Kibble mechanism compels the search for the existence of new physics beyond the Standard Model.

The LHC physics programme is what is good in life. And it shall continue onwards and upwards, but mostly upwards. You have now reached *die Sollbruchstelle*[2]. I thank you for reading and I wish you well on your quest.

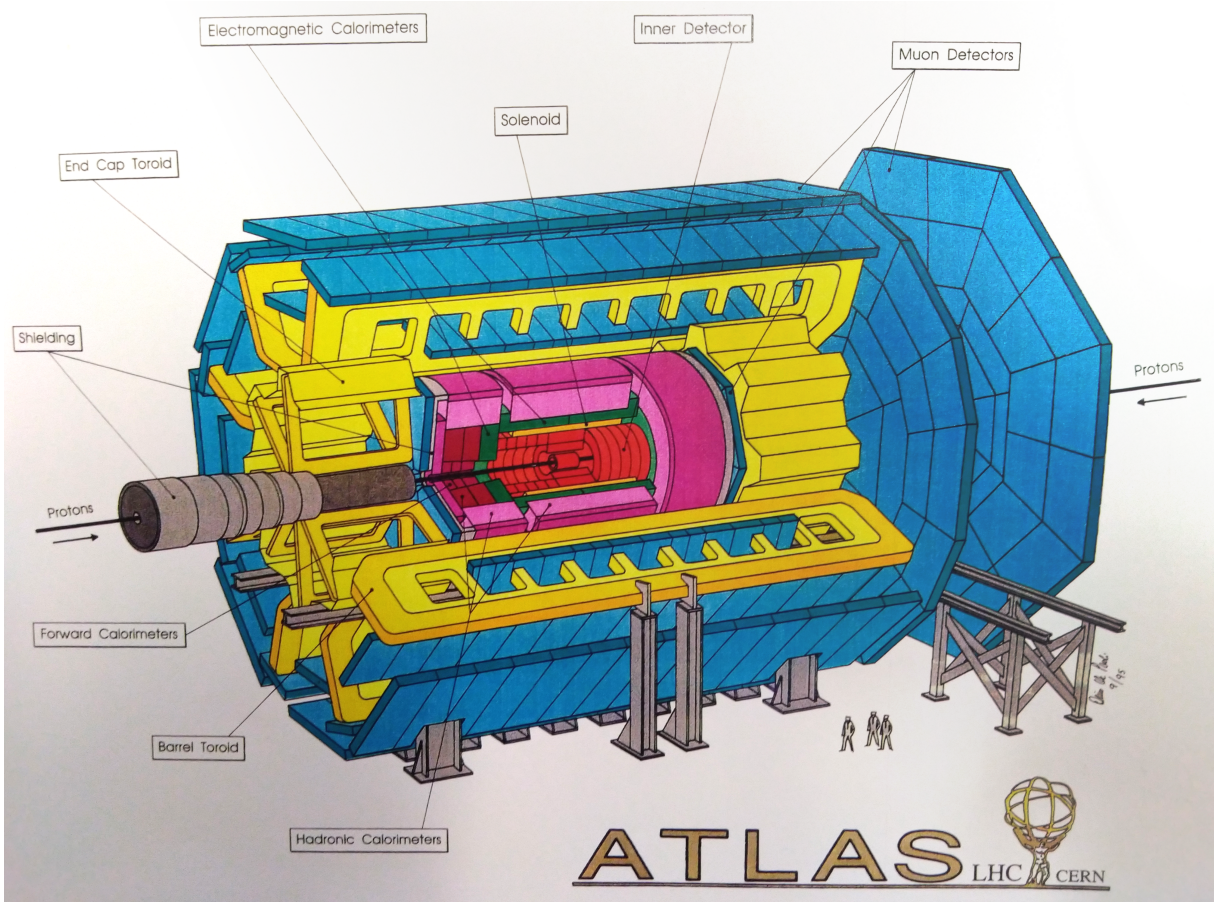


Figure 10.3: 1995 diagram of ATLAS, a reminder of the multi-generational nature of our explorations and an acknowledgement of past masters [440].

[2] *Sollbruchstelle* [predetermined breaking point] – term kindly suggested by Andrea Knue.

References

[1] The ATLAS Collaboration, *Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Physical Review D, 97 (7) (April 2018), doi:10.1103/PhysRevD.97.072016, URL https://doi.org/10.1103/PhysRevD.97.072016

[2] The ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Physics Letters B, 784, 173–191 (September 2018), doi:10.1016/j.physletb.2018.07.035, URL https://doi.org/10.1016/j.physletb.2018.07.035

[3] The CMS Collaboration, *Observation of $t\bar{t}H$ Production*, Physical Review Letters, 120 (23) (June 2018), doi:10.1103/physrevlett.120.231801, URL https://doi.org/10.1103/physrevlett.120.231801

[4] The ATLAS Collaboration, *ATLAS Computing Acknowledgements*, Technical Report ATL-GEN-PUB-2016-002, CERN, Geneva (July 2016), URL https://cds.cern.ch/record/2202407

[5] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Physical Review Letters, 13 (9), 321–323 (August 1964), doi:10.1103/PhysRevLett.13.321, URL https://doi.org/doi/10.1103/PhysRevLett.13.321

[6] P. W. Higgs, *Broken Symmetries, Massless Particles and Gauge Fields*, Physics Letters, 12, 132–133 (September 1964), doi:10.1016/0031-9163(64)91136-9, URL https://doi.org/10.1016/0031-9163(64)91136-9

[7] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Physical Review Letters, 13 (16), 508–509 (October 1964), doi:10.1103/PhysRevLett.13.508, URL https://doi.org/10.1103/PhysRevLett.13.508

[8] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Physical Review Letters, 13 (20), 585–587 (November 1964), doi:10.1103/PhysRevLett.13.585, URL https://doi.org/10.1103/PhysRevLett.13.585

[9] J. R. Ellis, M. K. Gaillard and D. V. Nanopoulos, *A Phenomenological Profile of the Higgs Boson*, Nuclear Physics B, 106 (CERN-TH-2093) (1976), doi:10.1016/0550-3213(76)90382-5, URL https://doi.org/10.1016/0550-3213(76)90382-5

[10] L. L. Camilleri, D. C. Cundy, P. Darriulat, J. R. Ellis, J. Field, H. Fischer, E. Gabathuler et al., *Physics with Very High-Energy e^+ e^- Colliding Beams*, CERN-76-18, CERN, Geneva (November 1976), doi:10.5170/CERN-1976-018, URL https://doi.org/10.5170/CERN-1976-018

[11] The 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), 1–29 (2012), ISSN 0370-2693, doi:10.1016/j.physletb.2012.08.020, URL https://doi.org/10.1016/j.physletb.2012.08.020

[12] The 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), 30–61 (2012), doi:10.1016/j.physletb.2012.08.021, URL https://doi.org/10.1016/j.physletb.2012.08.021

[13] C. Englert, A. Freitas, M. M. Mühlleitner, T. Plehn, M. Rauch, M. Spira and K. Walz, *Precision measurements of Higgs couplings: implications for new physics scales*, Journal of Physics G: Nuclear and Particle Physics, 41 (11) (2014), doi:10.1088/0954-3899/41/11/113001, URL https://doi.org/10.1088/0954-3899/41/11/113001

[14] The ATLAS Collaboration, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ TeV and 8 TeV*, Journal of High Energy Physics, 2016 (8), 45 (August 2016), ISSN 1029-8479, doi:10.1007/JHEP08(2016)045, URL https://doi.org/10.1007/JHEP08(2016)045

[15] *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, Nature, 607 (7917), 52–59 (July 2022), doi:10.1038/s41586-022-04893-w, URL https://doi.org/10.1038/s41586-022-04893-w

[16] M. Cepeda, S. Gori, P. Ilten, M. Kado, F. Riva, R. Abdul Khalek, A. Aboubrahim et al., *Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC*, CERN Yellow Reports: Monographs, 7 (CERN-LPCC-2018-04) (December 2018), doi:10.23731/CYRM-2019-007.221, URL http://cds.cern.ch/record/2650162

[17] The ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (ATLAS-CONF-2016-080) (August 2016), URL https://cds.cern.ch/record/2206255

[18] G. Aad, J. Adelman, N. Asbah, S. K. Boutle, W. D. Breaden Madden, T. P. Calvet, M. Casolino et al., *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Technical Report ATL-COM-PHYS-2016-116, CERN, Geneva (February 2016), URL https://cds.cern.ch/record/2130799

[19] F. Derue, *Estimation of fake lepton background for top analyses using the Matrix Method with the 2015 dataset at $\sqrt{s} = 13$ TeV with AnalysisTop-2.3.41*, Technical Report ATL-COM-PHYS-2016-198, CERN, Geneva (Febuary 2016), URL https://cds.cern.ch/record/2135116

[20] D. Whiteson, *SUSY Data Set*, Machine Learning Repository (July 2014), URL https://archive.ics.uci.edu/ml/datasets/SUSY

[21] G. A. Stewart, W. D. Breaden Madden, H. J. Maddocks, T. Harenberg, M. Sandhoff and B. Sarrazin, *ATLAS Job Transforms: A Data Driven Workflow Engine*, Journal of Physics: Conference Series, 513 (3), 032094 (2014), doi:10.1088/1742-6596/513/3/032094, URL https://www.doi.org/10.1088/1742-6596/513/3/032094

[22] Y. Abreu et al., *Commissioning and Operation of the Readout System for the SoLid Neutrino Detector* (2018), doi:10.48550/arXiv.1812.05425, URL https://doi.org/10.48550/arXiv.1812.05425

[23] Y. Abreu, Y. Amhis, L. Arnold, G. Ban, W. Beaumont, M. Bongrand, D. Boursette et al., *A novel segmented-scintillator antineutrino detector*, Journal of Instrumentation, 12 (04), P04024–P04024 (April 2017), doi:10.1088/1748-0221/12/04/p04024, URL https://iopscience.iop.org/article/10.1088/1748-0221/12/04/P04024

[24] L. N. Kalousis, *The SoLid experiment*, Journal of Physics: Conference Series, 888, 012181 (September 2017), doi:10.1088/1742-6596/888/1/012181, URL https://iopscience.iop.org/article/10.1088/1742-6596/888/1/012181

[25] G. Kirby, *Higgs production in association with top quarks at the LHC* (2016), URL http://theses.gla.ac.uk/7550

[26] *Large Hadron Collider exhibition* (February 2013), URL https://external.parliament.scot/visitandlearn/58283.aspx

[27] An Comhchoiste um Ghnó, Fiontar agus Nuálaíocht, *An Cás ar son Chomhaltas na hÉireann de CERN*, (032 BEI 021) (Samhain 2019), URL https://data.oireachtas.ie/ie/oireachtas/committee/dail/32/joint_committee_on_business_enterprise_and_innovation/reports/2019/2019-11-13_report-on-the-case-for-irish-membership-of-cern_en.pdf

[28] *Membership of CERN: Motion*, Seanad Éireann debate, 291 (1) (December 2022), URL https://www.oireachtas.ie/en/debates/debate/seanad/2022-12-15/10

[29] C. O'Raifeartaigh, *A happy new year for Ireland and Cern*, The Irish Times (December 2022), URL https://www.irishtimes.com/science/2022/12/29/a-happy-new-year-for-ireland-and-cern

[30] K. Anthony, *ATLAS Colouring Book*, (ATLAS-OUTREACH-2016-001) (November 2016), URL https://cds.cern.ch/record/2228941

[31] W. D. Breaden Madden, *TTHbbLeptonic 2015*, CERN Document Server (December 2015), doi:10.17181/cds.2658216, URL https://videos.cern.ch/record/2658216

[32] The ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Technical Report ATLAS-CONF-2017-076, CERN, Geneva (November 2017), URL https://cds.cern.ch/record/2291393

[33] The ATLAS Collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, Physical Review D, 97, 072003 (April 2018), doi:10.1103/PhysRevD.97.072003, URL https://doi.org/10.1103/PhysRevD.97.072003

[34] The CMS Collaboration, *Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying τ leptons at $\sqrt{s} = 13$ TeV, journal = Journal of High Energy Physics, reportNumber = CMS-HIG-17-018,*, 2018 (8) (August 2018), doi:10.1007/jhep08(2018)066, URL https://doi.org/10.1007/jhep08(2018)066

[35] W. D. Breaden Madden, *Searching for Higgs bosons in $t\bar{t}H$, $H \to b\bar{b}$ with ATLAS in LHC in Run-2: a brief introduction to the $t\bar{t}H$, $H \to b\bar{b}$ analysis, its infrastructure and discovery prospects*, Joint annual HEPP and APP conference (March 2016), URL https://indico.cern.ch/event/507870/contributions/1186956

[36] B. Sowden, *Search for the Higgs Boson decaying to two b-quarks, produced in association with leptonically decaying top quarks in Run 2 at ATLAS*, Joint annual HEPP and APP conference (March 2016), URL https://indico.cern.ch/event/507870/contributions/1186957

[37] S. J. Crawley, *Search for SM Higgs bosons produced in association with a t quark pair and decaying to two b quarks in pp collisions at 13 TeV in ATLAS, in the boosted regime*, Joint annual HEPP and APP conference (March 2016), URL https://indico.cern.ch/event/507870/contributions/1186958

[38] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Addison-Wesley (1995), ISBN 9780201503975

[39] H. Georgi, *Lie Algebras In Particle Physics*, CRC Press (May 2018), doi:10.1201/9780429499210, URL https://doi.org/10.1201/9780429499210

[40] T. Gowers, J. Barrow-Green and I. Leader (editors), *The Princeton Companion to Mathematics*, Princeton University Press (July 2010), doi:10.2307/j.ctt7sd01, URL https://doi.org/10.2307/j.ctt7sd01

[41] E. Norrbin and T. Sjöstrand, *Production and Hadronization of Heavy Quarks*, The European Physical Journal C, 17 (1), 137–161 (October 2000), doi:10.1007/s100520000460, URL https://doi.org/10.1007/s100520000460

[42] S. Bethke, *Determination of the QCD coupling α_s*, Journal of Physics G: Nuclear and Particle Physics, 26 (7), R27–R66 (June 2000), doi:10.1088/0954-3899/26/7/201, URL https://doi.org/10.1088/0954-3899/26/7/201

[43] M. Tanabashi, K. Hagiwara, K. Hikasa, K. Nakamura, Y. Sumino, F. Takahashi, J. Tanaka et al., *Review of Particle Physics*, Physical Review D, 98, 030001 (August 2018), doi:10.1103/PhysRevD.98.030001, URL https://www.doi.org/10.1103/PhysRevD.98.030001

[44] G. Dissertori, I. G. Knowles and M. Schmelling, *Quantum Chromodynamics: High Energy Experiments and Theory*, number 115 in International Series of Monographs on Physics, Oxford University Press (2009), ISBN 0199566410

[45] M. Tegmark, A. Aguirre, M. J. Rees and F. Wilczek, *Dimensionless constants, cosmology, and other dark matters*, Physical Review D, 73 (2) (January 2006), doi:10.1103/physrevd.73.023505, URL https://doi.org/10.1103/physrevd.73.023505

[46] M. Gell-Mann, *The Quark and the Jaguar: Adventures in the Simple and the Complex*, Little, Brown and Company (1994), ISBN 0316907413

[47] The CDF Collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, Physical Review Letters, 74 (FERMILAB-PUB-95-022-E, CDF-PUB-TOP-PUBLIC-3040, ANL-HEP-PR-95-44), 2626–

2631 (1995), doi:10.1103/PhysRevLett.74.2626, URL https://doi.org/10.1103/PhysRevLett.74.2626

[48] The DØ Collaboration, *Observation of the top quark*, Physical Review Letters, 74 (FERMILAB-PUB-95-028-E, D0-2698, D0-PUB-95-7), 2632–2637 (1995), doi:10.1103/PhysRevLett.74.2632, URL https://doi.org/10.1103/PhysRevLett.74.2632

[49] K. Kodama et al., *Observation of tau neutrino interactions*, Physics Letters B, 504 (FERMILAB-PUB-00-335-E), 218–224 (2001), doi:10.1016/S0370-2693(01)00307-0, URL https://doi.org/10.1016/S0370-2693(01)00307-0

[50] The LHCb Collaboration, *Test of lepton universality in beauty-quark decays*, (LHCB-PAPER-2021-004) (March 2021), URL http://cds.cern.ch/record/2758740

[51] B. Abi, T. Albahri, S. Al-Kilani, D. Allspach, L. P. Alonzi, A. Anastasi, A. Anisenkov et al., *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, Physical Review Letters, 126, 141801 (April 2021), doi:10.1103/PhysRevLett.126.141801, URL https://link.aps.org/doi/10.1103/PhysRevLett.126.141801

[52] *The Hunt for the Higgs Boson*, Science Scotland, (3) (2004), ISSN 1743-6222, URL https://www.sciencescotland.org/downloads/science_scotland_3.pdf

[53] A. Robinson, *The great high-energy write-off*, Physics World (December 2015), URL https://physicsworld.com/a/the-great-high-energy-write-off

[54] P. W. Anderson, *Plasmons, Gauge Invariance, and Mass*, Physical Review, 130 (1), 439–442 (April 1963), doi:10.1103/PhysRev.130.439, URL https://doi.org/10.1103/PhysRev.130.439

[55] J. I. Friedman, *The Road to the Nobel Prize* (July 2008), URL https://web.archive.org/web/20081009075526/http://www.hueuni.edu.vn/hueuni/en/news_detail.php?NewsID=1606

[56] *Now Presenting: A Visualization of the Proton* (5 2022), URL https://www.jlab.org/news/stories/now-presenting-visualization-proton

[57] R. E. Taylor, *The discovery of the point-like structure of matter*, Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 359 (1779), 225–240 (February 2001), doi:10.1098/rsta.2000.0723, URL https://doi.org/10.1098/rsta.2000.0723

[58] E. D. Bloom, D. H. Coward, H. DeStaebler, J. Drees, G. Miller, L. W. Mo, R. E. Taylor et al., *High-Energy Inelastic e-p Scattering at 6° and 10°*, Physical Review Letters, 23 (16), 930–934 (October 1969), doi:10.1103/physrevlett.23.930, URL https://doi.org/10.1103/physrevlett.23.930

[59] A. Ali, C. Hambrock and M. J. Aslam, *Tetraquark interpretation of the BELLE data on the anomalous $\upsilon(1S)\pi^+\pi^-$ and $\upsilon(2S)\pi^+\pi^-$ production near the $\upsilon(5S)$ resonance*, Physical Review Letters, 104 (16) (April 2010), doi:10.1103/PhysRevLett.104.162001, URL https://doi.org/10.1103/PhysRevLett.104.162001

[60] The LHCb Collaboration, *Observation of new resonances decaying to $J/\psi K^+$ and $J/\psi\phi$*, Physical Review Letters, 127 (8) (August 2021), doi:10.1103/PhysRevLett.127.082001, URL https://doi.org/10.1103/PhysRevLett.127.082001

[61] The LHCb Collaboration, *Evidence for Exotic Hadron Contributions to $\Lambda_b^0 \to J/\psi p\pi^-$ Decays*, Physical Review Letters, 117 (August 2016), doi:10.1103/PhysRevLett.117.082003, URL https://doi.org/10.1103/PhysRevLett.117.082003

[62] *A New Map of All the Particles and Forces*, Quanta Magazine (October 2020), URL https://www.quantamagazine.org/a-new-map-of-the-standard-model-of-particle-physics-20201022

[63] N. O'C, *Weak Decay*, Wikipedia (September 2013), URL https://en.wikipedia.org/wiki/File:Weak_Decay_(flipped).svg

[64] P. Zyla et al., *Review of Particle Physics*, Progress of Theoretical and Experimental Physics, (083C01) (2020), URL https://pdg.lbl.gov/2020

[65] F. Wilczek, *The Standard Model transcended*, Nature, 394 (6688), 13–15 (July 1998), doi:10.1038/27758, URL https://doi.org/10.1038/27758

[66] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, Helvetica Physica Acta, 6, 110–127 (February 1933), URL http://adsabs.harvard.edu/abs/1933AcHPh...6..110Z

[67] F. Zwicky, *On the Masses of Nebulae and of Clusters of Nebulae*, The Astrophysical Journal, 86 (3), 217–246 (October 1937), doi:10.1086/143864, URL http://adsabs.harvard.edu/doi/10.1086/143864

[68] J. Dietrich P., N. Werner, D. Clowe, A. Finoguenov, T. Kitching, L. Miller and A. Simionescu, *A filament of dark matter between two clusters*, Nature, 487 (7406), 202–204 (July 2012), doi:10.1038/nature11224, URL https://www.nature.com/articles/nature11224

[69] B. A. Dobrescu and P. J. Fox, *Quark and lepton masses from top loops*, Journal of High Energy Physics, 2008 (08), 100–100 (August 2008), doi:10.1088/1126-6708/2008/08/100, URL https://doi.org/10.1088/1126-6708/2008/08/100

[70] S. Weinberg, *Models of lepton and quark masses*, Physical Review D, 101 (3) (February 2020), doi:10.1103/physrevd.101.035020, URL https://doi.org/10.1103/physrevd.101.035020

[71] J. Rafelski, G. S. B. Müller and W. Greiner, *Critical Discussion of the Vacuum Polarization Measurements in Muonic Atoms*, Annalen der Physik, 88, 419–453 (December 1974), doi: 10.1016/0003-4916(74)90177-8, URL https://doi.org/10.1016/0003-4916(74)90177-8

[72] S. L. Adler, R. F. Dashen and S. B. Treiman, *Comments on proposed explanations for the muonic-atom x-ray discrepancy*, Physical Review D, 10 (11), 3728–3735 (December 1974), doi:10.1103/PhysRevD.10.3728, URL https://doi.org/doi/10.1103/PhysRevD.10.3728

[73] K. Sato and H. Sato, *Primordial Higgs Mesons and Cosmic Background Radiations*, Progress of Theoretical Physics, 54 (3), 912–913 (May 1975), doi:10.1143/PTP.54.912, URL http://ptp.ipap.jp/link?PTP/54/912

[74] K. Sato and H. Sato, *Higgs Meson Emission from a Star and a Constraint on its Mass*, Progress of Theoretical Physics, 54 (5), 1564–1565 (July 1975), doi:10.1143/PTP.54.1564, URL http://ptp.ipap.jp/link?PTP/54/1564

[75] B. W. Lee, C. Quigg and H. Thacker, *Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass*, Physical Review D, 16, 1519–1531 (September 1977), doi:10.1103/PhysRevD.16.1519, URL https://doi.org/10.1103/PhysRevD.16.1519

[76] J. D. Bjorken, *Weak Interactions*, (SLAC-PUB-2134) (June 1978), URL https://inspirehep.net/literature/130285, Invited talk presented at 13^{th} Rencontre de Moriond, Les Arcs, France (12–24 March 1978)

[77] J. Ellis, G. L. Fogli and E. Lisi, *The Higgs boson mass from precision electroweak data*, Zeitschrift für Physik C: Particles and Fields, 69 (4), 627–633 (1996), URL https://doi.org/10.1007/s002880050066

[78] J. Ellis, M. K. Gaillard and D. V. Nanopoulos, *A Historical Profile of the Higgs Boson*, (CERN-PH-TH-2012-009) (January 2012), URL https://cds.cern.ch/record/1420264

[79] J. Ellis, M. K. Gaillard and D. V. Nanopoulos, *An Updated Historical Profile of the Higgs Boson*, (CERN-PH-TH-2015-098) (April 2015), doi:10.1142/9789814733519_0014, URL https://cds.cern.ch/record/2012465

[80] Tevatron New Physics Higgs Working Group, *Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data*, FERMILAB-CONF-12-318-E, CDF-NOTE-10884, D0-NOTE-6348 (September 2018), doi:10.48550/arXiv.1207.0449, URL https://doi.org/10.48550/arXiv.1207.0449

[81] W. D. Breaden Madden, *Search for charged Higgs Bosons decaying via $H^+ \to \tau_{lep.} + \nu$ in $t\bar{t}$ events at $\sqrt{s} = 7$ TeV in ATLAS* (2012), URL http://theses.gla.ac.uk/3752

[82] The ATLAS Collaboration, *LHC Higgs Cross Section WG Picture Gallery* (December 2016), URL https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWGCrossSectionsFigures

[83] The ATLAS Collaboration, *SM Higgs production cross sections at $\sqrt{s} = 13$–14 TeV (CERN Report 3)* (December 2016), URL https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV2014

[84] S. Heinemeyer, C. Mariotti, G. Passarino, R. Tanaka, J. R. Andersen, P. Artoisenet, E. A. Bagnaschi et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*, number CERN-2013-004 in CERN Yellow Reports: Monographs (July 2013), doi:10.5170/CERN-2013-004, URL https://cds.cern.ch/record/1559921

[85] The ATLAS Collaboration, *A combination of measurements of Higgs boson production and decay using up to 139 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment*, (ATLAS-CONF-2020-027) (August 2020), URL http://cds.cern.ch/record/2725733

[86] The CMS Collaboration, *Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic distributions in the dilepton final state in proton-proton collisions at $\sqrt{13}$ TeV*, 102 (9) (November 2020), doi:10.1103PhysRevD.102.092013, URL https://doi.org/10.1103/physrevd.102.092013

[87] F. Bezrukov and M. Shaposhnikov, *Why should we care about the top quark Yukawa coupling?*, Journal of Experimental and Theoretical Physics, 120 (CERN-PH-TH-2014-218), 335–343 (2015), doi:10.1134/S1063776115030152, URL https://doi.org/10.1134/S1063776115030152

[88] J. N. Ng and P. Zakarauskas, *QCD-parton calculation of conjoined production of Higgs bosons and heavy flavors in $p\bar{p}$ collisions*, Physical Review D, 29 (5), 876–886 (March 1984), doi:10.1103/physrevd.29.876, URL https://doi.org/10.1103/physrevd.29.876

[89] Z. Kunszt, *Associated production of heavy Higgs boson with top quarks*, Nuclear Physics B, 247 (2), 339–359 (December 1984), doi:10.1016/0550-3213(84)90553-4, URL https://doi.org/10.1016/0550-3213(84)90553-4

[90] S. Dawson, L. H. Orr, L. Reina and D. Wackeroth, *Next-to-leading order QCD corrections to $pp \to t\bar{t}H$ at the CERN Large Hadron Collider*, Physical Review D, 67 (7) (April 2003), doi:10.1103/physrevd.67.071503, URL https://doi.org/10.1103/physrevd.67.071503

[91] W. Beenakker, S. Dittmaier, M. Krämer, B. Plümper, M. Spira and P. M. Zerwas, *Higgs Radiation Off Top Quarks at the Tevatron and the LHC*, Physical Review Letters, 87 (20) (October 2001), doi:10.1103/physrevlett.87.201805, URL https://doi.org/10.1103/physrevlett.87.201805

[92] D. de Florian, C. Grojean, F. Maltoni, C. Mariotti, A. Nikitenko, M. Pieri, P. Savard et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, number CERN-2017-002-M in CERN Yellow Reports: Monographs, CERN (October 2016), doi:10.23731/cyrm-2017-002, URL https://doi.org/10.23731/cyrm-2017-002

[93] The ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, The European Physical Journal C, 75 (7) (July 2015), doi:10.1140/epjc/s10052-015-3543-1, URL https://doi.org/10.1140/epjc/s10052-015-3543-1

[94] The ATLAS Collaboration, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV*, (ATLAS-CONF-2015-044) (September 2015), URL http://cds.cern.ch/record/2052552

[95] C. D. White, *The SAGEX Review on Scattering Amplitudes, Chapter 12: Amplitudes and collider physics* (March 2022), doi:10.48550/arXiv.2203.13023, URL https://doi.org/10.48550/arXiv.2203.13023

[96] G. Brianti, *The LHC machine* (March 1992), URL http://cds.cern.ch/record/247469, Conference: Towards the LHC Experimental Programme: General Meeting on LHC Physics and Detectors, Evian-les-Bains, France, 5–8 March 1992

[97] C. L. Smith, *Genesis of the Large Hadron Collider*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373 (2032) (November 2014), doi:10.1098/rsta.2014.0037, URL https://doi.org/10.1098/rsta.2014.0037

[98] *In the heart of CERN in 1967*, (CERN-VIDEO-1967-001-001) (1967), doi:10.17181/cds.43113, URL https://doi.org/10.17181/cds.43113

[99] The ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section using dilepton events in pp collisions at $\sqrt{s} = 5.02$ TeV with the ATLAS detector* (March 2021), URL http://cds.cern.ch/record/2754223

[100] M. Degtyarev, *The incident at the Large Hadron Collider experiment* (July 2018), URL https://www.behance.net/gallery/67661801/The-incident-at-the-Large-Hadron-Collider-experiment

[101] P. Traczyk, *59 new hadrons and counting* (March 2021), URL https://home.cern/news/news/physics/59-new-hadrons-and-counting

[102] C. De Melis, *The CERN accelerator complex*, (OPEN-PHO-ACCEL-2016-001) (January 2016), URL https://cds.cern.ch/record/2119882

[103] *CERN OVERVIEW animation*, CERN Document Server, (OPEN-VIDEO-2025-024-001) (June 2015), doi:10.17181/cds.2020780, URL https://videos.cern.ch/record/2020780

[104] AC Team, *Diagram of an LHC dipole magnet*, (CERN-DI-9906025) (June 1999), URL https://cds.cern.ch/record/40524

[105] AC Team, *Computer-generated diagram of an LHC dipole*, (CERN-DI-9809007) (September 1998), URL https://cds.cern.ch/record/39731

[106] R. Alemany, M. Lamont and S. Page, *Functional specification: LHC Modes*, Technical Report LHC-OP-ES-0005-10-00, CERN (11 2007), URL https://lhc-commissioning.web.cern.ch/lhc-commissioning/systems/data-exchange/doc/LHC-OP-ES-0005-10-00.pdf

[107] W. D. Breaden Madden, *ATLAS LHC first collisions 2015*, CERN Document Server (May 2015), doi:10.17181/cds.2689648, URL https://videos.cern.ch/record/2689648

[108] *How the barn was born*, Physics Today, 25 (7) (1972), doi:10.1063/1.3070918, URL https://doi.org/10.1063/1.3070918

[109] The CDF Collaboration, *Charged jet evolution and the underlying event in proton-antiproton collisions at 1.8 TeV*, Physical Review D, 65 (9) (April 2002), doi:10.1103/physrevd.65.092002, URL https://doi.org/10.1103/physrevd.65.092002

[110] The ATLAS Collaboration, *Luminosity Public Results Run-2: Delivered Luminosity versus time for 2011-2018* (October 2018), URL https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2

[111] The ATLAS Collaboration, *Luminosity Public Results Run-2: Number of Interactions per Crossing* (April 2021), URL https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2

[112] The ATLAS Collaboration, *Luminosity determination in pp collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*, Technical Report ATLAS-CONF-2019-021, CERN (June 2019), URL http://cds.cern.ch/record/2677054

[113] C. Weiner, *Oral History Interviews – George Gamow*, American Institute of Physics (April 1968), URL https://www.aip.org/history-programs/niels-bohr-library/oral-histories/4325

[114] The ALICE Collaboration, *Letter of Intent for A Large Ion Collider Experiment*, Technical Report CERN-LHCC-93-016, LHCC-I-4, CERN, Geneva (March 1993), URL https://cds.cern.ch/record/290825

[115] The ALICE Collaboration, *ALICE: Technical Proposal for A Large Ion Collider Experiment at the CERN LHC*, number CERN-LHCC-95-71 in LHC Technical Proposal, CERN, Geneva (December 1995), URL https://cds.cern.ch/record/293391

[116] The LHCb Collaboration, *LHC-B: Letter of Intent*, Technical Report CERN-LHCC-95-5, LHCC-I-8, CERN, Geneva (August 1995), URL https://cds.cern.ch/record/290868

[117] The LHCb Collaboration, *LHCb: Technical Proposal: A Large Hadron Collider Beauty Experiment for Precision Measurements of CP Violation and Rare Decays*, number CERN-LHCC-98-004 in Technical Proposal, CERN, Geneva (1998), ISBN 92-9083-123-5, URL https://cds.cern.ch/record/622031

[118] The CMS Collaboration, *CMS: The Compact Muon Solenoid: Letter of Intent*, Technical Report CERN-LHCC-92-003, LHCC-I-1, CERN, Geneva (October 1992), URL https://cds.cern.ch/record/290808

[119] The CMS Collaboration, *Technical Proposal*, number CERN-LHCC-94-38 in LHC Technical Proposal, CERN, Geneva (December 1994), URL http://cds.cern.ch/record/290969

[120] P. R. Norton, *Expression of Interest: The Ascot detector at the LHC*, 40 p (March 1992), URL http://cds.cern.ch/record/1076511, Conference: Towards the LHC Experimental Programme: General Meeting on LHC Physics and Detectors, Evian-les-Bains, France, 5–8 March 1992

[121] P. Jenni, *Expression of Interest: EAGLE: Experiment for Accurate Gamma, Lepton and Energy measurements:* (March 1992), URL http://cds.cern.ch/record/247456, Conference: Towards the LHC Experimental Programme: General Meeting on LHC Physics and Detectors, Evian-les-Bains, France, 5–8 March 1992

[122] CERN, *General Meeting on LHC Physics and Detectors*, CERN, Geneva (March 1992), URL http://cds.cern.ch/record/236265

[123] The ATLAS Collaboration, *ATLAS: Letter of Intent for a General-Purpose pp Experiment at the Large Hadron Collider at CERN*, number CERN-LHCC-92-004 in Letter of Intent, CERN, Geneva (October 1992), URL http://cds.cern.ch/record/291061

[124] The ATLAS Collaboration, *ATLAS: Technical Proposal for a General-Purpose pp Experiment at the Large Hadron Collider at CERN*, number CERN-LHCC-94-43, LHCC-P-2 in LHC Technical Proposal, CERN, Geneva (December 1994), URL https://cds.cern.ch/record/290968

[125] The ATLAS Collaboration, *Expected Performance of the ATLAS Experiment: Detector, Trigger and Physics*, CERN-OPEN-2008-020, CERN, Geneva (December 2008), URL http://cds.cern.ch/record/1125884

[126] J. Pequenao, *Computer generated image of the whole ATLAS detector*, (CERN-GE-0803012) (March 2008), URL https://cds.cern.ch/record/1095924

[127] The ATLAS Collaboration, *ATLAS Inner Detector Technical Design Report Volume 2*, number CERN-LHCC-97-017 in Technical Design Report ATLAS, CERN (April 1997), URL http://cds.cern.ch/record/331064

[128] M. Aleksa, F. Bergsma, L. Chevalier, P. A. Giudici, J. C. Hart, A. Kehrli, M. Losasso et al., *Results of the ATLAS solenoid magnetic field map*, Journal of Physics: Conference. Series, 110, 092018 (2008), doi:10.1088/1742-6596/110/9/092018, URL https://cds.cern.ch/record/1178488

[129] J. Pequenao, *Computer generated image of the ATLAS inner detector*, (CERN-GE-0803014) (March 2008), URL https://cds.cern.ch/record/1095926

[130] The ATLAS Collaboration, *A measurement of material in the ATLAS tracker using secondary hadronic interactions in 7 TeV pp collisions*, (CERN-EP-2016-137) (September 2016), doi:10.1088/1748-0221/11/11/P11020, URL https://doi.org/10.1088/1748-0221/11/11/P11020

[131] The ATLAS Collaboration, *Alignment of the ATLAS Inner Detector in Run-2*, The European Physical Journal C (July 2020), doi:10.1140/epjc/s10052-020-08700-6, URL https://cds.cern.ch/record/2724037

[132] M. Capeans, G. Darbo, K. Einsweiller, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme et al., *ATLAS Insertable B-Layer Technical Design Report*, (ATLAS-TDR-19) (September 2010), URL https://cds.cern.ch/record/1291633?ln=en

[133] C. Marcelloni De Oliveira, *IBL installation into the inner detector of the ATLAS Experiment side C*, (ATLAS-PHO-COLLAB-2014-008) (May 2014), URL https://cds.cern.ch/record/1702006

[134] W. D. Breaden Madden, *Shift Summary for ID General desk*, Electronic Logbook for the information storage of ATLAS (ELisA) (May 2015), URL https://atlasop.cern.ch/elisa/display/260508

[135] W. D. Breaden Madden, *Shift Summary for ID General desk*, Electronic Logbook for the information storage of ATLAS (ELisA) (June 2015), URL https://atlasop.cern.ch/elisa/display/264667

[136] K. Lantzsch, *Operational Experience with the ATLAS Pixel Detector* (September 2016), URL https://cds.cern.ch/record/2213084

[137] L. A. Bella, *Status of the ATLAS Liquid Argon calorimeter and its performance after two years of LHC operation*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 718, 60–62 (August 2013), doi:10.1016/j.nima.2012.11.077, URL https://doi.org/10.1016/j.nima.2012.11.077

[138] *ATLAS electromagnetic calorimeter layer*, (CERN-EX-9308048_09) (August 1993), URL https://cds.cern.ch/record/39737

[139] M. Brice, *Electromagnetic calorimeter end-cap*, (CERN-EX-0308005) (August 2003), URL http://cds.cern.ch/record/636673

[140] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation, 3 (8) (August 2008), doi:10.1088/1748-0221/3/08/S08003, URL https://doi.org/10.1088/1748-0221/3/08/S08003

[141] A. Barriuso Poy, H. Boterenbrood, H. J. Burckhart, J. Cook, V. Filimonov, S. Franz, O. Gutzwiller et al., *The detector control system of the ATLAS experiment*, Journal of Instrumentation, 3 (May 2008), URL https://cds.cern.ch/record/1107601

[142] A. Ruiz-Martinez and The ATLAS Collaboration, *The Run-2 ATLAS Trigger System*, Technical Report ATL-DAQ-PROC-2016-003, CERN (February 2016), doi:10.1088/1742-6596/762/1/012003, URL http://cds.cern.ch/record/2133909

[143] I. Foster and C. Kesselman, *The Grid 2: Blueprint for a New Computing Infrastructure* (1998), ISBN 9780080521534

[144] *Tier centres* (2019), URL http://wlcg-public.web.cern.ch/tier-centres

[145] K. Anthony, *Behind very great results lies great computing*, (ATLAS-PHO-ach-2015-050) (November 2015), URL https://atlas.cern/updates/atlas-news/behind-very-great-results-lies-great-computing

[146] N. Quadri, *ATLAS News: Computing Process Diagram*, (ATLAS-PHO-ach-2015-050) (November 2015), URL http://cds.cern.ch/record/2116437

[147] A. Buckley, T. Eifert, M. Elsing, D. Gillberg, K. Koeneke, A. Krasznahorkay, E. Moyse et al., *Implementation of the ATLAS Run 2 event data model*, Journal of Physics: Conference Series, 664 (7), 072045 (December 2015), doi:10.1088/1742-6596/664/7/072045, URL https://doi.org/10.1088/1742-6596/664/7/072045

[148] P. Mato, *GAUDI LHCb Data Processing Applications Framework: Architecture Design Document*, Technical report, CERN, Geneva (November 1998), URL https://cds.cern.ch/record/691746

[149] LHCb Computing Group, *GAUDI – The Software Architecture and Framework for building LHCb Data Processing Applications*, International Conference on Computing in High Energy Physics (February 2014), URL http://chep2000.pd.infn.it/abs/abs_a152.htm

[150] C. Leggett, D. Quarrie, M. Marino, P. Calafiura and W. Lavrijsen, *The Athena Control Framework in Production, New Developments and Lessons Learned*, International Conference on Computing in High Energy Physics (September 2000), URL https://indico.cern.ch/event/0/contributions/1294172

[151] G. A. Stewart, W. D. Breaden Madden, H. J. Maddocks, T. Harenberg, M. Sandhoff and B. Sarrazin, *ATLAS Job Transforms: A Data Driven Workflow Engine*, Journal of Physics: Conference Series, 513, 032094 (2014), URL http://cds.cern.ch/record/2026326

[152] D. Golubkov, B. Kersevan, A. Klimentov, A. Minaenko, P. Nevski, A. Vaniachine and R. Walker, *ATLAS Grid Data Processing: system evolution and scalability*, Journal of Physics: Conference Series, 396 (3), 032049 (December 2012), doi:10.1088/1742-6596/396/3/032049, URL https://iopscience.iop.org/article/10.1088/1742-6596/396/3/032049

[153] G. M. Amdahl, *Validity of the single processor approach to achieving large scale computing capabilities*, in *Proceedings of the April 18–20, 1967, Spring Joint Computer Conference on - AFIPS '67 (Spring)*, ACM Press (1967), doi:10.1145/1465482.1465560, URL https://doi.org/10.1145/1465482.1465560

[154] G. A. Stewart, J. Baines, T. Bold, P. Calafiura, A. Dotti, S. A. Farrell, C. Leggett et al., *Multi-threaded software framework development for the ATLAS experiment*, Journal of Physics: Conference Series, 762, 012024 (October 2016), doi:10.1088/1742-6596/762/1/012024, URL https://doi.org/10.1088/1742-6596/762/1/012024

[155] C. Leggett, J. Baines, T. Bold, P. Calafiura, S. Farrell, P. van Gemmeren, D. Malon et al., *AthenaMT: upgrading the ATLAS software framework for the many-core world with multi-threading*, Journal of Physics: Conference Series, 898, 042009 (October 2017), doi:10.1088/1742-6596/898/4/042009, URL https://doi.org/10.1088/1742-6596/898/4/042009

[156] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin et al., *TensorFlow: A system for large-scale machine learning*, in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283 (2016), URL https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

[157] *TensorFlow Release Version 0.8.0*, GitHub repository (April 2016), URL https://github.com/tensorflow/tensorflow/releases/tag/v0.8.0rc0

[158] F. Chollet et al., *Keras: Deep Learning for humans* (2015), URL https://keras.io

[159] R. Kotikalapudi et al., *keras-vis: Keras Visualization Toolkit* (2017), URL https://github.com/raghakot/keras-vis

[160] M. Kotila, *Talos: Hyperparameter Optimization for Keras*, GitHub repository (May 2018), URL https://github.com/autonomio/talos

[161] R. Stallman, *GNU General Public License Version 3* (June 2007), URL https://www.gnu.org/licenses/gpl

[162] J. Hale, *Deep Learning Framework Power Scores 2018* (September 2018), URL https://towardsdatascience.com/deep-learning-framework-power-scores-2018-23607ddf297a

[163] A. Hocker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes et al., *TMVA 4: Toolkit for Multivariate Data Analysis with ROOT: Users Guide*, Technical Report CERN-OPEN-2007-007, CERN, Geneva (March 2007), URL https://cds.cern.ch/record/1019880

[164] The NNPDF Collaboration, *Parton distributions for the LHC run II*, Journal of High Energy Physics, (4) (April 2015), doi:10.1007/jhep04(2015)040, URL https://doi.org/10.1007/jhep04(2015)040

[165] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin and C.-P. Yuan, *New parton distributions for collider physics*, Physical Review D, 82 (7) (October 2010), doi:10.1103/physrevd.82.074024, URL https://doi.org/10.1103/physrevd.82.074024

[166] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Höche, H. Hoeth, F. Krauss et al., *General-purpose event generators for LHC physics*, Physics Reports, 504 (5), 145–233 (July 2011), doi:10.1016/j.physrep.2011.03.005, URL https://doi.org/10.1016/j.physrep.2011.03.005

[167] D. J. Gross and F. Wilczek, *Asymptotically Free Gauge Theories. I*, Physical Review D, 8 (10), 3633–3652 (November 1973), doi:10.1103/PhysRevD.8.3633, URL https://doi.org/doi/10.1103/PhysRevD.8.3633

[168] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Physical Review Letters, 30 (26), 1343–1346 (June 1973), doi:10.1103/PhysRevLett.30.1343, URL https://doi.org/doi/10.1103/PhysRevLett.30.1343

[169] F. Wilczek, *Nobel Lecture: Asymptotic freedom: From paradox to paradigm*, Reviews of Modern Physics, 77 (3), 857–870 (September 2005), doi:10.1103/RevModPhys.77.857, URL https://doi.org/doi/10.1103/RevModPhys.77.857

[170] B. R. Webber, *A QCD model for jet fragmentation including soft gluon interference*, Nuclear Physics B, 238 (3), 492–528 (1984), URL http://www.sciencedirect.com/science/article/pii/055032138490333X

[171] B. Andersson, G. Gustafson, G. Ingelman and T. Sjöstrand, *Parton fragmentation and string dynamics*, Physics Reports, 97 (2), 31–145 (1983), doi:10.1016/0370-1573(83)90080-7, URL https://doi.org/10.1016/0370-1573(83)90080-7

[172] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai et al., *GEANT4–a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 506 (3), 250–303 (2003), ISSN 0168-9002, doi:10.1016/S0168-9002(03)01368-8, URL http://www.sciencedirect.com/science/article/pii/S0168900203013688

[173] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. A. Dubois, M. Asai, G. Barrand et al., *Geant4 developments and applications*, IEEE Transactions on Nuclear Science, 53, 270–278 (February 2006), doi:10.1109/tns.2006.869826, URL https://doi.org/10.1109/tns.2006.869826

[174] The ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, The European Physical Journal C, 70 (CERN-PH-EP-2010-044), 823–874 (September 2010), doi:10.1140/epjc/s10052-010-1429-9, URL https://doi.org/10.1140/epjc/s10052-010-1429-9

[175] *Recent developments in Geant4*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 835, 186 – 225 (2016), URL https://doi.org/10.1016/j.nima.2016.06.125

[176] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse and A. Salzburger, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, Technical Report ATL-SOFT-PUB-2007-007, CERN, Geneva (March 2007), URL http://cds.cern.ch/record/1020106

[177] W. Waltenberger, R. Frühwirth and P. Vanlaer, *Adaptive vertex fitting*, Journal of Physics G: Nuclear and Particle Physics, 34 (12) (2007), doi:10.1088/0954-3899/34/12/N01, URL http://doi.org/10.1088/0954-3899/34/12/N01

[178] A. Buckley, D. Kar and K. Nordström, *Fast simulation of detector effects in Rivet*, 8 (2) (February 2020), doi:10.21468/scipostphys.8.2.025, URL https://doi.org/10.21468/scipostphys.8.2.025

[179] The ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, Technical Report ATLAS-CONF-2016-024, CERN, Geneva (June 2016), URL http://cds.cern.ch/record/2157687

[180] The ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13\ TeV$, The European Physical Journal C, 76 (March 2016), doi: 10.1140/epjc/s10052-016-4120-y, URL https://doi.org/10.1140/epjc/s10052-016-4120-y

[181] The ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at* $\sqrt{s} = 13\ TeV$, Technical Report ATL-PHYS-PUB-2015-015, CERN, Geneva (July 2015), URL https://cds.cern.ch/record/2037613

[182] The ATLAS Collaboration, *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*, Technical Report ATLAS-CONF-2015-037, CERN, Geneva (August 2015), URL https://cds.cern.ch/record/2044941

[183] The ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, Technical Report ATL-PHYS-PUB-2015-022, CERN, Geneva (July 2015), URL https://cds.cern.ch/record/2037697

[184] The ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, Technical Report ATL-PHYS-PUB-2016-012, CERN, Geneva (June 2016), URL https://cds.cern.ch/record/2160731

[185] The ATLAS Collaboration, *Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at* $\sqrt{s} = 13\ TeV$, Technical Report ATL-PHYS-PUB-2015-027, CERN (July 2015), URL http://cds.cern.ch/record/2037904

[186] The ATLAS Collaboration, *TopCommonObjects13TeV* (2017), URL https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TopCommonObjects13TeV

[187] The ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at* $\sqrt{s} = 13\ TeV$, Technical Report ATL-PHYS-PUB-2015-026, CERN, Geneva (July 2015), URL http://cds.cern.ch/record/2037717

[188] The ATLAS Collaboration, *ATLAS Muon Combined Performance in 2016* (July 2016), URL https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/MUON-2016-004

[189] E. Ritsch, *Fast Calorimeter Punch-Through Simulation for the ATLAS Experiment* (September 2011), URL http://cds.cern.ch/record/1388275

[190] N. A. Asbah, *Search for the Production of a Standard Model Higgs Boson in Association with Top-Quarks and Decaying into a Pair of Bottom-Quarks with 13 TeV ATLAS Data* (2018), URL https://cds.cern.ch/record/2320703

[191] The DØ Collaboration, *Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using kinematic characteristics of lepton + jets events*, Physical Review D, 76 (9) (November 2007), doi:10.1103/physrevd.76.092007, URL https://doi.org/10.1103/physrevd.76.092007

[192] A. H. Knue, *Measurement of the W-boson Polarisation in Top-Quark Decays with the ATLAS Detector* (2013), doi:11858/00-1735-0000-0001-BB8C-B, URL https://doi.org/11858/00-1735-0000-0001-BB8C-B

[193] W. D. Breaden Madden, *Lepton fakes estimation update: tag-and-probe sample, selection validation and prototype algorithm*, ATLAS Top Backgrounds: Fakes and W/Z+jets (October 2015), URL https://indico.cern.ch/event/449612/contributions/1957379

[194] G. P. Salam, *Towards Jetography*, European Physical Journal C, 67, 637–686 (May 2010), doi:10.1140/epjc/s10052-010-1314-6, URL https://doi.org/10.1140/epjc/s10052-010-1314-6

[195] R. D. Field and R. P. Feynman, *Quark elastic scattering as a source of high-transverse-momentum mesons*, Physical Review D, 15, 2590–2616 (May 1977), doi:10.1103/PhysRevD.15.2590, URL https://doi.org/10.1103/PhysRevD.15.2590

[196] R. D. Field and R. P. Feynman, *A Parametrization of the Properties of Quark Jets*, Nuclear Physics B, 136 (CALT-68-618), 1–76 (April 1978), doi:10.1016/0550-3213(78)90015-9, URL https://doi.org/10.1016/0550-3213(78)90015-9

[197] J. E. Huth et al., *Toward a Standardization of Jet Definitions*, (FERMILAB-CONF-90-249-E) (December 1990), URL https://lss.fnal.gov/archive/preprint/fermilab-conf-90-249-e.shtml

[198] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, Journal of High Energy Physics, 4 (LPTHE-07-03) (April 2008), doi:10.1088/1126-6708/2008/04/063, URL https://doi.org/10.1088/1126-6708/2008/04/063

[199] The ATLAS Collaboration, *Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV in ATLAS 2010 data*, (ATLAS-CONF-2011-032) (March 2011), URL https://cds.cern.ch/record/1337782

[200] D. Karr, *Experimental Particle Physics: Understanding the measurements and searches at the Large Hadron Collider*, IOP Publishing (August 2019), ISBN 978-0-7503-2112-9, doi:10.1088/2053-2563/ab1be6ch2, URL https://doi.org/10.1088/2053-2563/ab1be6ch2

[201] W. H. Parker and W. D. Williams, *Photoexcitation of Quasiparticles in Nonequilibrium Superconductors*, Physical Review D, 29, 924–927 (October 1972), doi:10.1103/PhysRevLett.29.924, URL https://doi.org/10.1103/PhysRevLett.29.924

[202] N. Christ, B. Hasslacher and A. H. Mueller, *Light-Cone Behavior of Perturbation Theory*, Physical Review D, 6, 3543–3562 (December 1972), doi:10.1103/PhysRevD.6.3543, URL https://doi.org/10.1103/PhysRevD.6.3543

[203] T. Sjöstrand and M. Bengtsson, *The Lund Monte Carlo for Jet Fragmentation and e^+e^- Physics – JETSET Version 6.3 – An Update*, Computer Physics Communications, 43 (3), 367–379 (February 1987), doi:10.1016/0010-4655(87)90054-3, URL https://doi.org/10.1016/0010-4655(87)90054-3

[204] S. D. Ellis, *On defining the undefinable: jets*, 17 (10), 1552–1560 (October 1991), doi:10.1088/0954-3899/17/10/011, URL https://doi.org/10.1088/0954-3899/17/10/011

[205] P. Weber, *Jet Algorithms*, LHC Physics: Experimental Potential and Detector Performance (April 2009), URL http://www.kip.uni-heidelberg.de/atlas/seminars/SS2009_JC/jet_algorithms.pdf

[206] The ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Physical Review D, 96, 072002 (October 2017), doi:10.1103/PhysRevD.96.072002, URL https://doi.org/10.1103/PhysRevD.96.072002

[207] The ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV*, The European Physical Journal C, 73 (3) (March 2013), doi:10.1140/epjc/s10052-013-2304-2, URL https://doi.org/10.1140/epjc/s10052-013-2304-2

[208] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan et al., *Calorimeter Clustering Algorithms: Description and Performance*, Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN (April 2008), URL https://cds.cern.ch/record/1099735

[209] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, The European Physical Journal C, 72 (3) (March 2012), doi:10.1140/epjc/s10052-012-1896-2, URL https://doi.org/10.1140/epjc/s10052-012-1896-2

[210] M. Wobisch and T. Wengler, *Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering*, in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998–1999*, PITHA-99-16, 270–279 (1998), doi:10.48550/arXiv.hep-ph/9907280, URL https://doi.org/10.48550/arXiv.hep-ph/9907280

[211] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Physics Letters B, 659 (1), 119 – 126 (2008), ISSN 0370-2693, doi:https://doi.org/10.1016/j.physletb.2007.09.077, URL http://www.sciencedirect.com/science/article/pii/S0370269307011094

[212] The ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, Technical Report ATLAS-CONF-2014-018, CERN, Geneva (May 2014), URL http://cds.cern.ch/record/1700870

[213] G. Aad, J.-F. Arguin, H. Arnold, Y. Bai, P. Berta, J. E. Blanco, V. Boisvert et al., *Performance of b-Jet Identification in the ATLAS Experiment*, Technical Report ATL-COM-PHYS-2012-824, CERN, Geneva (June 2012), URL https://cds.cern.ch/record/1455862

[214] A. Abdesselam et al., *Boosted Objects: A Probe of Beyond the Standard Model Physics*, European Physical Journal J., C71 (SLAC-PUB-15081, FERMILAB-PUB-10-617-CMS) (June 2011), doi:10.1140/epjc/s10052-011-1661-y, URL https://doi.org/10.1140/epjc/s10052-011-1661-y

[215] The ATLAS Collaboration, *Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC*, European Physical Journal J., C79 (5), 375 (August 2019), doi:10.1140/epjc/s10052-019-6847-8, URL https://doi.org/10.1140/epjc/s10052-019-6847-8

[216] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, Journal of High Energy Physics, 03 (MIT-CTP-4191), 015 (March 2011), doi:10.1007/JHEP03(2011)015, URL https://doi.org/10.1007/JHEP03(2011)015

[217] A. Banfi, G. P. Salam and G. Zanderighi, *Phenomenology of event shapes at hadron colliders*, Journal of High Energy Physics, 2010 (6) (June 2010), doi:10.1007/jhep06(2010)038, URL https://doi.org/10.1007/jhep06(2010)038

[218] S. Kluth, P. A. Movilla Fernández, S. Bethke, C. Pahl and P. Pfeifenschneider, *A Measurement of the QCD Color Factors using Event Shape Distributions at $\sqrt{s} = 14$ to 189 GeV*, European Journal of Physics C, 21 (2), 199–210 (June 2001), doi:10.1007/s100520100742, URL https://doi.org/10.1007/s100520100742

[219] S. Bethke, *α_s 2002*, Nuclear Physics B – Proceedings Supplements, 121, 74–81 (June 2003), doi:10.1016/s0920-5632(03)01817-6, URL https://doi.org/10.1016/s0920-5632(03)01817-6

[220] The CMS Collaboration, *Shape, transverse size, and charged-hadron multiplicity of jets in pp collisions at* $\sqrt{s} = 7\ TeV$, Journal of High Energy Physics, 2012 (6) (June 2012), doi:10.1007/jhep06(2012)160, URL https://doi.org/10.1007/jhep06(2012)160

[221] J. Gallicchio and M. D. Schwartz, *Seeing in Color: Jet Superstructure*, Physical Review Letters, 105 (2) (July 2010), doi:10.1103/physrevlett.105.022001, URL https://doi.org/10.1103/physrevlett.105.022001

[222] C. G. Lester and D. J. Summers, *Measuring masses of semi-invisibly decaying particle pairs produced at hadron colliders*, Physics Letters B, 463 (1), 99–103 (September 1999), doi:10.1016/s0370-2693(99)00945-4, URL https://doi.org/10.1016/s0370-2693(99)00945-4

[223] C. Röver, C. Messenger and R. Prix, *Bayesian versus frequentist upper limits*, Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland, 17–20 January 2011, (CERN-2011-006) (September 2011), doi:10.5170/CERN-2011-006.158, URL https://doi.org/10.5170/CERN-2011-006.158

[224] J. Neyman and E. S. Pearson, *On the Problem of the most Efficient Tests of Statistical Hypotheses*, Philosophical Transactions of the Royal Society of London, 231, 289–337 (January 1933), doi:10.1098/rsta.1933.0009, URL http://rsta.royalsocietypublishing.org/content/231/694-706/289.short

[225] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, The Annals of Mathematical Statistics, 9 (1), 60–62 (March 1938), ISSN 00034851

[226] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Transactions of the American Mathematical Society, 54 (3), 426–482 (November 1943), ISSN 00029947

[227] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C, 71 (February 2011), doi:10.1140/epjc/s10052-011-1554-0, URL https://doi.org/10.1140/epjc/s10052-011-1554-0

[228] The ATLAS Collaboration, the CMS Collaboration and the LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011*, Technical Report CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, CERN, Geneva (August 2011), doi:10.1088/0954-3899/28/10/313, URL https://doi.org/10.1088/0954-3899/28/10/313

[229] O. Behnke, K. Kröninger, G. Schott and T. Schörner-Sadenius (editors), *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods* (June 2013), doi:10.1002/9783527653416, URL https://doi.org/10.1002/9783527653416

[230] A. L. Read, *Modified frequentist analysis of search results (the CL_s method)*, (CERN-OPEN-2000-205) (January 2000), doi:10.5170/CERN-2000-005.81, URL http://cdsweb.cern.ch/record/451614

[231] A. L. Read, *Presentation of search results: the CL_s technique*, Journal of Physics G: Nuclear and Particle Physics, 28 (10), 2693 (2002), doi:10.1088/0954-3899/28/10/313, URL https://doi.org/10.1088/0954-3899/28/10/313

[232] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, The European Physical Journal C, 70, 525–530 (October 2010), doi:10.1140/epjc/s10052-010-1470-8, URL https://doi.org/10.1140/epjc/s10052-010-1470-8

[233] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert and J. Winter, *Event generation with SHERPA 1.1*, Journal of High Energy Physics, (2), 007–007 (February 2009), doi:10.1088/1126-6708/2009/02/007, URL https://doi.org/10.1088/1126-6708/2009/02/007

[234] 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), 152–155 (April 2001), doi:10.1016/s0168-9002(01)00089-4, URL https://doi.org/10.1016/s0168-9002(01)00089-4

[235] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, Journal of High Energy Physics, (7) (July 2014), doi:10.1007/jhep07(2014)079, URL https://doi.org/10.1007/jhep07(2014)079

[236] The ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, Technical report, CERN, Geneva (November 2014), URL https://cds.cern.ch/record/1966419

[237] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, Journal of High Energy Physics, 2013 (3) (March 2013), doi:10.1007/jhep03(2013)015, URL https://doi.org/10.1007/jhep03(2013)015

[238] R. Raitio and W. W. Wada, *Higgs-boson production at large transverse momentum in quantum chromodynamics*, Physical Review D, 19 (3), 941–944 (February 1979), doi:10.1103/physrevd.19.941, URL https://doi.org/10.1103/physrevd.19.941

[239] W. Beenakker, S. Dittmaier, M. Krämer, B. Plümper, M. Spira and P. Zerwas, *NLO QCD corrections to production in hadron collisions*, Nuclear Physics B, 653 (1-2), 151–203 (March 2003), doi:10.1016/s0550-3213(03)00044-0, URL https://doi.org/10.1016/s0550-3213(03)00044-0

[240] Y. Zhang, W.-G. Ma, R.-Y. Zhang, C. Chen and L. Guo, *QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider*, Physics Letters B, 738, 1–5 (November 2014), doi:10.1016/j.physletb.2014.09.022, URL https://doi.org/10.1016/j.physletb.2014.09.022

[241] S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao and M. Zaro, *Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons*, Journal of High Energy Physics, 2015 (6) (June 2015), doi:10.1007/jhep06(2015)184, URL https://doi.org/10.1007/jhep06(2015)184

[242] A. Djouadi, J. Kalinowski and M. Spira, *HDECAY: a program for Higgs boson decays in the Standard Model and its supersymmetric extension*, Computer Physics Communications, 108 (1), 56–74 (January 1998), doi:10.1016/s0010-4655(97)00123-9, URL https://doi.org/10.1016/s0010-4655(97)00123-9

[243] P. Nason, *A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms*, Journal of High Energy Physics, 2004 (11), 040–040 (November 2004), doi:10.1088/1126-6708/2004/11/040, URL https://doi.org/10.1088/1126-6708/2004/11/040

[244] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, Journal of High Energy Physics, 2007 (11), 070–070 (November 2007), doi:10.1088/1126-6708/2007/11/070, URL https://doi.org/10.1088/1126-6708/2007/11/070

[245] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, Journal of High Energy Physics, 2010 (6) (June 2010), doi:10.1007/jhep06(2010)043, URL https://doi.org/10.1007/jhep06(2010)043

[246] J. M. Campbell, R. K. Ellis, P. Nason and E. Re, *Top-pair production and decay at NLO matched with parton showers*, Journal of High Energy Physics, 2015 (4) (April 2015), doi:10.1007/jhep04(2015)114, URL https://doi.org/10.1007/jhep04(2015)114

[247] The ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, Technical report, CERN, Geneva (September 2016), URL https://cds.cern.ch/record/2216168

[248] 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), 2930–2938 (November 2014), doi:10.1016/j.cpc.2014.06.021, URL https://doi.org/10.1016/j.cpc.2014.06.021

[249] M. Cacciari, M. Czakon, M. Mangano, A. Mitov and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Physics Letters B, 710 (4-5), 612–622 (April 2012), doi:10.1016/j.physletb.2012.03.013, URL https://doi.org/10.1016/j.physletb.2012.03.013

[250] M. Guzzi, P. Nadolsky, E. Berger, H.-L. Lai, F. Olness and C.-P. Yuan, *CT10 parton distributions and other developments in the global QCD analysis* (January 2011), URL https://doi.org/10.48550/arXiv.1101.0561

[251] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky and W.-K. Tung, *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, Journal of High Energy Physics, (7), 012–012 (July 2002), doi:10.1088/1126-6708/2002/07/012, URL https://doi.org/10.1088/1126-6708/2002/07/012

[252] M. Bähr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Plätzer et al., *Herwig++ physics and manual*, The European Physical Journal C, 58 (4), 639–707 (November 2008), doi:10.1140/epjc/s10052-008-0798-9, URL https://doi.org/10.1140/epjc/s10052-008-0798-9

[253] P. Z. Skands, *Tuning Monte Carlo generators: The Perugia tunes*, Physical Review D, 82 (7) (October 2010), doi:10.1103/physrevd.82.074018, URL https://doi.org/10.1103/physrevd.82.074018

[254] S. Frixione, E. Laenen, P. Motylinski, C. White and B. R. Webber, *Single-top hadroproduction in association with a W boson*, Journal of High Energy Physics, (07) (July 2008), doi:10.1088/1126-6708/2008/07/029, URL https://doi.org/10.1088/1126-6708/2008/07/029

[255] The ATLAS Collaboration, *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, Technical report, CERN, Geneva (October 2014), URL https://cds.cern.ch/record/1951336

[256] G. Aad, J. Adelman, N. Asbah, S. K. Boutle, N. Brahimi, T. P. Calvet, R. Caminal Armadans et al., *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into* $b\bar{b}$ *in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Technical Report ATL-COM-PHYS-2017-079, CERN (May 2018), URL https://cds.cern.ch/record/2244360

[257] J. A. Raine, *Evidence for the production of a Higgs boson in association with two top quarks with the ATLAS detector: A search in the* $H \to b\bar{b}$ *channel and in combination with other Higgs boson decays at* $\sqrt{s} = 13$ *TeV* (2018), URL http://cds.cern.ch/record/2316953

[258] The ATLAS Collaboration, *Measurement of the Higgs boson decaying to b-quarks produced in association with a top-quark pair in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, (ATLAS-CONF-2020-058) (November 2020), URL http://cds.cern.ch/record/2743685

[259] The ATLAS Collaboration, *Search for flavour-changing neutral current top quark decays* $t \to Hq$ *in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, Journal of High Energy Physics, (CERN-

PH-EP-2015-229) (September 2015), doi:10.1007/JHEP12(2015)061, URL https://cds.cern.ch/record/2053792

[260] The ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, The European Physical Journal C, 75 (CERN-PH-EP-2015-047) (March 2015), URL https://cds.cern.ch/record/2001975

[261] *ROOT: A Data Analysis Framework* (November 2017), URL https://root.cern.ch

[262] R. Brun, *Re: What does ROOT stand for?*, RootTalk (May 1998), URL https://root.cern.ch/root/roottalk/roottalk98/0718.html

[263] R. Brun, *40 Years of Large Scale Data Analysis in HEP - the HBOOK, PAW and ROOT Story. 40 Years of Large Scale Data Analysis in HEP – the HBOOK, PAW and ROOT Story* (October 2017), URL http://cds.cern.ch/record/2287161

[264] A. Buckley, J. Butterworth, D. Grellscheid, H. Hoeth, L. Lönnblad, J. Monk, H. Schulz et al., *Rivet user manual*, Computer Physics Communications, 184 (12), 2803–2819 (June 2013), doi:10.1016/j.cpc.2013.05.021, URL https://doi.org/10.1016/j.cpc.2013.05.021

[265] E. Conte and B. Fuks, *Confronting new physics theories to LHC data with MADANALYSIS 5*, International Journal of Modern Physics A, 33 (28), 1830027 (October 2018), doi:10.1142/s0217751x18300272, URL https://doi.org/10.1142/s0217751x18300272

[266] A. Koppanyi and M. Metcalf, *A schematic comparison of HTV and its successor PAW*, CERN Computer Newsletter, (CERN-CNL-186) (1986), URL http://cds.cern.ch/record/1292499

[267] R. K. Böck, R. Brun, O. Couet, J. C. Marin, R. Nierhaus, L. Pape, N. Saumon et al., *PAW – Towards a Physics Analysis Workstation*, Computer Physics Communications, 45 (CERN-DD-87-5) (February 1987), doi:10.1016/0010-4655(87)90154-8, URL https://cds.cern.ch/record/175134

[268] *Control Data at CERN*, (CERN-VIDEO-1965-002-001) (1965), doi:10.17181/cds.43172, URL https://doi.org/10.17181/cds.43172

[269] *TTHbbLeptonic-02-04-23-03*, CERN Central SVN Service (January 2017), URL https://svnweb.cern.ch/cern/wsvn/atlasphys-hsg8/Physics/Higgs/HSG8/AnalysisCode/TTHbbLeptonic/tags/TTHbbLeptonic-02-04-23-03

[270] *Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Journal of High Energy Physics, (30) (January 2019), doi:10.1007/jhep01(2019)030, URL https://doi.org/10.1007/jhep01(2019)030

[271] The ATLAS Collaboration, *ATLAS Data Access Policy*, CERN Open Data Portal (2014), doi: 10.7483/OPENDATA.ATLAS.T9YR.Y7MZ, URL http://opendata.cern.ch/record/413

[272] W. Breaden Madden, *TTHbbLeptonic example run 2015-09-15T1612Z*, asciinema (September 2015), URL https://asciinema.org/a/dsmpyeis419jm3fdb5co76qzo

[273] W. Breaden Madden, *Quick Update: Run-2 TTHbbLeptonic AnalysisTop package status, 8 TeV–13 TeV comparisons*, ATLAS Htop(bb) (April 2015), URL https://indico.cern.ch/event/364577/contribution/1/material/slides/0.pdf

[274] J. A. Raine, *$t\bar{t}H$ Run 2 Analysis Framework Update*, ATLAS Htop(bb) (April 2015), URL https://indico.cern.ch/event/364576/contributions/1779310

[275] S. Guindon, E. Shabalina, J. Adelman, M. Alhroob, A. Basye, J. Bouffard, M. Casolino et al., *Search for the Standard Model Higgs boson produced in association with top quarks in pp collisions at 8 TeV with the ATLAS detector at the LHC*, (ATL-COM-PHYS-2014-1471) (November 2014), URL https://cds.cern.ch/record/1744007

[276] The ATLAS Collaboration, *Measurements of top-quark pair differential cross-sections in the lepton + jets channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, 2017 (11) (November 2017), doi:10.1007/jhep11(2017)191, URL https://doi.org/10.1007/jhep11(2017)191

[277] The ATLAS Collaboration, *Measurements of $t\bar{t}$ differential cross-sections of highly boosted top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, 98 (1) (July 2018), doi:10.1103/physrevd.98.012003, URL https://doi.org/10.1103/physrevd.98.012003

[278] M. Lassnig, V. Garonne, G. A. Stewart, M. Barisits, T. Beermann, R. Vigne, C. Serfon et al., *The ATLAS Data Management Software Engineering Process*, Technical report, CERN, Geneva (November 2013), doi:10.1088/1742-6596/513/5/052017, URL https://cds.cern.ch/record/1630190

[279] The ATLAS Collaboration, *TRExFitter documentation* (2021), URL https://trexfitter-docs.web.cern.ch/trexfitter-docs

[280] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling* (June 2003), doi:10.48550/arxiv.physics/0306116, URL https://doi.org/10.48550/arxiv.physics/0306116

[281] L. Moneta, K. Belasco, K. Cranmer, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke et al., *The RooStats Project*, PoS, ACAT2010 (September 2010), doi:10.22323/1.093.0057, URL https://doi.org/10.22323/1.093.0057

[282] G. Schott, *RooStats for Searches* (March 2012), doi:10.48550/arXiv.1203.1547, URL https://doi.org/10.48550/arXiv.1203.1547

[283] J. Albrecht et al., *A Roadmap for HEP Software and Computing R&D for the 2020s*, (HSF-CWP-2017-01, HSF-CWP-2017-001, FERMILAB-PUB-17-607-CD) (2017), doi:10.1007/s41781-018-0018-8, URL https://doi.org/10.1007/s41781-018-0018-8

[284] *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Physics Letters B, 784 (CERN-EP-2018-138), 173–191 (September 2018), ISSN 0370-2693, doi:10.1016/j.physletb.2018.07.035, URL http://cds.cern.ch/record/2621167

[285] K. Cranmer, G. Lewis, L. Moneta, A. Shibata and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, Technical Report CERN-OPEN-2012-016 (2012), URL https://cds.cern.ch/record/1456844

[286] F. James and M. Winkler, *Root Reference Guide: Minuit2* (2021), URL https://root.cern.ch/doc/master/Minuit2Page.html

[287] F. James and M. Roos, *Minuit – a system for function minimization and analysis of the parameter errors and correlations*, Computer Physics Communications, 10 (6), 343–367 (December 1975), doi:10.1016/0010-4655(75)90039-9, URL https://doi.org/10.1016/0010-4655(75)90039-9

[288] M. Fenton, *Boosting to the top: measurements of boosted top quarks and Higgs bosons with the ATLAS detector at the Large Hadron Collider* (2019), URL http://theses.gla.ac.uk/75176

[289] A. Romero, D. Whiteson, M. Fenton, J. Collado and P. Baldi, *Safety of Quark/Gluon Jet Classification* (March 2021), doi:10.48550/arXiv.2103.09103, URL https://doi.org/10.48550/arXiv.2103.09103

[290] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher and D. J. Schwab, *A high-bias, low-variance introduction to Machine Learning for physicists*, Physics Reports, 810, 1–124 (May 2019), doi:10.1016/j.physrep.2019.03.001, URL https://doi.org/10.1016/j.physrep.2019.03.001

[291] F. Chollet, *Deep Learning with Python*, Manning Publications, Shelter Island, NY (2018), ISBN 9781617294433

[292] F. Chollet, *On the Measure of Intelligence*, arXiv e-prints, arXiv:1911.01547 (November 2019), doi:10.48550/arXiv.1911.01547, URL https://doi.org/10.48550/arXiv.1911.01547

[293] C. Verkerk (editor), *1991 CERN School of Computing: Ystads Saltsjöbad Hotel, Ystad, Sweden 23 August – 2 September 1991*, CERN, CERN, Geneva (May 1992), doi:10.5170/CERN-1992-002, URL https://cds.cern.ch/record/215507

[294] C. van der Malsburg, *Frank Rosenblatt: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, in *Brain Theory*, 245–248, Springer Berlin Heidelberg (1986), doi:10.1007/978-3-642-70911-1_20, URL https://doi.org/10.1007/978-3-642-70911-1_20

[295] C. M. Bishop, *Neural Networks for Pattern Recognition*, Advanced Texts in Econometrics, Oxford University Press (1996), ISBN 9780198538646

[296] L. Breiman, J. H. Friedman, R. A. Olshen and C. J. Stone, *Classification And Regression Trees*, Routledge (October 2017), doi:10.1201/9781315139470, URL https://doi.org/10.1201/9781315139470

[297] The OPAL Collaboration, *A Measurement of the Production of $D^{*\pm}$ Mesons on the Z^0 Resonance*, Zeitschrift für Physik C, 67 (December 1994), doi:10.1007/BF01564819, URL http://cds.cern.ch/record/275063

[298] The ALEPH Collaboration, *Determination of $|V_{ub}|$ from the measurement of the inclusive charmless semileptonic branching ratio of b-hadrons*, The European Physical Journal C, 6 (May 1998), doi:10.1007/s100529801031, URL https://cds.cern.ch/record/355437

[299] W. S. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, The Bulletin of Mathematical Biophysics, 5 (4), 115–133 (1943), doi:10.1007/BF02478259, URL https://doi.org/10.1007/BF02478259

[300] D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory*, John Wiley & Sons, Inc. (1949)

[301] F. Rosenblatt, *The Perceptron: A Perceiving and Recognizing Automaton (Project PARA)*, (85-460-1) (January 1957), URL https://blogs.umass.edu/brain-wars/files/2016/03/rosenblatt-1957.pdf

[302] J. C. Hay, B. E. Lynch and D. R. Smith, *Mark I Perceptron Operator's Manual (Project PARA)*, (VG-1196-6-5) (February 1960), URL https://archive.org/details/DTIC_AD0236965

[303] M. Olazaran, *A Sociological Study of the Official History of the Perceptrons Controversy*, Social Studies of Science, 26 (3), 611–659 (August 1996), doi:10.1177/030631296026003005, URL https://doi.org/10.1177/030631296026003005

[304] J. R. Quinlan, *Induction of decision trees*, Machine Learning, 1 (1), 81–106 (March 1986), doi: 10.1007/bf00116251, URL https://doi.org/10.1007/bf00116251

[305] J. L. Hodges, *The significance probability of the Smirnov two-sample test*, Arkiv för Matematik, 3 (43), 469–486 (January 1958), doi:10.1007/bf02589501, URL https://doi.org/10.1007/bf02589501

[306] R. M. Tavares, A. Mendelsohn, Y. Grossman, C. H. Williams, M. Shapiro, Y. Trope and D. Schiller, *A Map for Social Navigation in the Human Brain*, Neuron, 87 (1), 231–243 (July 2015), doi: 10.1016/j.neuron.2015.06.011, URL https://doi.org/10.1016/j.neuron.2015.06.011

[307] A. Speakman, *Place cells in the brain: evidence for a cognitive map*, Science Progress, 71 (4 (284)), 511–530 (1987), ISSN 00368504, 20477163, URL http://www.jstor.org/stable/43420697

[308] T. Mikolov, W.–t. Yih and G. Zweig, *Linguistic Regularities in Continuous Space Word Representations*, in *Human Language Technologies – North American Chapter of the ACL (Association for Computational Linguistics) (HLT-NAACL)*, Association for Computational Linguistics (June 2013), URL https://aclanthology.org/N13-1090

[309] P. Baldi, P. Sadowski and D. Whiteson, *Searching for exotic particles in high-energy physics with deep learning*, Nature Communications, 5 (1) (July 2014), doi:10.1038/ncomms5308, URL https://doi.org/10.1038/ncomms5308

[310] L. van der Maaten and G. Hinton, *Visualizing Data using t-SNE*, Journal of Machine Learning Research, 9, 2579–2605 (November 2008), URL http://www.jmlr.org/papers/v9/vandermaaten08a.html

[311] L. van der Maaten, *Visualizing Data Using t-SNE*, Google Tech Talks (July 2013), URL https://www.youtube.com/watch?v=RJVL80Gg3lA

[312] L. McInnes, J. Healy, N. Saul and L. Großberger, *UMAP: Uniform Manifold Approximation and Projection*, Journal of Open Source Software, 3 (29), 861 (September 2018), doi:10.21105/joss.00861, URL https://doi.org/10.21105/joss.00861

[313] L. McInnes, J. Healy and J. Melville, *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*, arXiv e-prints, arXiv:1802.03426 (February 2018), doi:10.48550/arXiv.1802.03426, URL https://doi.org/10.48550/arXiv.1802.03426

[314] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. Courville, *Improved Training of Wasserstein GANs* (March 2017), doi:10.48550/arXiv.1704.00028, URL https://doi.org/10.48550/arXiv.1704.00028

[315] C. Shimmin, P. Sadowski, P. Baldi, E. Weik, D. Whiteson, E. Goul and A. Søgaard, *Decorrelated Jet Substructure Tagging using Adversarial Neural Networks*, Physical Review D, 96 (7) (October 2017), doi:10.1103/PhysRevD.96.074034, URL https://doi.org/10.1103/PhysRevD.96.074034

[316] A. Aurisano, A. Radovic, D. Rocco, A. Himmel, M. D. Messier, E. Niner, G. Pawloski et al., *A Convolutional Neural Network Neutrino Event Classifier*, Journal of Instrumentation, 11 (9) (September 2016), doi:10.1088/1748-0221/11/09/P09001, URL https://www.doi.org/10.1088/1748-0221/11/09/P09001

[317] A. Alves, *Stacking machine learning classifiers to identify Higgs bosons at the LHC*, Journal of Instrumentation, 12 (5) (May 2017), doi:10.1088/1748-0221/12/05/t05005, URL https://doi.org/10.1088/1748-0221/12/05/t05005

[318] J. N. Morgan and J. A. Sonquist, *Problems in the Analysis of Survey Data, and a Proposal*, Journal of the American Statistical Association, 58 (302), 415–434 (June 1963), doi:10.1080/01621459.1963.10500855, URL https://doi.org/10.1080/01621459.1963.10500855

[319] R. A. Fisher, *The Use of Multiple Measurements in Taxonomic Problems*, Annals of Eugenics, 7 (2), 179–188 (September 1936), doi:10.1111/j.1469-1809.1936.tb02137.x, URL https://doi.org/10.1111/j.1469-1809.1936.tb02137.x, Articles published by the *Annals of Eugenics* are included here solely for scholarly use. The work of eugenicists was pervaded by prejudice against racial, ethnic, disabled and other groups. The referencing of this material for scholarly research purposes is not an endorsement of those views nor is it a promotion of eugenics in any way.

[320] J. H. Friedman, *Greedy Function Approximation: A Gradient Boosting Machine*, Annals of Statistics, 29, 1189–1232 (2000), doi:10.2307/2699986, URL https://statweb.stanford.edu/~jhf/ftp/trebst.pdf

[321] M. Shoaran, B. A. Haghi, M. Taghavi, M. Farivar and A. Emami-Neyestanak, *Energy-Efficient Classification for Resource-Constrained Biomedical Applications*, IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 8 (4), 693–707 (December 2018), doi:10.1109/jetcas.2018.2844733, URL https://doi.org/10.1109/jetcas.2018.2844733

[322] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, Computing Research Repository (March 2016), URL http://arxiv.org/abs/1603.02754

[323] T. Chen and T. He, *Boosted Trees for Higgs Discovery* (May 2015), URL https://indico.cern.ch/event/382895/contributions/910921

[324] A. L. Hodgkin and A. F. Huxley, *A quantitative description of membrane current and its application to conduction and excitation in nerve*, The Journal of Physiology, 117 (4), 500–544 (August 1952), URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1392413

[325] T. Heimburg and A. D. Jackson, *On soliton propagation in biomembranes and nerves*, Proceedings of the National Academy of Sciences of the United States of America, 102, 9790–9795 (August 2005), doi:10.1073/pnas.0503823102, URL https://doi.org/10.1073/pnas.0503823102

[326] R. Appali, U. van Rienen and T. Heimburg, *A Comparison of the Hodgkin–Huxley Model and the Soliton Theory for the Action Potential in Nerves*, in *Advances in Planar Lipid Bilayers and Liposomes Volume 16*, 275–299, Elsevier (2012), doi:10.1016/b978-0-12-396534-9.00009-x, URL https://doi.org/10.1016/b978-0-12-396534-9.00009-x

[327] C. Peterson, T. S. Rögnvaldsson and L. Lönnblad, *JETNET 3.0: A Versatile Artificial Neural Network Package*, Computer Physics Communications, 81 (December 1993), doi:10.1016/0010-4655(94)90120-1, URL http://cds.cern.ch/record/259970

[328] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by back-propagating errors*, Nature, 323 (6088), 533–536 (October 1986), doi:10.1038/323533a0, URL https://doi.org/10.1038/323533a0

[329] F. van Veen, *A mostly complete chart of Neural Networks*, The Azimov Institute (September 2016), URL http://www.asimovinstitute.org/neural-network-zoo

[330] E. Hoffer, I. Hubara and D. Soudry, *Train longer, generalize better: closing the generalization gap in large batch training of neural networks*, arXiv e-prints, arXiv:1705.08741 (May 2017), doi:10.48550/arXiv.1705.08741, URL https://doi.org/10.48550/arXiv.1705.08741

[331] S. L. Smith, P.-J. Kindermans, C. Ying and Q. V. Le, *Don't Decay the Learning Rate, Increase the Batch Size*, arXiv e-prints, arXiv:1711.00489 (November 2017), doi:10.48550/arXiv.1711.00489, URL https://doi.org/10.48550/arXiv.1711.00489

[332] A. Krizhevsky, I. Sutskever and G. E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, Communications of the Association for Computing Machinery, 60 (6), 84–90 (May 2017), doi:10.1145/3065386, URL https://doi.org/10.1145/3065386

[333] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning*, Nature, 521 (7553), 436–444 (May 2015), doi:10.1038/nature14539, URL https://doi.org/10.1038/nature14539

[334] P. Sadowski, *Deep Learning for Experimental Physics* (2016), URL https://escholarship.org/uc/item/5qb7z23t

[335] H. W. Lin, M. Tegmark and D. Rolnick, *Why Does Deep and Cheap Learning Work So Well?*, Journal of Statistical Physics, 168 (6), 1223–1247 (July 2017), doi:10.1007/s10955-017-1836-5, URL https://doi.org/10.1007/s10955-017-1836-5

[336] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, *Backpropagation Applied to Handwritten Zip Code Recognition*, Neural Computation, 1 (4), 541–551 (December 1989), doi:10.1162/neco.1989.1.4.541, URL https://doi.org/10.1162/neco.1989.1.4.541

[337] C. S. Sherrington, *Observations on the scratch-reflex in the spinal dog.*, The Journal of Physiology, 34 (1-2), 1–50 (March 1906), doi:10.1113/jphysiol.1906.sp001139, URL https://doi.org/10.1113/jphysiol.1906.sp001139, This research may have involved unethical practices and is included here solely for scholarly use. The referencing of this material for scholarly research purposes is not an endorsement of unethical methodology in any way.

[338] H. K. Hartline, *The Response of Single Optic Nerve Fibres of the Vertebrate Eye to Illumination of the Retina*, American Journal of Physiology, 121 (2), 400–415 (January 1938), doi:10.1152/ajplegacy.1938.121.2.400, URL https://doi.org/10.1152/ajplegacy.1938.121.2.400, This research may have involved unethical practices and is included here solely for scholarly use. The referencing of this material for scholarly research purposes is not an endorsement of unethical methodology in any way.

[339] T. Hannagan, A. Agrawal, L. Cohen and S. Dehaene, *Emergence of a compositional neural code for written words: Recycling of a convolutional neural network for reading*, 118 (46) (November 2021), doi:10.1073/pnas.2104779118, URL https://doi.org/10.1073/pnas.2104779118

[340] S. Bai, J. Z. Kolter and V. Koltun, *An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling* (March 2018), doi:10.48550/arXiv.1803.01271, URL https://doi.org/10.48550/arXiv.1803.01271

[341] Y. Chen, H. Fang, B. Xu, Z. Yan, Y. Kalantidis, M. Rohrbach, S. Yan et al., *Drop an Octave: Reducing Spatial Redundancy in Convolutional Neural Networks with Octave Convolution*, arXiv e-prints, arXiv:1904.05049 (April 2019), doi:10.48550/arXiv.1904.05049, URL https://doi.org/10.48550/arXiv.1904.05049

[342] M. Weiler, P. Forré, E. Verlinde and M. Welling, *Coordinate Independent Convolutional Networks – Isometry and Gauge Equivariant Convolutions on Riemannian Manifolds* (2021), doi:10.48550/arXiv.2106.06020, URL https://doi.org/10.48550/arXiv.2106.06020

[343] C. Aguayo, *Machine Learning Chapter: Text Classification* (February 2019), URL https://docs.google.com/presentation/d/1NQpJtkD8PhMmER4fw4WO0VwaD9I-s20gWnypwbkx5wk

[344] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan et al., *Going Deeper with Convolutions*, Computing Research Repository (September 2014), URL http://arxiv.org/abs/1409.4842

[345] T. Lin, P. Goyal, R. B. Girshick, K. He and P. Dollár, *Focal Loss for Dense Object Detection*, Computing Research Repository, abs/1708.02002 (2017), URL http://arxiv.org/abs/1708.02002

[346] G. Louppe, *Adversarial Games for Particle Physics*, Accelerating the Search for Dark Matter with Machine Learning (January 2018), URL https://indico.cern.ch/event/664842/contributions/2720905

[347] M. Naaman, *On the tight constant in the multivariate Dvoretzky–Kiefer–Wolfowitz inequality*, Statistics & Probability Letters, 173, 109088 (June 2021), doi:10.1016/j.spl.2021.109088, URL https://doi.org/10.1016/j.spl.2021.109088

[348] G. David Marvin, *Signal Detection Theory and Psychophysics*, John Wiley & Sons (1966), ISBN 0471324205

[349] B. Knyazev, M. Drozdzal, G. W. Taylor and A. Romero-Soriano, *Parameter Prediction for Unseen Deep Architectures* (2021), doi:10.48550/arXiv.2110.13100, URL https://doi.org/10.48550/arXiv.2110.13100

[350] T. Dozat, *Incorporating Nesterov Momentum into Adam*, International Conference on Learning Representations (ICLR) Workshop, URL http://cs229.stanford.edu/proj2015/054_report.pdf

[351] D. P. Kingma and J. L. Ba, *Adam: a Method for Stochastic Optimization*, International Conference on Learning Representations (ICLR) (2015), doi:11245/1.505367, URL https://doi.org/11245/1.505367

[352] S. Ruder, *An overview of gradient descent optimization algorithms* (September 2016), doi:10.48550/arXiv.1609.04747, URL https://doi.org/10.48550/arXiv.1609.04747

[353] N. Developers, *NumPy Reference*, SciPy (January 2019), URL https://docs.scipy.org/doc/numpy/reference

[354] A. Thompson, *An evolved circuit, intrinsic in silicon, entwined with physics.* (1997), doi:10.1007/3-540-63173-9_61, URL https://doi.org/10.1007/3-540-63173-9_61

[355] D. Guest, K. Cranmer and D. Whiteson, *Deep Learning and Its Application to LHC Physics*, 68 (1), 161–181 (October 2018), doi:10.1146/annurev-nucl-101917-021019, URL https://doi.org/10.1146/annurev-nucl-101917-021019

[356] V. Estrade, C. Germain, I. Guyon and D. Rousseau, *Systematics aware learning: a case study in High Energy Physics*, 3rd ATLAS Machine Learning Workshop (October 2018), URL https://docs.google.com/presentation/d/1NQpJtkD8PhMmER4fw4WO0VwaD9I-s20gWnypwbkx5wk

[357] V. Estrade, C. Germain, I. Guyon and D. Rousseau, *Systematics aware learning: a case study in High Energy Physics* (April 2018), URL https://hal.inria.fr/hal-01715155

[358] Y. Gal, *Uncertainty in Deep Learning* (2017), URL http://idiscover.lib.cam.ac.uk/permalink/f/t9gok8/44CAM_ALMA21582084170003606

[359] L.-G. Xia, *QBDT, a new boosting decision tree method with systematical uncertainties into training for High Energy Physics*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 930, 15–26 (June 2019), doi:10.1016/j.nima.2019.03.088, URL https://doi.org/10.1016/j.nima.2019.03.088

[360] I. Osband, Z. Wen, M. Asghari, M. Ibrahimi, X. Lu and B. V. Roy, *Epistemic Neural Networks*, Computing Research Repository (July 2021), doi:10.48550/arXiv.2107.08924, URL https://doi.org/10.48550/arXiv.2107.08924

[361] P. de Castro and T. Dorigo, *INFERNO: Inference-Aware Neural Optimisation*, Computer Physics Communications, 244, 170–179 (November 2019), doi:10.1016/j.cpc.2019.06.007, URL https://doi.org/10.1016/j.cpc.2019.06.007

[362] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Improving neural networks by preventing co-adaptation of feature detectors*, Computing Research Repository, abs/1207.0580 (2012), URL http://arxiv.org/abs/1207.0580

[363] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, Journal of Machine Learning Research, 15, 1929–1958 (2014), URL http://jmlr.org/papers/v15/srivastava14a.html

[364] A. Krizhevsky, I. Sutskever and G. E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, in *Proceedings of the 25th International Conference on Neural Information Processing Systems – Volume 1*, NIPS '12, 1097–1105, Curran Associates Inc. (2012), URL http://dl.acm.org/citation.cfm?id=2999134.2999257

[365] S. Park and N. Kwak, *Analysis on the Dropout Effect in Convolutional Neural Networks*, 189–204, Springer International Publishing (2017), ISBN 978-3-319-54184-6, doi:10.1007/978-3-319-54184-6_12, URL https://doi.org/10.1007/978-3-319-54184-6_12

[366] G. Ghiasi, T.-Y. Lin and Q. V. Le, *DropBlock: A regularization method for convolutional networks*, arXiv e-prints, arXiv:1810.12890 (October 2018), doi:10.48550/arXiv.1810.12890, URL https://doi.org/10.48550/arXiv.1810.12890

[367] Y. Gal and Z. Ghahramani, *Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference*, arXiv e-prints, arXiv:1506.02158 (June 2015), doi:10.48550/arXiv.1506.02158, URL https://doi.org/10.48550/arXiv.1506.02158

[368] P. Sollich and A. Krogh, *Learning with ensembles: How overfitting can be useful* (1995), URL http://papers.nips.cc/paper/1044-learning-with-ensembles-how-overfitting-can-be-useful.pdf, Neural Information Processing Systems, Colorado, November 1995

[369] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, Adaptive Computation and Machine Learning, The MIT Press (2016), ISBN 0262035618

[370] D. H. Wolpert, *Stacked generalization*, Neural Networks, 5 (2), 241–259 (January 1992), doi:10.1016/s0893-6080(05)80023-1, URL https://doi.org/10.1016/s0893-6080(05)80023-1

[371] *IJCAI-17 Workshop on Explainable AI (XAI) Proceedings*, International Joint Conference on Artificial Intelligence 2017 (August 2017), URL http://www.intelligentrobots.org/files/IJCAI2017/IJCAI-17_XAI_WS_Proceedings.pdf

[372] A. Mahendran and A. Vedaldi, *Visualizing Deep Convolutional Neural Networks Using Natural Pre-images*, International Journal of Computer Vision, 120 (3), 233–255 (May 2016), doi:10.1007/s11263-016-0911-8, URL https://doi.org/10.1007/s11263-016-0911-8

[373] A. M. Nguyen, A. Dosovitskiy, J. Yosinski, T. Brox and J. Clune, *Synthesizing the preferred inputs for neurons in neural networks via deep generator networks*, Computing Research Repository, abs/1605.09304 (2016), doi:10.48550/arXiv.1605.09304, URL https://doi.org/10.48550/arXiv.1605.09304

[374] K. Simonyan, A. Vedaldi and A. Zisserman, *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*, arXiv e-prints, arXiv:1312.6034 (December 2013), doi:10.48550/arXiv.1312.6034, URL https://doi.org/10.48550/arXiv.1312.6034

[375] M. D. Zeiler and R. Fergus, *Visualizing and Understanding Convolutional Networks*, arXiv e-prints, arXiv:1311.2901 (November 2013), doi:10.48550/arXiv.1311.2901, URL https://doi.org/10.48550/arXiv.1311.2901

[376] D. Müllner, *Modern hierarchical, agglomerative clustering algorithms*, arXiv e-prints, arXiv:1109.2378 (September 2011), doi:10.48550/arXiv.1109.2378, URL https://doi.org/10.48550/arXiv.1109.2378

[377] X. Liu, X.-H. Zhu, P. Qiu and W. Chen, *A Correlation-Matrix-Based Hierarchical Clustering Method for Functional Connectivity Analysis*, Journal of Neuroscience Methods, 211 (1), 94–102 (October 2012), doi:10.1016/j.jneumeth.2012.08.016, URL https://doi.org/10.1016/j.jneumeth.2012.08.016

[378] D. Freedman and P. Diaconis, *On the histogram as a density estimator: L22 theory*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 57 (4), 453–476 (December 1981), doi:10.1007/bf01025868, URL https://doi.org/10.1007/bf01025868

[379] L. de Oliveira, *Learning Particle Physics by Example: Accelerating Science with Generative Adversarial Networks*, GPUTech (May 2017), URL http://on-demand.gputechconf.com/gtc/2017/presentation/s7666-luke-de-oliveira-particle-physics-generative-adversarial-networks.pdf

[380] M. Paganini, L. de Oliveira and B. Nachman, *Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters*, Physical Review Letters, 120 (4) (January 2018), doi:10.1103/physrevlett.120.042003, URL https://doi.org/10.1103/physrevlett.120.042003

[381] D. Belayneh, F. Carminati, A. Farbin, B. Hooberman, G. Khattak, M. Liu, J. Liu et al., *Calorimetry with Deep Learning: Particle Simulation and Reconstruction for Collider Physics*, European Physical Journal C, 80 (December 2019), doi:10.1140/epjc/s10052-020-8251-9, URL http://cds.cern.ch/record/2706000

[382] T. A. Lehrer, *We Will All Go Together When We Go*, An Evening Wasted with Tom Lehrer (March 1959), URL https://www.youtube.com/watch?v=frAEmhqdLFs

[383] P. T. Komiske, E. M. Metodiev and J. Thaler, *Energy flow polynomials: a complete linear basis for jet substructure*, Journal of High Energy Physics, 2018 (4) (April 2018), doi:10.1007/jhep04(2018)013, URL https://doi.org/10.1007/jhep04(2018)013

[384] T. Faucett, J. Thaler and D. Whiteson, *Mapping machine-learned physics into a human-readable space*, Physical Review D, 103 (3) (February 2021), doi:10.1103/physrevd.103.036020, URL https://doi.org/10.1103/physrevd.103.036020

[385] A. Biekötter, A. Knochel, M. Krämer, D. Liu and F. Riva, *Vices and virtues of Higgs effective field theories at large energy*, Physical Review D, 91, 055029 (2015), doi:10.1103/PhysRevD.91.055029, URL https://doi.org/10.1103/PhysRevD.91.055029

[386] The ATLAS Collaboration, *Combination of searches for invisible Higgs boson decays with the ATLAS experiment*, Technical Report ATLAS-CONF-2018-054, CERN, Geneva (November 2018), URL https://cds.cern.ch/record/2649407

[387] Y. LeCun, Facebook (February 2019), URL https://www.facebook.com/yann.lecun/posts/10155803693337143

[388] K. Albertsson, P. Altoe, D. Anderson, J. Anderson, M. Andrews, J. P. A. Espinosa, A. Aurisano et al., *Machine Learning in High Energy Physics Community White Paper*, Journal of Physics: Conference Series, 1085 (2018), doi:10.1088/1742-6596/1085/2/022008, URL https://doi.org/10.1088/1742-6596/1085/2/022008, Proceedings, 18$^{\text{th}}$ International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2017): Seattle, WA, USA, August 21-25, 2017

[389] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, *Jet substructure classification in high-energy physics with deep neural networks*, Physical Review D, 93 (9) (May 2016), doi:10.1103/physrevd.93.094034, URL https://doi.org/10.1103/physrevd.93.094034

[390] Y. Lu, A. Romero, M. J. Fenton, D. Whiteson and P. Baldi, *Resolving Extreme Jet Substructure* (2022), doi:10.48550/arXiv.2202.00723, URL https://doi.org/10.48550/arXiv.2202.00723

[391] The ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, Technical report, CERN, Geneva (March 2017), URL https://cds.cern.ch/record/2255226

[392] G. Louppe, K. Cho, C. Becot and K. Cranmer, *QCD-Aware Recursive Neural Networks for Jet Physics*, ArXiv e-prints (February 2017), doi:10.48550/arXiv.1702.00748, URL https://doi.org/10.48550/arXiv.1702.00748

[393] M. Faucci Giannelli, S. Palazzo, A. Suresh, D. Fiacco, E. Kioseoglou, S. Giovinazzo and D. Morozova, *Fast simulation of the ATLAS calorimeter system with Generative Adversarial Networks*, (ATL-COM-SOFT-2020-040) (March 2020), URL https://cds.cern.ch/record/2712552

[394] C. Englert, P. Galler, P. Harris and M. Spannowsky, *Machine learning uncertainties with adversarial neural networks*, The European Physical Journal C, 79 (1) (January 2019), doi:10.1140/epjc/s10052-018-6511-8, URL https://doi.org/10.1140/epjc/s10052-018-6511-8

[395] Y. LeCun, *Deep Learning Hardware: Past, Present, and Future*, International Solid-State Circuits Conference (2019), URL http://yann.lecun.com/exdb/publis/pdf/lecun-isscc-19.pdf

[396] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan and S. Ganguli, *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*, in F. Bach and D. Blei (editors), *Proceedings of the 32nd International Conference on Machine Learning*, *Proceedings of Machine Learning Research*, volume 37, Proceedings of Machine Learning Research (July 2015), URL https://proceedings.mlr.press/v37/sohl-dickstein15.html

[397] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam and P. Vandergheynst, *Geometric Deep Learning: Going beyond Euclidean data*, IEEE Signal Processing Magazine, 34, 18–42 (July 2017), doi:10.1109/MSP.2017.2693418, URL https://doi.org/10.1109/MSP.2017.2693418

[398] A. G. Baydin, L. Shao, W. Bhimji, L. Heinrich, L. Meadows, J. Liu, A. Munk et al., *Etalumis: Bringing Probabilistic Programming to Scientific Simulators at Scale*, Computing Research Repository, abs/1907.03382 (July 2019), URL http://arxiv.org/abs/1907.03382

[399] S. Dallmeier-Tiessen and T. Šimko, *Open science: A vision for collaborative, reproducible and reusable research*, CERN Courier (March 2019), URL https://cerncourier.com/open-science-a-vision-for-collaborative-reproducible-and-reusable-research

[400] X. Chen, S. Dallmeier-Tiessen, R. Dasler, S. Feger, P. Fokianos, J. B. Gonzalez, H. Hirvonsalo et al., *Open is not enough*, Nature Physics, 15 (2), 113–119 (November 2018), doi:10.1038/s41567-018-0342-2, URL https://doi.org/10.1038/s41567-018-0342-2

[401] J. Shiers, *Data preservation is a journey*, CERN Courier (May 2016), URL https://cerncourier.com/data-preservation-is-a-journey

[402] T. Smith and F. Flückiger, *Licensing the Web*, URL https://home.cern/science/computing/birth-web/licensing-web

[403] M. Giampietro, *Hardware joins the open movement*, CERN Courier (June 2011), URL https://cerncourier.com/hardware-joins-the-open-movement

[404] E. Ormancey, *Migrating to open-source technologies* (June 2019), URL https://home.cern/news/news/computing/migrating-open-source-technologies

[405] E. Ormancey and T. Smith, *The Road to new CERN Office and Collaborative Applications*, Microsoft Alternatives project (MALT) (2019), URL https://cern.ch/malt

[406] T. Smith, M. Barroso Lopez, A. R. Purcell and E. M. Porcari, *MALT and the status of recent licensing negotiations with Microsoft – proposed next steps*, (CERN-IT-Note-2022-001) (2021), URL https://cds.cern.ch/record/2800622

[407] The ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*, Technical Report CERN-LHCC-2017-020, ATLAS-TDR-029, CERN, Geneva (September 2017), URL http://cds.cern.ch/record/2285584

[408] The ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Strip Detector*, Technical Report CERN-LHCC-2017-005, ATLAS-TDR-025, CERN, Geneva (April 2017), URL http://cds.cern.ch/record/2257755

[409] The ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer*, Technical Report CERN-LHCC-2017-017, ATLAS-TDR-026, CERN, Geneva (September 2017), URL http://cds.cern.ch/record/2285580

[410] The ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS LAr Calorimeter*, Technical Report CERN-LHCC-2017-018, ATLAS-TDR-027, CERN, Geneva (September 2017), URL http://cds.cern.ch/record/2285582

[411] The ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter*, Technical Report CERN-LHCC-2017-019, ATLAS-TDR-028, CERN, Geneva (September 2017), URL http://cds.cern.ch/record/2285583

[412] The ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, Technical Report CERN-LHCC-2017-021, ATLAS-TDR-030, CERN, Geneva (September 2017), URL http://cds.cern.ch/record/2285585

[413] CERN, *High-Luminosity Large Hadron Collider (HL-LHC) Technical Design Report V. 0.1*, CERN Yellow Reports: Monographs, 4 (2017), doi:10.23731/cyrm-2017-004, URL https://e-publishing.cern.ch/index.php/CYRM/issue/view/40

[414] O. Brüning and L. Rossi, *The High Luminosity Large Hadron Collider: The New Machine for Illuminating the Mysteries of Universe*, volume 24, World Scientific (February 2015), doi:10.1142/9581, URL https://doi.org/10.1142/9581

[415] P. Azzi, S. Farry, P. Nason, A. Tricoli, D. Zeppenfeld, R. Abdul Khalek, J. Alimena et al., *Report from Working Group 1: Standard Model Physics at the HL-LHC and HE-LHC*, CERN Yellow Re-

ports: Monographs, 7 (CERN-LPCC-2018-03) (December 2018), doi:10.23731/CYRM-2019-007.1, URL http://cds.cern.ch/record/2650160

[416] W. D. Breaden Madden, *CERN Commences Study on Future Circular Colliders*, The Glasgow Insight into Science and Technology (February 2014), URL https://the-gist.org/2014/02/cern-commences-study-on-future-circular-colliders

[417] M. Benedikt, M. Capeans Garrido, F. Cerutti, B. Goddard, J. Gutleber, J. M. Jimenez, M. Mangano et al., *Future Circular Collider*, Technical Report CERN-ACC-2018-0058, CERN, Geneva (December 2018), URL https://cds.cern.ch/record/2651300

[418] E. Ter Laak, *Schematic of an 80-100 km long circular tunnel. Schéma d'un tunnel circulaire de 80 à 100 km de long*, (OPEN-PHO-CHART-2014-003) (February 2014), URL https://cds.cern.ch/record/1646785

[419] The ILD Concept Group, *The International Large Detector: Letter of Intent* (March 2009), doi:10.48550/arXiv.1006.3396, URL https://doi.org/10.48550/arXiv.1006.3396

[420] A. Blondel et al., *Letter of Intent: An International Muon Ionization Cooling Experiment (MICE)*, (MICE-NOTE-1) (November 2001), URL https://inspirehep.net/literature/857075

[421] N. Bartosik, A. Bertolin, L. Buonincontri, M. Casarsa, F. Collamati, A. Ferrari, A. Ferrari et al., *Detector and Physics Performance at a Muon Collider*, Journal of Instrumentation, 15 (05) (May 2020), doi:10.1088/1748-0221/15/05/p05001, URL https://doi.org/10.1088/1748-0221/15/05/p05001

[422] D. Neuffer and V. Shiltsev, *On the feasibility of a pulsed 14 TeV c.m.e. muon collider in the LHC tunnel*, Journal of Instrumentation, 13 (10) (October 2018), doi:10.1088/1748-0221/13/10/t10003, URL https://doi.org/10.1088/1748-0221/13/10/t10003

[423] R. Munroe, *Telescope Names*, xkcd, (1294), URL https://xkcd.com/1294

[424] T. R. Junk and L. Lyons, *Reproducibility and Replication of Experimental Particle Physics Results*, Harvard Data Science Review, 2 (4) (December 2020), doi:10.1162/99608f92.250f995b, URL https://doi.org/10.1162/99608f92.250f995b

[425] P. M. Greenfield, *Cultural Change Over Time: Why Replicability Should Not Be the Gold Standard in Psychological Science*, 12 (5), 762–771 (September 2017), doi:10.1177/1745691617707314, URL https://doi.org/10.1177/1745691617707314

[426] T. Berners-Lee, *Information Management: A Proposal* (March 1989), URL https://www.w3.org/History/1989/proposal.html

[427] J. Ingram, C. J. Hand and G. Maciejewski, *Social isolation during COVID-19 lockdown impairs cognitive function*, Applied Cognitive Psychology, 35 (4), 935–947 (March 2021), doi:10.1002/acp.3821, URL https://doi.org/10.1002/acp.3821

[428] G. Pietrabissa and S. G. Simpson, *Psychological Consequences of Social Isolation During COVID-19 Outbreak*, Frontiers in Psychology, 11 (September 2020), doi:10.3389/fpsyg.2020.02201, URL https://doi.org/10.3389/fpsyg.2020.02201

[429] F. Dordei, E. Ben Haim, F. Cicala, S. Klaver, R. Kopecna, V. Lukashenko and N. Skidmore, *Results of the survey on the effects of the Covid-19 pandemic on LHCb scientists*, Technical report, CERN, Geneva (May 2021), URL https://cds.cern.ch/record/2752585

[430] *Universal Declaration of Human Rights*, United Nations General Assembly (December 1948), URL https://www.un.org/sites/un2.un.org/files/udhr.pdf

[431] R. Bregman, *Utopia for Realists*, De Correspondent (November 2016), ISBN 978-9082520347

[432] D. Graeber, *Debt: The First 5,000 Years*, Melville House (2011), ISBN 978-1-933633-86-2, URL https://theanarchistlibrary.org/library/david-graeber-debt-the-first-five-thousand-years

[433] L. Mayor, *Where people and particles collide*, Physics World (March 2016), URL https://physicsworld.com/a/where-people-and-particles-collide

[434] E. Gibney, *Excluded, intimidated and harassed: LGBT physicists face discrimination*, Nature (March 2016), doi:10.1038/nature.2016.19614, URL https://www.nature.com/news/excluded-intimidated-and-harassed-lgbt-physicists-face-discrimination-1.19614

[435] N. Ackerman, T. Atherton, A. R. Avalani, C. A. Berven, T. Laskar, A. Neunzert, D. S. Parno et al., *LGBT+ Inclusivity in Physics and Astronomy: A Best Practices Guide* (April 2018), doi:10.48550/arXiv.1804.08406, URL https://doi.org/10.48550/arXiv.1804.08406

[436] C. O'Luanaigh, *The birds and the beams: Biodiversity at CERN* (September 2015), URL https://home.cern/news/news/cern/birds-and-beams-biodiversity-cern

[437] CERN, *Environment Report 2017–2018*, CERN Environmental-Report, 1 (102-101-PB) (September 2020), URL https://e-publishing.cern.ch/index.php/CERN_Environment_Report/issue/download/102/83

[438] E. Strubell, A. Ganesh and A. McCallum, *Energy and Policy Considerations for Deep Learning in NLP* (2019), doi:10.18653/v1/p19-1355, URL https://doi.org/10.18653/v1/p19-1355

[439] R. Schwartz, J. Dodge, N. A. Smith and O. Etzioni, *Green AI*, Computing Research Repository, abs/1907.10597 (2019), doi:10.48550/arXiv.1907.10597, URL https://doi.org/10.48550/arXiv.1907.10597

[440] *CMS and ATLAS get the green light*, CERN Bulletin, (10), 2–3 (1996), URL http://cds.cern.ch/record/1974283

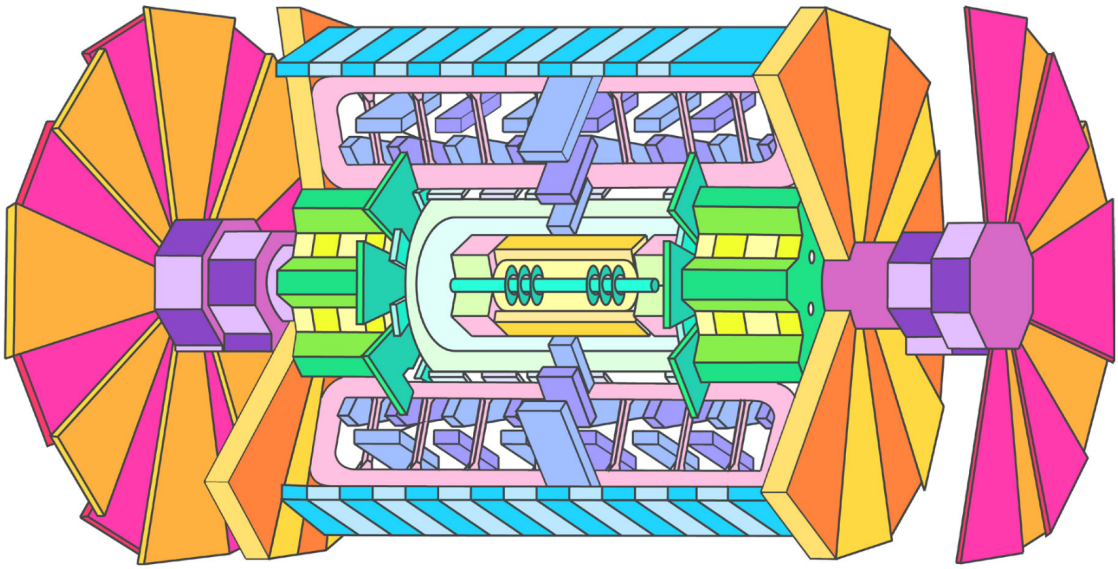


Figure 10.4: ATLAS, as depicted in *The ATLAS Experiment Colouring Book* [30].