Fiducial cross-section measurements of the production of a prompt photon in association with a top-quark pair at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC

Dissertation

for the award of the degree
“Doctor of Philosophy”PhD Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen

within the ProPhys doctoral program
of the Georg-August University School of Science (GAUSS)

submitted by

Joshua Wyatt Smith

from Cape Town, South Africa

Göttingen, 2018

Thesis Committee:

Prof. Dr. Arnulf Quadt
Prof. Dr. Stan Lai

Members of the Examination Board:

Reviewer:	Prof. Dr. Arnulf Quadt II. Physikalisches Institut, Georg-August-Universität Göttingen
Second Reviewer:	Prof. Dr. Stan Lai II. Physikalisches Institut, Georg-August-Universität Göttingen

Further Members of the Examination Board:

Prof. Laura Covi, PhD
Institut für Theoretische Physik, Georg-August-Universität Göttingen

Prof. Dr. Ariane Frey
II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Wolfram Kollatschny
Institut für Astrophysik, Georg-August-Universität Göttingen

Prof. Dr. Steffen Schumann
Institut für Theoretische Physik, Georg-August-Universität Göttingen

Date of the Oral Examination: 6th November 2018

Reference: II.Physik-UniGö-Diss-2018/01

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fiducial cross-section measurements of the production of a prompt photon in association with a top-quark pair at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC

Joshua Wyatt Smith

The cross sections for top-quark pair production in association with a photon are measured in a fiducial volume with the ATLAS detector at a centre-of-mass energy of 13 TeV. Results are presented using proton-proton collision data collected by the LHC during 2015 and 2016, amounting to a total of 36.1 fb^{-1}. This also presents the first $t\bar{t}\gamma$ cross-section measurements performed in the dilepton channels. Exactly one photon is required to have $p_{\text{T}} > 20$ GeV and be isolated based on track and calorimeter information. At least two (four) jets are required in the dilepton (single-lepton) channels, with at least one jet originating from a b-quark. Two separate neural network algorithms are used to help reduce the impact backgrounds play in the final measurements. The Prompt Photon Tagger is trained on information from energy deposits in the calorimeters to distinguish prompt photons from hadronic fake photons. The output of this neural network is fed into the Event-level Discriminator that uses event information to classify signal from the sum of all backgrounds. A maximum likelihood fit is performed on the output of the Event-level Discriminator to determine the fiducial cross section of the signal process. The fiducial cross section for the single-lepton and dilepton channel are measured to be $521 \pm 9(\text{stat.}) \pm 41(\text{sys.})$ fb and $69 \pm 3(\text{stat.}) \pm 4(\text{sys.})$ fb, respectively. In total, eight cross-section measurements are performed and all agree with theoretical next-to-leading-order predictions.

Acknowledgements

I would first like to thank Arnulf Quadt, my PhD supervisor, for the opportunity to be a part of the II. Physikalisches Institut in Göttingen. I learnt a tremendous amount during my time in Göttingen, CERN, and the various conferences around the world that I was lucky enough to participate in. I picked up a very broad range of skills that will certainly help me in my next endeavours, whatever those may be.

Thank you to Elizaveta Shabalina who has been incredibly helpful in the day-to-day supervision of this analysis. Her experience and willingness to try innovative techniques ensured that this measurement was interesting and unique.

A special thank you to Dan Guest for many fruitful discussions (often over pasta) on machine learning topics, and just general help and advice on many subjects during the course of my PhD. Also, thank you for not letting us fall off or get crushed by mountains. That is important.

Thanks to María Moreno Llácer who helped determine the strategy the analysis took, as well as provide assistance on the fitting procedures. Thank you to Yichen Li and Julien Caudron. As analysis contacts they moved the analysis forward and provided essential tips, comments and studies. Yichen was one of the main driving forces in making it possible to show these results at a public conference. Likewise, to the group members of the $t\bar{t}\gamma$ analysis, thank you for making this happen!

Thank you to Thomas Peiffer and Royer Edson Ticse Torres for interesting discussions and for proofreading this thesis. Many helpful comments were provided.

To my friends and colleagues in Göttingen, around the world, and those constantly scattered throughout the French and Swiss Alps (you should be working!), thank you for the general help and companionship that is needed when pursuing a PhD.

Thank you to Marie-Pier for the encouragement over the last year. You understand the demands of finishing a PhD thesis and have been nothing but supportive.

Finally, thank you to my family for the continued love and support throughout the years. It goes without saying that none of this would have been possible without you. For that I am grateful.

Contents

Prologue **1**

1. Introduction **3**
- 1.1. Brief summary of the Standard Model . . . 3
  - 1.1.1. The top quark . . . 6
  - 1.1.2. Top-quark pair production in association with a photon . . . 9
- 1.2. Machine learning in high energy physics . . . 12
  - 1.2.1. Core concepts . . . 13
  - 1.2.2. Tools and libraries . . . 14

2. The LHC and the ATLAS experiment **17**
- 2.1. The Large Hadron Collider . . . 17
- 2.2. The ATLAS detector . . . 19
  - 2.2.1. Magnets . . . 20
  - 2.2.2. The Inner Detector . . . 22
  - 2.2.3. Calorimeters . . . 24
  - 2.2.4. Muon spectrometers . . . 26
  - 2.2.5. Triggers . . . 27

3. Analysed data and Monte Carlo **29**
- 3.1. Data . . . 29
- 3.2. Monte Carlo datasets . . . 29
  - 3.2.1. Nominal signal and background MC samples . . . 32
  - 3.2.2. Signal sample NLO normalisation . . . 32

4. Object reconstruction and identification **33**
- 4.1. Electron and photon objects . . . 33
  - 4.1.1. Photons . . . 34
  - 4.1.2. Electrons . . . 37
- 4.2. Muons . . . 37
- 4.3. Jets . . . 38
  - 4.3.1. b-jets . . . 39
- 4.4. Missing transverse energy . . . 39
- 4.5. Object overlap removal procedure . . . 40
- 4.6. Event overlap removal and categorisation . . . 40

5. Top-quark pairs and photons at $\sqrt{s} = 13$ TeV 43
5.1. Analysis strategy . . . 43
5.1.1. Fiducial cross-section measurements . . . 43
5.1.2. Maximum likelihood fit . . . 45
5.1.3. Fit scenarios . . . 45
5.2. Signal region event selections . . . 46
5.3. $e \to \gamma$ fake background . . . 50
5.4. Hadronic fake background . . . 51
5.5. Non-prompt and fake lepton background . . . 53
5.6. Prompt photon backgrounds . . . 57
5.7. Pre-fit distributions . . . 61

6. Object- and event-level neural networks 65
6.1. The Prompt Photon Tagger . . . 65
6.1.1. Training . . . 65
6.1.2. Evaluation . . . 67
6.1.3. Scale factors and systematic uncertainties . . . 70
6.2. The Event-level Discriminator . . . 72
6.2.1. Variable selection . . . 72
6.2.2. Training . . . 76
6.2.3. Evaluation . . . 83

7. Systematic uncertainties 89
7.1. Signal and background modelling uncertainties . . . 89
7.1.1. QCD renormalisation and factorisation scales . . . 89
7.1.2. Parton Shower . . . 90
7.1.3. Initial and final state radiation . . . 90
7.1.4. PDF uncertainty . . . 90
7.1.5. Other background modelling uncertainties . . . 90
7.2. Experimental uncertainties . . . 91
7.3. Smoothing, symmetrisation and pruning . . . 92

8. Results 101
8.1. Asimov fits . . . 101
8.1.1. ELD binning optimisation . . . 101
8.1.2. Asimov fit cross checks . . . 102
8.2. Full fit results . . . 108
8.2.1. Signal strength . . . 108
8.2.2. Distributions, yields and uncertainties . . . 109
8.2.3. Fit cross checks . . . 115
8.3. Fiducial cross sections . . . 125

9. Conclusions 127

Bibliography 129

Appendices 141

A. Porting the ATLAS software stack to the ARM architecture 143
A.1. Hardware . . . 143
A.1.1. AthSimulation . . . 144

A.1.2. Power measurements 145
A.1.3. Outlook for ARM in HEP 145

B. Support material for the $t\bar{t}\gamma$ process 149
B.1. Validation plots for $t\bar{t}$ selections 149

C. Object- and event-level neural network support material 151
C.1. Event-level Descriminator 151

D. Systematic uncertainties support material 153
D.1. Cross-section closure test for small MC backgrounds 153

E. Results section support material 155
E.1. Final fits 155

Prologue

The study of high energy particle physics revolves around understanding our universe at the most fundamental level. At this level, the constituents of matter and the forces they feel (the electromagnetic, weak, and strong forces) are described by the Standard Model of particle physics. The Standard Model has been confirmed to incredibly high precision over countless experiments for more than 50 years. And yet, we know it is not the full picture. Perhaps the most exciting and profound part is that it can only explain around 4% of *everything* we know in the entire universe. The Large Hadron Collider at CERN was built to explore and probe the Standard Model. By colliding (mainly) protons at high energies, quarks, gluons, leptons, and bosons are created in abundance. The study of these particles and their interactions provide physicists with the data to continue to expand our knowledge of the universe.

The top quark is especially interesting for several reasons. Its very large mass and its strong coupling to the recently discovered Higgs boson seem to hint that this particle is special. The top quark is also the only quark to decay before it has a chance to hadronise. It presents us with a unique opportunity to study the direct fundamental properties of a quark through the decay products. By probing the top-photon coupling at increasing centre-of-mass energies we are continuously putting the Standard Model to the test. While making precision measurements useful for other analyses, we are essentially looking for a scale at which our model does not describe the data anymore. It must exist, we are just not sure where and how this new physics will manifest.

The analysis presented in this thesis measures the cross section for photons radiated from top-quark pairs, the so-called $t\bar{t}\gamma$ process. The data was collected in 2015 and 2016, and amounts to 36.1 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV. Similar measurements of the $t\gamma$ coupling have been performed by the CDF, ATLAS, and CMS experiments, all with less data and lower centre-of-mass energies. For the first time the $t\bar{t}\gamma$ cross-section measurements are performed in the dilepton channels, as well as the single-lepton channels. Modern machine learning algorithms play an important role in the strategies chosen for this analysis.

This thesis is laid out as follows: the introduction (Chapter 1) provides brief theoretical background material in the form of the Standard Model and machine learning in high energy physics. Chapters 2 - 4 lay the groundwork for the experimental setup as well as the experiment-wide algorithms essential for object reconstruction and identification. Chapter 5 describes the analysis strategy, signal, and background processes for the $\sqrt{s} = 13$ TeV analysis. Two unique neural networks using object- and event-level information are presented in Chapter 6. An overview of the systematic uncertainties used in the analysis is provided in Chapter 7, and the final results with in-depth discussions are presented in Chapter 8. This is followed by conclusions in Chapter 9.

In general the appendix is home to supporting material for the above studies. However, Appendix A presents an overview of work done during this PhD in porting the ATLAS software stack (Intel x86 based) to the ARM architecture. Appendices B - E contain support material for the analysis strategy, the object- and event-level neural networks, the systematic uncertainties, and the final results.

The analysis presented in this thesis [1] was first presented at the 11^{th} International Workshop on Top Quark Physics (TOP 2018).

CHAPTER 1

Introduction

1.1. Brief summary of the Standard Model

The Standard Model (SM) describes the behaviour of three out of the four fundamental forces we can observe. These are electromagnetism, the weak force, and the strong force. Many physicists have contributed to the SM over the last century. While the theory itself was formalised in the 1970's, the building blocks such as Quantum Electrodynamics (QED) [2–6] were created well before that.

The SM is the combination of three renormalisable, local gauge invariant quantum field theories, two of which (electromagnetism and the weak force) are unified into the electroweak force [7–10]. The vector fields of the local symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$ give rise to these three fundamental interactions. The first term ($SU(3)$) describes the strong interaction in the framework of Quantum Chromodynamics (QCD) [11–13]. The second two terms ($SU(2)_L \times U(1)_Y$) describe the electroweak interaction.

Particles can be described as excitations of scalar and vector fields. There are 12 spin-$\frac{1}{2}$ fermions and five force mediators or gauge bosons. Interactions between fermions are mediated by the exchange of gauge bosons. Photons (γ) mediate the electromagnetic force, $W^{\pm}$- and Z^0-bosons mediate the weak force, and gluons (g) carry the strong force.

The masses of the $W^{\pm}$- and Z^0-bosons, as well as the prediction of a new heavy scalar boson, arise through spontaneous symmetry breaking of the $SU(2)_L \times U(1)_Y$ gauge symmetry. This is achieved through the Brout-Englert-Higgs mechanism [14, 15], first predicted in 1964. By interacting with the heavy scaler field, fermions also acquire mass. In 2012 the Higgs-boson particle was discovered by the ATLAS and CMS collaborations at the LHC [16, 17].

The interaction between fermions and the photon can be described by the electromagnetic Lagrangian in the framework of QED:

$$\begin{aligned}\mathcal{L}_{\text{QED}} &= \mathcal{L}_{\text{fermion}} + \mathcal{L}_{\text{photon}} + \mathcal{L}_{\text{interaction}} \\ &= \bar{\psi}(i\partial\!\!\!/ + m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - q\bar{\psi}\gamma^{\mu}A_{\mu}\psi. \end{aligned} \tag{1.1}$$

The first term describes the kinematics of the spin-$\frac{1}{2}$ fermion field ($\psi/\bar{\psi}$). The second term describes the kinematics of the photon field A_μ, where $F_{\mu\nu}$ is the electromagnetic field tensor defined as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.

The last term describes the interaction between the photon and fermion field, and is directly proportional to the charge (q) of the fermion. Thus, probing this vertex probes the structure of the coupling as well as the coupling strength of photons to fermions. It is the particles which we can measure in collider experiments that enable us to infer properties about the associated interacting fields.

Of the fermions, six are called quarks and six are called leptons. They are ordered in three generations (essentially by mass), where each quark has an "up-type" (up (u), charm (c), top (t)) and "down-type" (down (d), strange (s), bottom (b)) particle. The leptons consist of charged particles (e, μ, τ) and their neutral counterparts, the neutrinos (ν_e, ν_μ, ν_τ). Each of the up-type quarks carries a fractional electric charge of $\frac{2}{3}e$,[1] while down-type quarks carry $-\frac{1}{3}e$. Charged leptons carry integer charge ($-1e$), while the neutrinos are neutral and only interact via the weak force. All quarks also carry colour charge labelled (somewhat arbitrarily) as red, green or blue. Due to colour confinement [18], quarks and gluons can not be observed as free particles but rather hadronise to form colour neutral composite particles called hadrons. Two examples of hadrons are baryons, which have three quarks (uud in the case of protons), and mesons ($u\bar{d}$ in the case of a charged pion)[2]. All of the observable matter in the universe is comprised of three particles from the first generation: u-quarks, d-quarks and electrons.

For each fermion there exists an antiparticle, essentially identical, except that additive quantum numbers (such as charge and weak isospin) have opposite values. For quarks and neutrinos these particles are denoted with bars (i.e. $\bar{u}$, $\bar{t}$, $\bar{\nu}_e$), while for charged leptons the charge is simply reversed (e^+, μ^+, τ^+).

Leptons, both charged and neutral, can interact via the weak interaction and so are organised into left-handed doublets and right-handed singlets, defined as:

$$\begin{pmatrix} \nu_l \\ l^- \end{pmatrix}_L, \qquad l^-_R.$$

Of the left-handed doublets, the charged leptons have a corresponding third-component of weak isospin[3], $I_3 = -\frac{1}{2}$, while the neutral leptons have $I_3 = \frac{1}{2}$. Right-handed singlets have $I_3 = 0$. The SM does not accommodate right-handed neutrinos, and all neutrinos in the SM are assumed to be massless. However, small oscillations between neutrino flavours have been observed, meaning their masses are actually non-zero [21].

Quarks are sensitive to the weak and strong forces and therefore also transform as doublets and singlets. However, there exists cross-generational coupling between the quarks. The representations can be written as:

$$\begin{pmatrix} q_{\text{u-type}} \\ q'_{\text{d-type}} \end{pmatrix}_L, \qquad q_{\text{u-type},R}, \qquad q_{\text{d-type},R},$$

where for the left-handed doublets the up-type quarks have $I_3 = \frac{1}{2}$ and the down-type quarks have $I_3 = -\frac{1}{2}$. For the right-handed singlets, $I_3 = 0$. The weak eigenstates (q') are translated to the mass eigenstates (q) of the down-type quarks via the CKM matrix [22, 23].

The SM of particle physics as described above is summarised in Figure 1.1. For each particle the charge, mass and spin is depicted and grouped according to generation number (and implicitly by I_3). Also included are the interaction terms represented by lines connecting various particles. For example, gluons have colour charge, as do quarks. Thus, gluons can couple to themselves as well as quarks. Photons can interact with any charged particle; the electron, muon, tau, all six quarks and the W-boson. The W- and Z-vector bosons[4] couple to themselves and the Higgs-boson (H). Since they lack colour charge, they do not couple to gluons.

---

[1] Throughout this thesis natural units are used. This involves setting $c = \hbar = 1$. Masses, energies and momenta are expressed in GeV. Additionally, charges are expressed as absolute values of the electron, e.

[2] Tetraquarks (two quarks and two antiquark-quarks) and pentaquarks (four quarks and one antiquark) have also been observed at the Large Hadron Collider [19, 20].

[3] This quantum number must be conserved in all weak interactions.

[4] In this thesis V-boson refers to the W- and Z-boson, and not the photon.

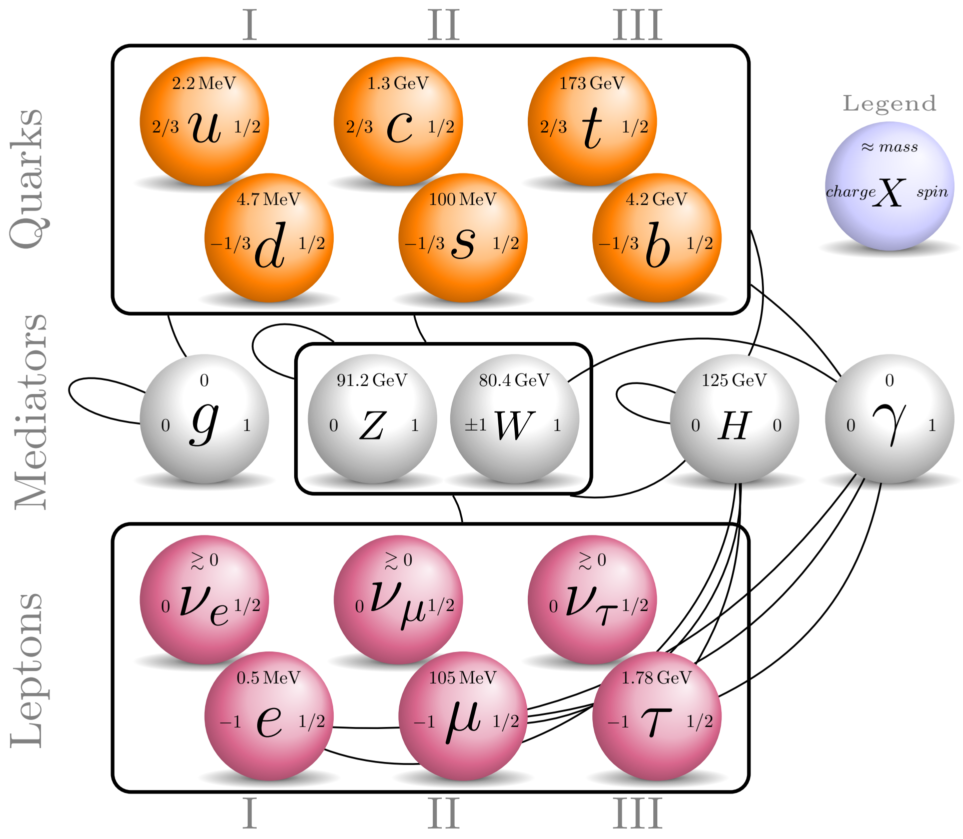

I
II
III
Quarks
Mediators
Leptons
Legend
≈ mass
charge X spin
2.2 MeV 2/3 u 1/2
1.3 GeV 2/3 c 1/2
173 GeV 2/3 t 1/2
4.7 MeV −1/3 d 1/2
100 MeV −1/3 s 1/2
4.2 GeV −1/3 b 1/2
0 0 g 1
91.2 GeV 0 Z 1
80.4 GeV ±1 W 1
125 GeV 0 H 0
0 0 γ 1
≳ 0 0 νe 1/2
≳ 0 0 νμ 1/2
≳ 0 0 ντ 1/2
0.5 MeV −1 e 1/2
105 MeV −1 μ 1/2
1.78 GeV −1 τ 1/2
I
II
III

Figure 1.1.: The Standard Model of particle physics ordered by generation and particle type; quarks, bosons (force mediators) and leptons. Also included are the possible interactions for each boson represented by connecting lines.

The predictions of the SM have been measured to high degrees of accuracy [24] and have yet to be conclusively proven wrong. However, there are gaps missing in the formalism of the SM. A few such examples are:

- Gravity is not included in the SM.
- The masses of particles are not predicted by the SM but rather must be experimentally determined (along with other parameters).
- Neutrinos do not have mass in the SM, yet experimental observation shows their ability to oscillate between different flavours. This depends on the mass squared differences, and thus requires them to have some mass [21].
- The SM describes only about 4% of the entire universe. The rest, 74% dark energy and 22% dark matter, are believed to exist through indirect experimental evidence [25–27]. We know that dark energy and dark matter exist, however, we know very little about them.

By interpreting our measurements in the framework of the SM, these gaps could manifest themselves as new particles or unexplained interactions.

1.1.1. The top quark

This section gives a brief overview of the properties of the top quark. A more thorough treatment can be seen in [28].

After the discovery of the b-quark at Fermilab in 1977 [29], a weak isospin partner was needed to fulfil the prediction of the SM, just as the previous two particle generations have weak isospin partners. The top quark was discovered in 1995 at the Tevatron at Fermilab in proton-antiproton collisions [30, 31].

Production

Top quarks are predominantly created in pairs (top-antitop) through the processes of quark-antiquark annihilation ($q\bar{q} \to t\bar{t}$) or gluon-gluon fusion ($gg \to t\bar{t}$) in proton-proton collisions. At $\sqrt{s} = 13$ TeV, approximately 90% of top-quark pairs are created through gluon-gluon fusion.

In 2009, 14 years after the discovery of the top quark, electroweak production of single-top quarks was observed in the s- and t-channels[5] at the Tevatron [33, 34]. Single-top quarks are also created at the LHC and have a cross section around two to three times lower than that of pair production. The dominant process (about 70%) is from the t-channel production, $ub \to dt$ or $\bar{d}b \to \bar{u}t$ via a virtual W-boson. Other processes include the s-channel, $u\bar{d} \to t\bar{b}$ (also via a virtual W-boson) as well as production of a top quark with a real W-boson, $gb \to Wt$.

In hadron colliders, top-quark pair production can be factorised into two "pieces" and convoluted with one another, known as the factorisation theorem [35]. As a function of the centre-of-mass energy (CME, $\sqrt{s}$) and the top-quark mass, the cross section of $t\bar{t}$ production from proton-proton collisions is described as

$$\sigma^{pp \to t\bar{t}}(\sqrt{s}, m_t) = \sum_{i,j=q,\bar{q},g} \int dx_i dx_j f_i(x_i, \mu_f^2) f_i(x_j, \mu_f^2) \cdot \hat{\sigma}^{ij \to t\bar{t}}(m_t, \sqrt{\hat{s}}, x_i, x_j, \alpha_s(\mu_R^2), \mu_R^2). \tag{1.2}$$

The term on the first line is governed by non-perturbative QCD and describes the modelling of the protons, where $f_i(x_i, \mu_f^2)$ is the parton distribution function (PDF) for parton i. The PDFs describe the fraction of momenta (x_i) the partons possess from the parent hadron. The PDFs are experimentally determined [36–38] and are evaluated as a function of the factorisation scale μ_f, which separates perturbative from non-perturbative QCD. The bottom term describes the interaction using perturbative QCD where the cross section is a function of the top-quark mass, the CME in the parton-parton rest frame ($\sqrt{\hat{s}}$), the strong coupling constant α_s, and the renormalisation scale μ_R. This term includes the matrix element calculation which can occur at various levels of precision.

Decay and backgrounds

The SM predicts that top quarks decay essentially 100% of the time to a W-boson and b-quark. Decays to the s- and d-quarks are strongly suppressed by the CKM matrix. The W^+-boson decays to a pair of up-type (u- or c-) and down-type ($\bar{d}$-, $\bar{s}$-, $\bar{b}$-) quarks 67% of the time, or a charged lepton (l^+) and the corresponding neutrino (ν_l) around 33% of the time (summed total for the three charged leptons). The decay of the W-boson categorises the channels in which we choose to make top-quark measurements. For instance $t \to W^+b \to l^+\nu b$ is categorised as a leptonic decay, whereas $t \to W^+b \to q\bar{q}b$ is classified as a "hadronic" decay.

The branching ratios for $t\bar{t}$ decay are shown in Figure 1.2. When both W-bosons decay to a quark-antiquark pair and a b-quark, this is classified as "all hadronic" and occurs 46% of the time. When one W-boson decays leptonically ($l = e, \mu, \tau$) this is classified as the single-lepton channel and occurs 45% of the time. If both W-bosons decay leptonically then this is called the dilepton channel and occurs 9% of the time. In the context of this thesis, an "inclusive" measurement includes all but the "all hadronic" channels. Figure 1.3 shows the diagrams for gluon-gluon initiated $t\bar{t}$ production for the single-lepton and dilepton decay channels.

---

[5] Here, s-, t-, u-channels refer to the Mandelstam variables [32].

The main irreducible[6] backgrounds in $t\bar{t}$ measurements arise from the production of a leptonically decaying W- or Z-boson with jets in the single-lepton and dilepton channels, respectively. The electroweak single-top processes can also significantly contribute to the $t\bar{t}$ background with extra jets from background QCD processes. However, the flavour of the jets is a powerful discriminant to help distinguish these backgrounds from the signal. As the CME increases the $t\bar{t}$ production cross section increases faster than that of the respective backgrounds, essentially turning the LHC into a top-quark factory. Thus, the signal to background ratio becomes larger.

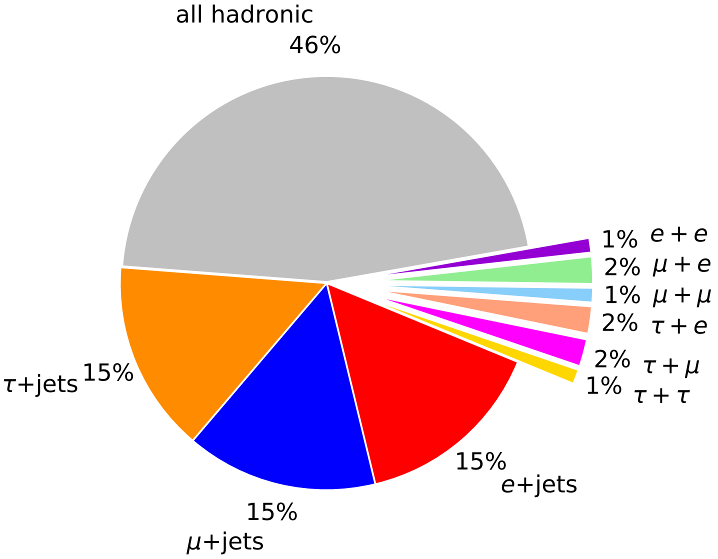


Figure 1.2.: Branching ratios for $t\bar{t}$ decay [24].

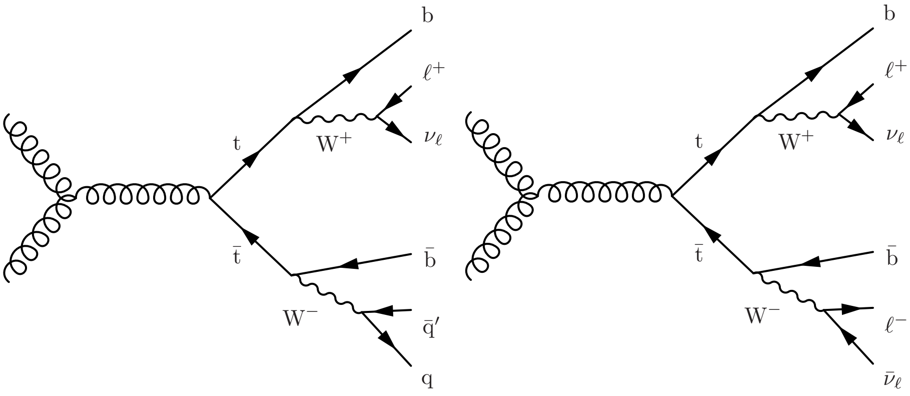
Figure 1.4 shows all $t\bar{t}$ cross section measurements performed at the Tevatron and LHC as a function of the CME. Theoretical calculations to which the measurements are compared are performed at next-to-next-to-leading order (NNLO) with next-to-next-to-leading logarithm (NNLL) order calculations for soft gluon resummation [39]. All measurements agree with the theoretical prediction.

Properties

Due to the very large mass of the top quark it is expected that it plays an important role in electroweak symmetry breaking. The lifetime is approximately 0.5×10^{-24}s, which is about an order of magnitude smaller than the timescale for the hadronisation process of quarks. Thus, it decays before it has a chance to hadronise. This means that the properties are transferred directly to the decay products. This makes for interesting measurements in which the underlying dynamics of quarks can be studied. A select few properties and results are highlighted, while a full overview can be found in [24].

---

[6] Irreducible backgrounds have the same final state as the signal



(a) single-lepton (b) dilepton

Figure 1.3.: Examples of the $t\bar{t}$ process for gluon-gluon production and single-lepton and dilepton decay channels.

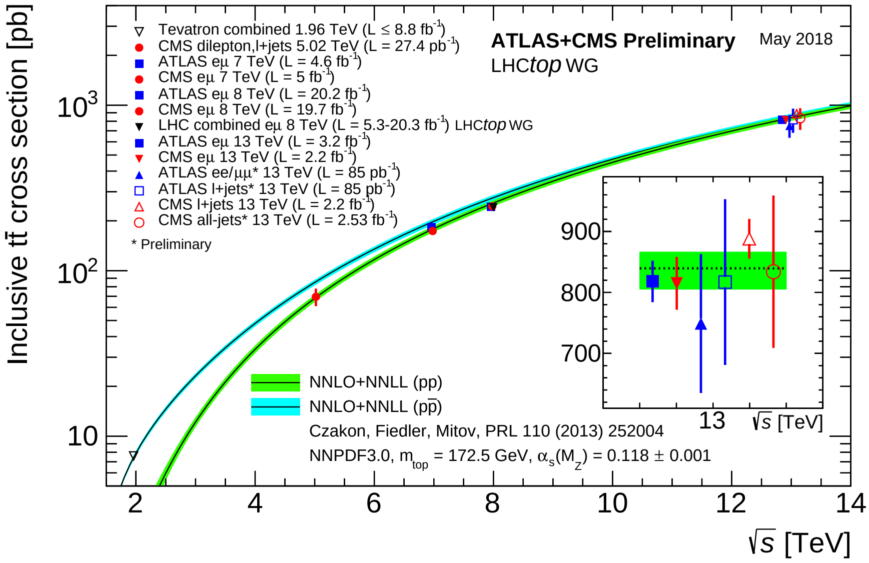


Figure 1.4.: A summary of all $t\bar{t}$ cross section measurements performed at the Tevatron and LHC as a function of the CME, compared to the NNLO QCD calculation with NNLL resummation. Measurements made at the same CME are slightly offset for clarity [40].

Direct mass measurements of the top quark have been performed at the Tevatron [41] and at ATLAS [42] and CMS [43, 44] at the LHC. A global average results in 173.0 ± 0.4 GeV [24]. Indirect measurements of the top-quark mass from cross sections have also been performed at DØ [45] at the Tevatron, as well as at ATLAS [46–48] and CMS [49, 50]. The global average results in 173.1 ± 0.9 GeV [24], consistent with the direct mass measurements.

After the discovery of what was believed to be the top quark, one hypothesis was that it could be an exotic quark with charge $-\frac{4}{3}e$ [51]. Measurements made at the Tevatron, and more recently at the LHC, strongly reject this hypothesis [52, 53].

The top-quark width is expected to be fairly large given that the lifetime is very small. Deviations from the expected decay width could hint at new physics. A global average results in $\Gamma_t = 1.41^{+0.19}_{-0.15}$ GeV [24], which agrees with theoretical predictions.

An interesting measurement of top-quark pairs is that of the spin correlation [54]. The SM predicts that the angles of the decay products from the top quark are correlated with the top polarisation. In $t\bar{t}$ pairs, the spin correlations between each of the top quarks can be inferred from angles between the respective decay products. At $\sqrt{s} = 7$ TeV the measurements made by ATLAS were consistent with the SM [55, 56]. However, in a recent measurement at $\sqrt{s} = 13$ TeV with the ATLAS detector, a 3.2σ deviation from the SM is observed [57]. It should be mentioned that the same mis-modelling is observed in the analysis presented in this thesis at a lower significance of 1.5σ. However, it will not be presented and differential distributions can be seen in [1].

In $q\bar{q} \rightarrow t\bar{t}$ events at leading order (LO), the top quark (antitop quark) moves in the direction of the incoming quark (antiquark). However, at next-to-leading order (NLO) interference terms from additional Feynman diagrams cause an asymmetry between this symmetrical angular production. The asymmetry can be exaggerated by Beyond SM (BSM) physics. Tension with the SM was first seen at measurements made at the Tevatron [58]. However recent measurements done at $\sqrt{s} = 8$ TeV in ATLAS and CMS [59–65] all agree with theoretical SM predictions. Interestingly, the charge asymmetry between top-quark pairs produced at the LHC is expected to be magnified in $t\bar{t}\gamma$ production [66], however this measurement has yet to be performed.

Lastly, a measurement of how the top quark couples to photons probes the electromagnetic coupling. This is the central theme in this thesis and is explained in the next section.

1.1.2. Top-quark pair production in association with a photon

A cross-section measurement of a prompt photon in association with a top-quark pair ($t\bar{t}\gamma$) can be interpreted as a direct probe of the electromagnetic coupling of the top quark as explained above. Anomalous top-quark couplings could manifest as shape discrepancies in various kinematic distributions or in cross-section measurements [67,68]. The results can also be interpreted in the framework of an effective field theory in the search for new physics [69, 70].

In addition to being radiated from a top quark, photons can be radiated from any of the charged particles resulting from top quark decay, as well as the incoming partons in proton-proton collisions. The processes that contribute to the background of $t\bar{t}$ are essentially the same for $t\bar{t}\gamma$, albeit with smaller cross sections. However, overall it is fake photons and leptons that contribute the most to the $t\bar{t}\gamma$ backgrounds. These are discussed further in Chapter 5.

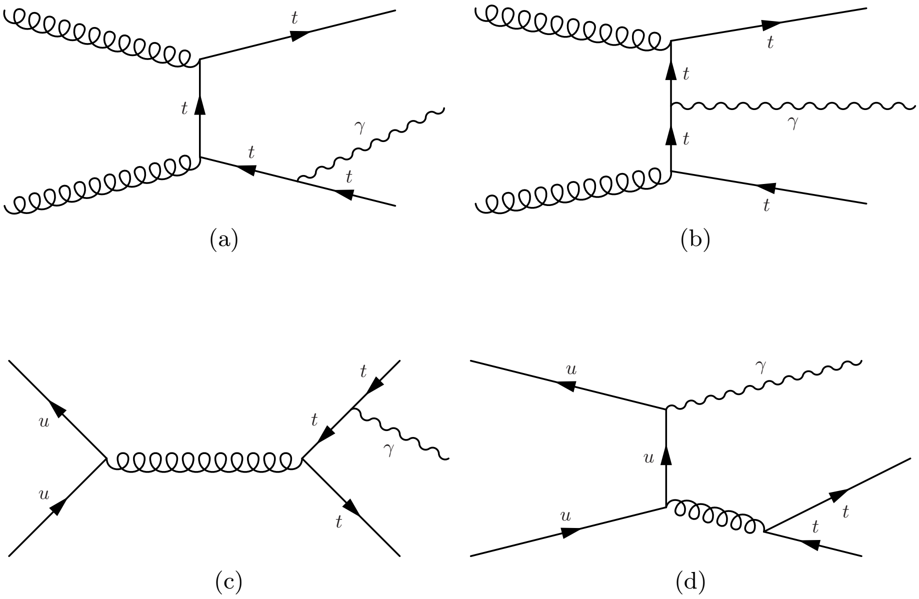
We can artificially construct two "processes" in which to classify photons considered as signal.

- *Radiative top production*: This is the sought after signal process, specifically when photons are radiated from off-shell top quarks (Figure 1.5a, 1.5b, 1.5c). An interference term can also enter in the form of initial state radiation (ISR) from the incoming quarks (Figure 1.5d). As mentioned previously, as the CME increases, so too does the cross section of $t\bar{t}$ production from gluon-gluon fusion. Thus, the contribution of photons radiated from the incoming quarks is only $\mathcal{O}(10\%)$.

- *Radiative top decay*: Photons are radiated from any of the charged particles from the decay of the top quark. This includes photons from b-quarks (Figure 1.6a), W-bosons (Figure 1.6b), and leptons (Figure 1.6c). Also included is the radiation from the on-shell top quark.

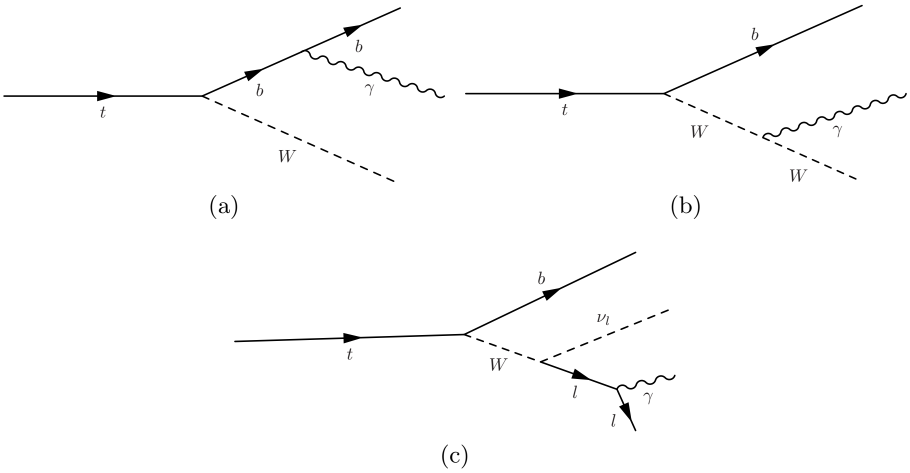
In all diagrams, the coupling of a quark to the photon is described by the last term in Equation 1.1. Distinguishing these processes is difficult (and in the case of ISR from top-quark radiation, an ill posed question) as all result in prompt photons. However, for radiative top-decay contributions we can exploit the kinematic properties and constituents of each event.

Placing cuts between leptons and photons in the η-ϕ plane helps reduce this contribution. This will be explained in Chapter 5.2.



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

Figure 1.5.: Examples of the $t\bar{t}\gamma$ process at LO from quark-antiquark annihilation and gluon-gluon fusion. Photons are radiated during top quark production or from initial partons. All diagrams are considered as signal.



(a) (b) (c)

Figure 1.6.: Example diagrams of top quark decays where photons are radiated from charged particles. All diagrams are considered as signal.

Status of $t\bar{t}\gamma$ analyses

Evidence of the $t\bar{t}\gamma$ process occurred at the CDF experiment at the Tevatron [71] at a CME of 1.96 TeV. A total dataset of 6.0 fb^{-1} from $p\bar{p}$ collisions was used to measure the cross section resulting in $\sigma_{t\bar{t}\gamma} = 180 \pm 80$ fb. A ratio of the $t\bar{t}$ to $t\bar{t}\gamma$ cross section was also measured resulting in $\sigma_{t\bar{t}}/\sigma_{t\bar{t}\gamma} = 0.024 \pm 0.009$. Measuring the ratio allows for smaller uncertainties due to various experimental effects either cancelling out or being reduced. Subsequent measurements have been carried out at the LHC in both the ATLAS and CMS experiments.

Observation at 5.3σ was seen at the LHC by ATLAS at $\sqrt{s} = 7$ TeV with a total dataset size of 4.59 fb^{-1} [72]. The measurement was carried out in the single-lepton channel resulting in a fiducial cross section, $\sigma_{t\bar{t}\gamma} \times \text{BR} = 64 \pm 8(\text{stat})\ ^{+17}_{-13}(\text{syst}) \pm 1(\text{lumi})$ fb per lepton flavour.

CMS performed an analysis at $\sqrt{s} = 8$ TeV with an integrated luminosity[7] of 19.7 fb^{-1}. The fiducial measurement was done in the single-lepton channel where they measured the ratio to the $t\bar{t}$ cross section, $\sigma_{t\bar{t}}/\sigma_{t\bar{t}\gamma} = (5.2 \pm 1.1) \times 10^{-4}$ [73]. The $t\bar{t}\gamma$ cross section is extrapolated to all single-lepton final states resulting in $\sigma_{t\bar{t}\gamma} \times \text{BR} = 515 \pm 108(\text{stat+syst})$ fb per single-lepton flavour.

ATLAS carried out another measurement at $\sqrt{s} = 8$ TeV in the single-lepton channel with an integrated luminosity of 20.2 fb^{-1} [74]. The cross section yielded $\sigma_{t\bar{t}\gamma} = 139 \pm 7(\text{stat.}) \pm 17(\text{syst.})$ fb. For the first time this analysis included differential measurements of the transverse momentum and pseudorapidity of the photon.

In summary, all the above measurements have been carried out in some form of the single-lepton channel, and all agree with SM predictions. The cross-section results scaled to the respective theoretical NLO predictions (represented by the dashed vertical line) are summarised in Figure 1.7.

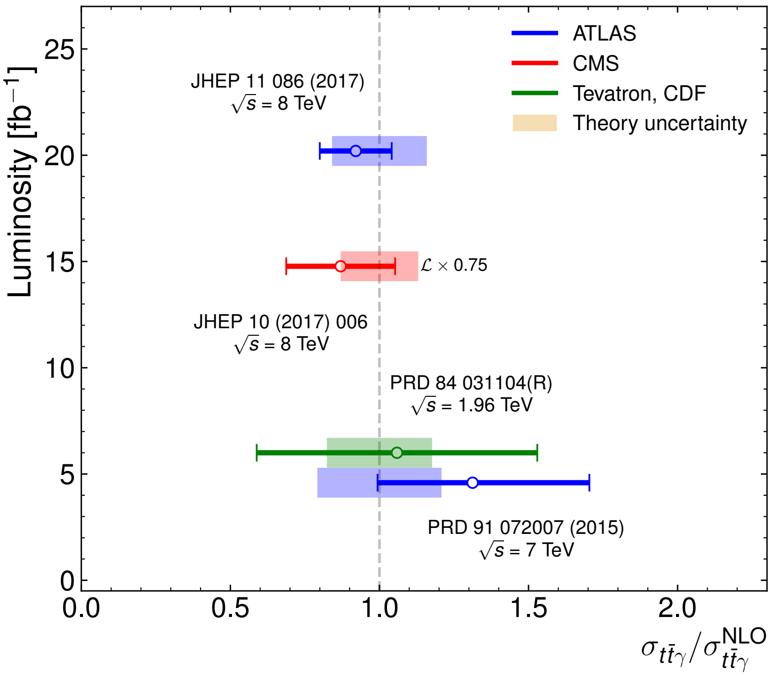


Figure 1.7.: Summary of the previous $t\bar{t}\gamma$ cross-section measurements performed at the Tevatron and LHC, scaled and compared to the respective theoretical predictions (represented by the dashed vertical line at one). The luminosity for the CMS measurement is scaled by 0.75 for better readability. The shaded bands represent the theoretical errors obtained from dedicated calculations for each analysis.

---

[7] The integrated luminosity is the total amount of data collected over a certain period. The instantaneous luminosity corresponds to the incoming number of particles per second and is determined by the accelerator properties.

1.2. Machine learning in high energy physics

The goal of all detector components and algorithms used in high energy physics (HEP) is essentially to identify particles more efficiently with smaller uncertainties. The goal of all analyses is to examine and extract meaningful information from the data.

Multivariate analyses (MVAs) in HEP take many inputs (called variables or features) and reduce this complexity to find new features in data. Some examples that CERN has conventionally used in the past are Likelihoods, Kalman fitters [75], Boosted Decision Trees (BDTs, which the Higgs boson discovery of 2012 used) and neural networks (NNs). The latter two belong to a class of algorithms called Machine Learning (ML). When ML algorithms are used two steps are essential: training a model which is time and central processing unit (CPU) intensive, and inference of the model. Inference refers to the process of applying the trained algorithm to unseen data. This is a CPU in-expensive task. ML techniques for training (and inference) have progressed significantly in the last few years with specific emphasis on NNs. A new class, *Deep neural networks* (DNNs), has large implications for the workflows in HEP[8]. With the huge amount of data collected (and still to be collected) at the LHC, deeper NNs are possible, thus further exploring the fundamentals of physics interactions.

Some topics in which ML can help HEP are highlighted in dedicated reports [76, 77] as well as detailed studies and research papers. A selection of some general themes (with examples where available) are explained below.

- (Fast) Simulation [78]: Simulation plays a crucial role as it enables us to compare our collected data to a theory (Chapter 3.2). The events in the simulated data must have gone through the same interactions as the real data do through the detectors. That means each particle and its interaction with matter needs to be simulated accurately. This is an incredibly CPU intensive task, especially in calorimeters where particle showers occur. An area of research is to learn these “shower shape variables” through DNNs, or more specifically Generative Adversarial Networks (GANs) [79]. Inference of already simulated datasets could save significant amounts of CPU time.

- Triggering: The decision to accept or reject events is made by the online and offline components of a triggering system in a detector (Chapter 2.2.5). Key to the success is the speed and accuracy in which decisions are made to keep only a tiny fraction of interesting events. Also key is how to define what constitutes an interesting event. DNNs essentially have the capability of increasing the efficiency of various triggers by using all reconstructed particle information and kinematics. This could mitigate the possibility of throwing away interesting events based on a more fuller description of the constituents of the event.

- Object (and event) reconstruction and identification [80–88]: A large proportion of DNN studies focus around jet physics. This involves studying the quark/gluon jet substructure that leaves different signatures in calorimeters, an ideal problem for Recurrent Neural Networks (RNN) with Long Short Term Memory (LSTM) units or Convolutional Neural Networks (CNN). These algorithms have originally been developed for the field of natural language processing and image recognition. This thesis makes use of novel techniques for object identification (Chapter 6.1) and event classification (Chapter 6.2).

- Uncertainty mitigation [89]: An interesting area of research involving NNs it to mitigate the systematic uncertainties that arise in any HEP analyses. The key lies in “pivoting” between sensitivity or robustness in a classification problem. This introduces a novel technique using GANs.

---

[8] The term “Deep” is very subjective. Essentially, it can mean many trainable parameters from hundreds of thousands to tens of millions or even larger. For this reason DNN and NN are often used interchangeably.

- Anomaly detection [90]: From the nominal operation of each component of a detector, to storing massive amounts of data on hardware, everything needs constant monitoring. In the case of failure an appropriate response needs to be carried out. Conventionally this is done to some degree by electronic sensors, but ultimately always ends up at an "expert" taking action. By detecting or even forecasting anomalies based on enormous amounts of data, algorithms could detect and fix, or even just notify the correct expert significantly faster than humans. Another type of anomaly detection applies to physics processes. By training a NN on simulation (which is a theory such as the SM) we create output distributions. These outputs are non-linear transformations and so can represent non-trivial properties that are otherwise hidden. By inferring on unseen collected data, assumptions can be made on the validity of the theory.

The above points generally rely on some well simulated datasets that provide the labels that are used in training. In cases where labels are not available (for example training on actual data) or when the labels are not implicitly trusted, novel methods have been developed [91–94]. While some of these techniques are still in early development, we can imagine the advantages that DNNs without labels may provide. An example would be not telling the network what to look for, but rather looking for patterns that occur naturally in the data, which can be interpreted as re-discovering fundamental physics with very little or no bias to a given theory.

1.2.1. Core concepts

A brief introduction to some core concepts, common algorithms, and architectures used in this thesis will be presented here.

The typical goal of an NN algorithm is to search for a function which maps a given space or set $\mathbf{X}$ (which are some observed data points with associated features) to lower dimensional space $\mathbf{Y}$ (which are the target labels). In doing so, a loss function or cost function is optimised.

The transformation of an input variable ($\boldsymbol{x}$) into a hidden state ($\mathbf{h}_i$) is expressed as

$$\boldsymbol{h}_{i+1} = g_i(W_i \boldsymbol{h}_i + \boldsymbol{b}_i), \tag{1.3}$$

where i represents the ith transformation. For the first transformation, $\boldsymbol{h}_0 = \boldsymbol{x}$. W_i is called a weight matrix and $\mathbf{b}_i$ is the bias vector. These are the parameters that are updated in the process of optimising a loss function, the so-called training of the NN. For this thesis, weight matrices and bias vectors are initialised uniformly in the range $[0, 1]$. g_i is called the activation function. This is often a non-linear function with the only requirement being that it needs to be differentiable. This thesis makes use of three different activation functions;

- ReLu: $g(z) = max(0, z)$
- Sigmoid: $g(z) = \frac{1}{1+e^{-z}}$
- Softmax: $g(z) = \frac{e^{z_j}}{\sum_{k=1} e^{z_k}}$, for j number of inputs to final layer, with k final outputs.

For all given weights and biases we want an algorithm that describes $\boldsymbol{y}$ for all inputs $\boldsymbol{x}$. In this thesis, ideally, signal and background events are classified such that

$$y = \begin{cases} 0, & \text{if background} \\ 1, & \text{signal.} \end{cases} \tag{1.4}$$

Thus, *binary cross entropy* is used as the loss function and defined as:

$$\text{Loss}(\boldsymbol{y},\boldsymbol{p}) = -\boldsymbol{y}\log(\boldsymbol{p}) + (1-\boldsymbol{y})\log(1-\boldsymbol{p}), \tag{1.5}$$

where $\boldsymbol{y}$ are the given labels and $\boldsymbol{p}$ are the predictions made by the network. The loss becomes small (close to 0) when $\boldsymbol{y}$ is roughly equal to the output, $\boldsymbol{p}$. The training stage of a NN involves minimising the loss function using gradient descent. The significant advancements in ML that we have seen since the 1990's is largely due to the invention of *back-propagation* [95]. By taking partial derivatives, weights and biases in all hidden layers are able to be updated with respect to the loss function in significantly less time. This allows for the training of deeper networks.

Overtraining occurs when the NN has learned too much about the training dataset. Generalisation to unseen data could result in artificial trends emerging based on the learned fluctuations in the training set. Two regularisation methods in the form of layers are used to reduce any overtraining that may occur (in addition to the general k-fold cross validation checks presented later).

- Batch Normalisation [96]: During training, the distribution of inputs at each layer changes as the parameters of the previous layer change. This leads to slower training as the overall learning rate is reduced due to parameter initialisation. It is also difficult to train models with saturating non-linearities (such as for a sigmoid or *tanh* activation layers). This is referred to internal covariate shift. The solution is surprisingly simple and effective: normalise each batch so that it has a mean of 0 and a standard deviation of 1.
- Dropout [97]: A dropout layer sacrifices some performance for the generalisation of the NN. During training, a percentage (supplied as a hyper-parameter in which to optimise) of neurons are evicted from the training epoch[9]. This ensures the NN does not become too dependent on specific neurons, which can lead to overtraining. Essentially what this technique is doing is training many slightly different NNs simultaneously.

1.2.2. Tools and libraries

An underlying theme in this thesis is the use of industry standard, open-source machine learning libraries in ATLAS. While this may not sound profound, it is. It represents a large change to the workflow of how ML is used within HEP at CERN, specifically within ATLAS. By using modern libraries such as TensorFlow [98], Theano [99] and Keras [100], new approaches to ML can be achieved making the ideas presented above possible.

This thesis presents work on an analysis that was one of the first studies to use modern NN libraries on ATLAS data and not on phenomenological datasets that most of the studies presented above use. It presents a challenge because the ATLAS software stack is large (about 6.5 million lines) with the vast majority written in C++. It was built at a time when machine learning in HEP was not foremost in people's minds. Modern ML libraries generally have python (and only much more recently, C++) interfaces, which would need significant restructuring of our code to be able to use. For this reason ROOT [101] (the standard "swiss army knife" of tools in HEP) and the TMVA library [102] have been popular in HEP, simply because the code was designed around ROOT data structures.

ROOT and TMVA were developed at CERN and are maintained by developers based at CERN. However, there are good reasons to use industry standard (yet open-source) libraries. For instance, at time of writing, TensorFlow and Keras have around 1600 and 700 contributors to their respective codebases since their "founding" in 2015.[10] In comparison, ROOT has just over 200 contributors spanning over about 18 years[11]. A larger user-base ensures access to cutting edge algorithms, better documentation and more helpful forums.

---

[9] An epoch refers to one full cycle of training on the training dataset.

[10] Taken from their respective Github code repositories.

[11] Taken from `www.openhub.net`.

A strategy that has propelled the use of ML in HEP is based on the premise that training and inference of a NN are two steps that are completely independent of each other. The training stage can take place in any environment, often making use of graphical processing units (GPUs), which considerably speeds up the training of complex networks. Inference has more restrictions, and in the case of ATLAS needs to have a low CPU footprint and be written in C++. The analysis presented in this thesis makes extensive use of Lightweight Trained Neural Network (LWTNN) [103], which was created to bridge the gap between training using modern ML libraries, and inference of the NN within ATLAS (and CERN in general). It has enabled rich ML studies using novel techniques that are coming to fruition.

CHAPTER 2

The LHC and the ATLAS experiment

The European Organisation for Nuclear Research (CERN) is currently the largest scientific laboratory in the world. CERN is a high energy particle physics laboratory founded in 1954, where many important discoveries have been made. These include (among others) the discovery of neutral currents in 1973 [104], W- and Z- bosons [105, 106], direct charge-parity (CP) violation [107] and the Higgs boson in 2012 [16, 17]. Various accelerators have been commissioned at CERN with the latest being the Large Hadron Collider (LHC). This chapter discusses the LHC and describes the ATLAS detector which sits on one of the four main interaction points (IP) on the LHC ring.

2.1. The Large Hadron Collider

Figure 2.1 shows the schematics for how particles are injected into the LHC. The protons originate from a bottle of hydrogen where the electrons are stripped off and injected into the Linear accelerator 2 (Linac 2). They are accelerated to an energy of 50 MeV where they then enter the Booster and are accelerated up to 1.4 GeV. From there the protons enter the Proton Synchrotron (PS) where they are accelerated up to 25 GeV within a ring with a circumference of 628 m. The protons then enter the Super Proton Synchrotron (SPS), which is the second largest accelerator at CERN. It has a circumference of 7 km and was first switched on in 1976. It was at this accelerator that the W and Z bosons were discovered. Protons are accelerated up to 450 GeV before they are injected into the final accelerator: the LHC.

The LHC [108] is 27 km in circumference and crosses the border of Switzerland and France in multiple places. It is the final stage in the chain of accelerating particles to the design centre of mass energy ($\sqrt{s}$) of 14 TeV. It is made up of eight straight segments and eight arcs which contain dipole magnets to keep the particles in a circular orbit. There are 1232 dipole magnets and 392 quadrupole magnets that steer the beam along its 27 km trajectory. Special focusing magnets squeeze the beam at four IPs to induce collisions. The LHC was designed to accommodate 2808 proton bunches per beam (with 10^{11} protons in a bunch) and with a spacing of 25 ns per bunch. This means a peak instantaneous luminosity of around 1.2×10^{34} cm^{-2}s^{-1}, which translates to just under 1 billion collisions per second.

A detector sits at each of these IPs. There are the two general purpose detectors:

- ATLAS (A large Toroidal LHC ApparatuS) [109] is a general purpose detector designed to probe many aspects of particle physics from "bump hunting" (detection of new particles)

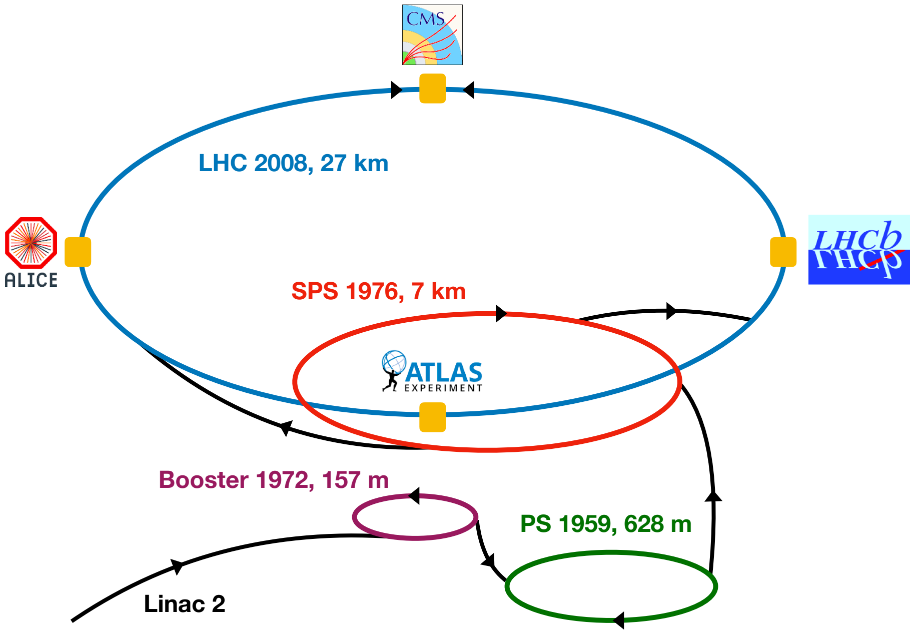

CMS
LHC 2008, 27 km
ALICE
LHCb
SPS 1976, 7 km
ATLAS EXPERIMENT
Booster 1972, 157 m
PS 1959, 628 m
Linac 2

Figure 2.1.: A simplified view of the CERN accelerator complex. Interaction points are shown as yellow squares. Each element in the chain serves to accelerate particles (black arrows) to higher energies, culminating at the LHC. This schematic applies to protons being injected into the LHC. For heavy ions, a slightly different initial injector scheme is used, which is not shown.

to SM precision measurements, to Beyond the SM (BSM) and supersymmetry (SUSY) searches. It is the largest HEP detector in the world. ATLAS will be described in detail in Section 2.2.

- CMS (Compact Muon Solenoid) [110] is a second general purpose detector with the same physics goals as ATLAS. The redundancy is needed to ensure that discoveries are verifiable. While the physics goals are the same, the detector itself and the software and algorithms are different to ATLAS.

There are two other detectors with more specific aims[1]:

- ALICE (A Large Ion Collider Experiment) [111] studies the properties of the universe an infinitesimal time after the Big Bang. Specifically, it studies the quark-gluon plasma that forms under conditions of extremely high temperatures and densities. They require heavy nuclei to be collided such as lead-lead, proton-lead or even Xenon atoms. Special runs of the LHC are requested where protons are switched to heavy ions.

- LHCb (LHC beauty) [112] focuses on b-physics (hadrons that contain b quarks) which play an important role in probing the asymmetries of nature. It could answer questions like the imbalance of matter versus anti-matter in our universe. LHCb takes advantage of the production asymmetry of the signatures of interest and only has detectors in a more forward region of one of the beam directions.

2.2. The ATLAS detector

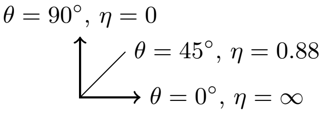
The ATLAS detector, shown in Figure 2.2, is 25 m tall and 44 m long, weighing about 7000 tons. It is symmetric and is made up of many sub-detectors. The IP is located at the centre of the detector, which defines the origin of the coordinate system used henceforth. The beam axis is defined by the z-axis with the x-y plane transverse to the beam axis. The polar angle, θ is measured from the beam axis and is often expressed in terms of the pseudorapidity[2], $\eta = -\ln\tan(\theta/2)$, while the azimuthal angle, ϕ is measured around the beam axis. A cone $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, can be expressed as the angular distance in the $\eta - \phi$ plane. Quantities such as the transverse momentum (p_{T}), transverse energy (E_{T}) and missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) are expressed in the x-y plane.

Figure 2.3 shows a cross section of the ATLAS detector and how it can be split into four main components (each of which have further sub-components). The inner detector (ID) is largely responsible for tracking. With millions of read-out channels and thus high resolution, this detector provides a snapshot of tracks for charged particles. Ionising gases or semiconductor material is used in these detectors. ATLAS has two critical magnet systems. The first surrounds the inner detector with a homogeneous 2 T magnetic field to detect charge type and particle momentum. The second critical magnet is the large toroidal magnet for which ATLAS is named. This toroidal magnet provides an in-homogeneous field of between 0.5 and 4 Tesla. The electromagnetic (EM) calorimeters measure energy from photons and electrons due to

---

[1] There are multiple smaller experiments that branch off from the LHC, but these four detectors are the only ones that sit at beam IPs.

[2] This quantity is preferred over θ since the differences in pseudorapidity are Lorentz invariant under boosts along the beam axis. The relationship can be shown as:


$\theta = 90°$, $\eta = 0$
$\theta = 45°$, $\eta = 0.88$
$\theta = 0°$, $\eta = \infty$

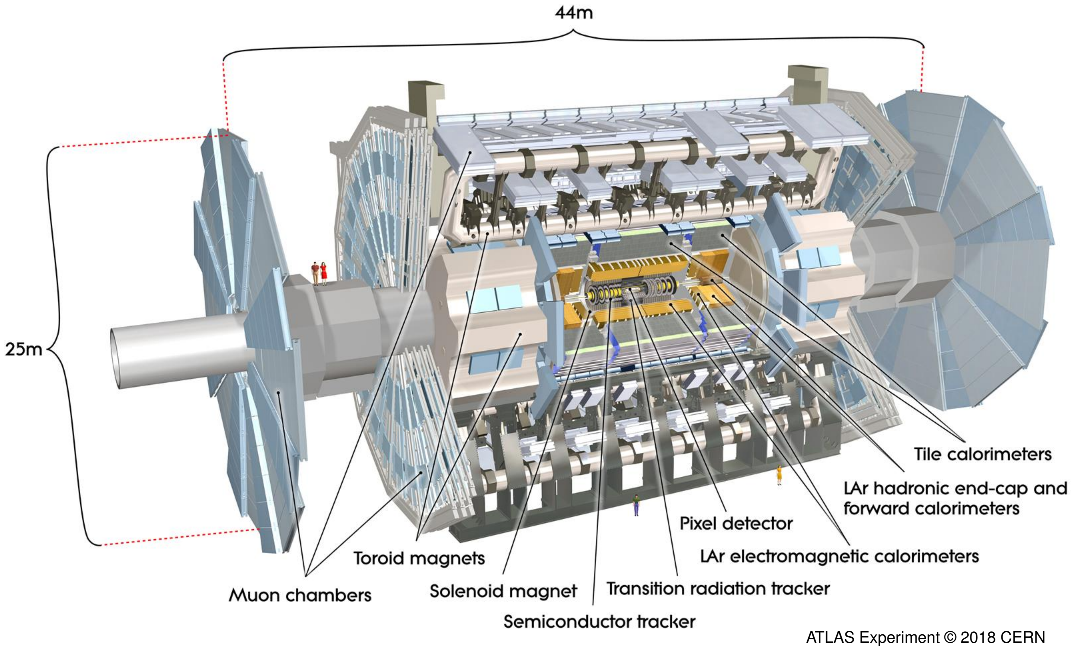

44m
25m
Tile calorimeters
LAr hadronic end-cap and forward calorimeters
Pixel detector
LAr electromagnetic calorimeters
Toroid magnets
Muon chambers
Solenoid magnet
Transition radiation tracker
Semiconductor tracker
ATLAS Experiment © 2018 CERN

Figure 2.2.: An overview of the ATLAS detector (including humans for scale) with some of the sub-detectors labelled.

their interactions with matter. Pair production and bremsstrahlung are such examples. The hadronic calorimeters measure the energy deposits of heavier particles (such as protons and neutrons) due to their interaction with matter. Muons are minimal ionising particles and so are able to traverse the entire detector. Thus, the muon detectors encapsulate the other detectors and form the outer layer. The interaction of neutrinos with matter is negligible. For this reason neutrinos pass through the detector and are reconstructed as “missing transverse energy” ($E_{\mathrm{T}}^{\mathrm{miss}}$), which is possible due to energy conservation. These components are described further in the next sections.

2.2.1. Magnets

Charged particles feel a force due to an electric and magnetic field (the Lorentz Force). Their bending radius is proportional to their momentum. Thus, magnets are essential to determine the momentum and charge type of various particles. ATLAS has two magnet systems which influenced much of the design for the rest of the experiment. These are the solenoid and toroid (with separate end-caps) magnets. Figure 2.4 shows the geometry of the magnet systems. At the centre lies the ID (the first four layers) which is surrounded by the solenoid magnet. The eight toroid coils as well as the end-cap coils are also shown.

The solenoid magnet has an outer diameter of 2.56 m and a length of 5.8 m and has its conductive windings between the inner and outer layer of width 12 mm. Care had to be taken to reduce the amount of material before the calorimeter, yet remain steady and lightweight. The conductive coils are made from Niobium-titanium (NbTi), which is reinforced with an aluminium alloy. The solenoid fully encompasses the ID and provides a homogeneous 2 T field, while only contributing a total of 0.66 radiation lengths, X_0[3]. The toroid magnet system consists of a barrel and end-cap segments. The barrel components consists of eight coils encased in

---

[3] The radiation length is a property of a material defined as the length at which a charged particle loses 0.368 (or $1/e$) of its energy when traversing the material.

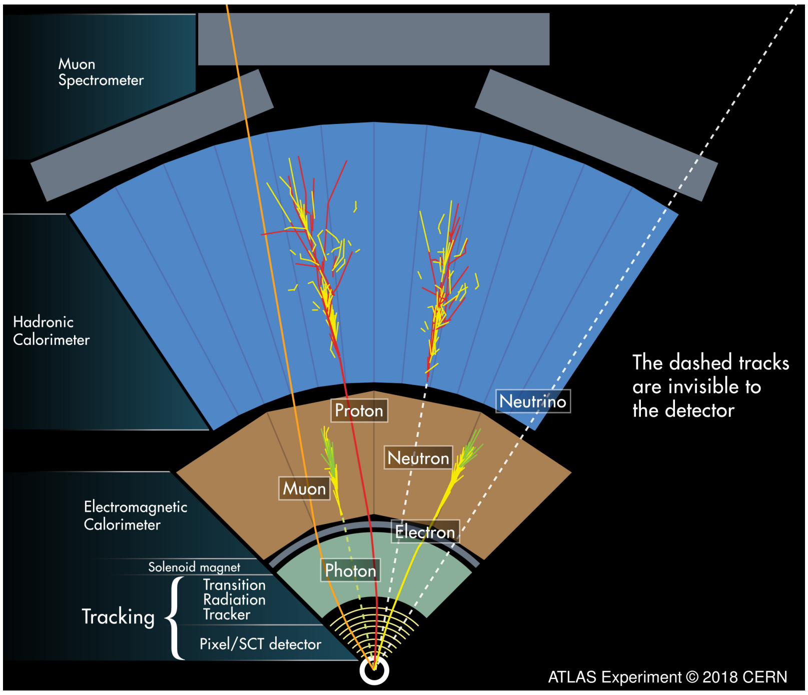

Muon Spectrometer
Hadronic Calorimeter
Electromagnetic Calorimeter
Solenoid magnet
Tracking
Transition Radiation Tracker
Pixel/SCT detector
Proton
Neutrino
Neutron
Muon
Electron
Photon
The dashed tracks are invisible to the detector
ATLAS Experiment © 2018 CERN

Figure 2.3.: A cross section of the ATLAS detector showing the main detector components and the different points at which particles interact with the detector. Deflection or curvature of particles due to the solenoid magnet are also shown. Deflection or curvature due to the toroid magnet are in and out of the page and so can not be seen.

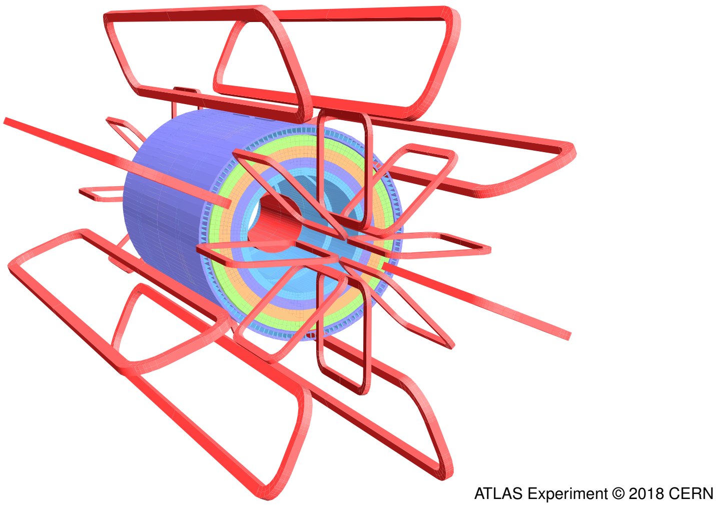

ATLAS Experiment © 2018 CERN

Figure 2.4.: The solenoid magnet (the last layer on the cylinder) surrounds the ID. The eight air-coils for the toroid magnets as well as the end-caps are also shown.

circular stainless-steel vacuum tubes. The overall barrel toroid length is 25.3 m while the outer diameter is 20.1 m. The net Lorentz force acting on each coil is approximately 1400 tonnes, directed inwards. The structural deflection of the barrel toroid had to be taken into account in the construction of magnet infrastructure. The end-cap toroids consists of eight flat, square coil units and eight wedges. These are joined to form a structure that can withstand a Lorentz force of 250 tonnes. Both the barrel and the end-caps toroids consist of pure Al-stabilised Nb/Ti/Cu conductor.

2.2.2. The Inner Detector

Three sub-components make up the Inner Detector (ID), all placed within a homogeneous 2 T magnetic field provided by the solenoid. This allows for accurate determination of charged particles' momenta. By order of distance from the beam-pipe these are the Pixel Detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). Figure 2.5a shows the location of each component. All components have a barrel and end-cap elements. Not shown is the Insertable B-Layer (IBL), which is part of the pixel layer and was added after Run 1. A crucial consideration in designing the ID was to minimise the material used in the construction, and thus ensure particles reach the calorimeters. Depending on the value of $|\eta|$, the radiation length varies from about 0.5 X_0 (for around $|\eta| < 0.7$) to approximately 2.5 X_0, the densest material being due to the barrel and end-cap infrastructure plates at approximately $|\eta| = 1.5$ and $|\eta| = 2.7$. Figure 2.5b shows a more detailed schematic of the ID, which includes these end-plates. Also shown are radii for each component from the beam-pipe, as well as the lengths of the detectors in the z-plane. The resolution for the ID as a function of the transverse momenta is $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}} \approx 0.05\% p_{\mathrm{T}}[\mathrm{GeV}] \oplus 1\%$.

Pixel detectors

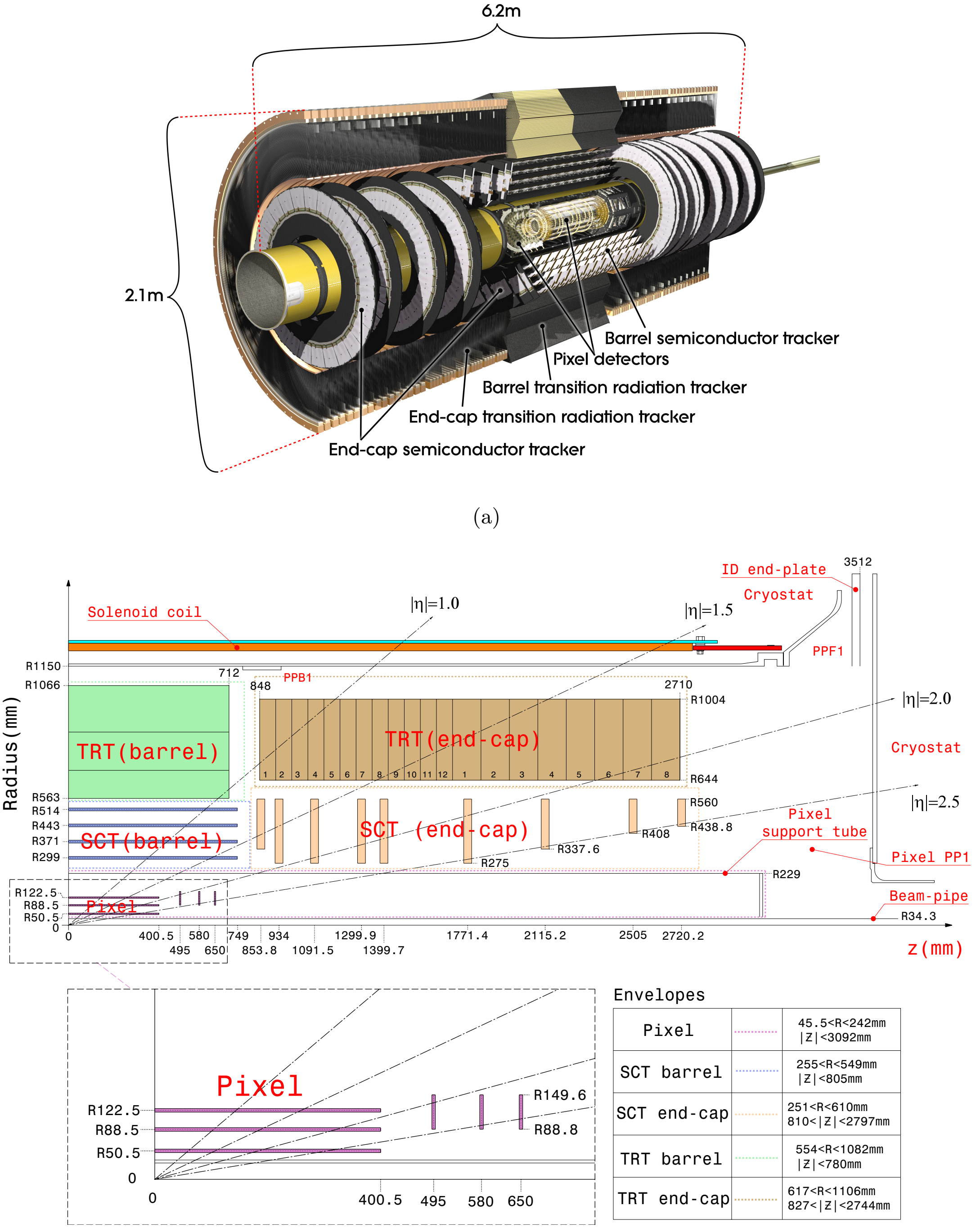
The Pixel detectors (based on silicon semiconductor technology) provide high resolution vertex measurements for charged particles within $|\eta| < 2.5$ and consist of more than 80 million pixels. Originally it consisted of three layers but after Run 1, ATLAS installed an Insertable B-Layer (IBL) [113], which contains smaller pixels and thus enhances the capabilities of detecting b-tagged jets. The IBL is placed 33.2 mm away from the beam-pipe, with the three original layers placed at 50.2 mm, 88.5 mm and 122.5 mm from the beam-pipe, as shown in Figure 2.5b. The older three layers and disks of the Pixel Detector modules have an accuracy of 10 µm in the R-ϕ plane and 115 µm in the z or radial plane for single hits. The IBL is accurate to 8 µm in the R-ϕ plane and 40 µm in the z direction.

Semiconductor Tracker

The SCT contributes to high precision measurements of track impact parameters and has a coverage of $|\eta| < 2.5$. It functions similar to the Pixel detectors in that it is based on silicon semiconductor technology. It is composed of four cylindrical barrel layers and 18 planar end-cap discs. A slight tilt and overlaying sensors allows for 2D measurements of particle hits. The SCT barrel and disk modules have an accuracy of 17 µm in the R-ϕ plane and 580 µm in the z or radial plane.

Transition Radiation Tracker

The TRT is made up of 4 mm diameter tubes of polyimide, cut to 144 cm for the barrel and 37 cm for the end-caps. Around 73 layers in the barrel and 160 straw planes in the end-cap provide high track curvature measurements through transition radiation up to a coverage of


6.2m
2.1m
Barrel semiconductor tracker
Pixel detectors
Barrel transition radiation tracker
End-cap transition radiation tracker
End-cap semiconductor tracker
ID end-plate
3512
Cryostat
Solenoid coil
|η|=1.0
|η|=1.5
PPF1
R1150
R1066
712
848
PPB1
2710
R1004
|η|=2.0
Radius (mm)
TRT (barrel)
TRT (end-cap)
Cryostat
1 2 3 4 5 6 7 8 9 10 11 12 1 2 3 4 5 6 7 8
R644
R563
R514
R443
R371
R299
SCT (barrel)
R560
|η|=2.5
SCT (end-cap)
R408
R438.8
Pixel support tube
R337.6
R275
Pixel PP1
R229
R122.5
R88.5
R50.5
0
Pixel
Beam-pipe
R34.3
0
400.5
495
580
650
749
853.8
934
1091.5
1299.9
1399.7
1771.4
2115.2
2505
2720.2
z (mm)
Pixel
R122.5
R88.5
R50.5
0
R149.6
R88.8
0
400.5
495
580
650
Envelopes
Pixel 45.5<R<242mm |Z|<3092mm
SCT barrel 255<R<549mm |Z|<805mm
SCT end-cap 251<R<610mm 810<|Z|<2797mm
TRT barrel 554<R<1082mm |Z|<780mm
TRT end-cap 617<R<1106mm 827<|Z|<2744mm

(a)

(b)

Figure 2.5.: a) An overview of the ATLAS ID with the different components. b) Quarter-slice of the ID showing the main detectors and dimensions. Also shown is the solenoid magnet. Not shown is the IBL which was added during the first long shutdown of the LHC in 2013. It would be shown in b) at a radius of $R = 33.2$ mm from the beam-pipe [109].

$|\eta| < 2.0$. The TRT is also able to discriminate between pions and electrons due to high thresholds on the signal energy. The barrel straws are placed parallel to the beam axis, while the end-cap disks place the straws perpendicular to the beam-axis. The straws are filled with a $Xe/CO_2/O_2$ gas mixture and the anodes consist of tungsten wires plated with gold. The TRT modules have an electron collection time of approximately 48 ns and a drift-time accuracy of around 130 µm.

2.2.3. Calorimeters

Figure 2.6 shows the layout of the EM and hadronic calorimeters in ATLAS. Both have full ϕ coverage around the beam axis with a total pseudorapidity coverage of $|\eta| < 4.9$. The EM calorimeters consist of a barrel (EMB) and end-cap (EMEC) components.

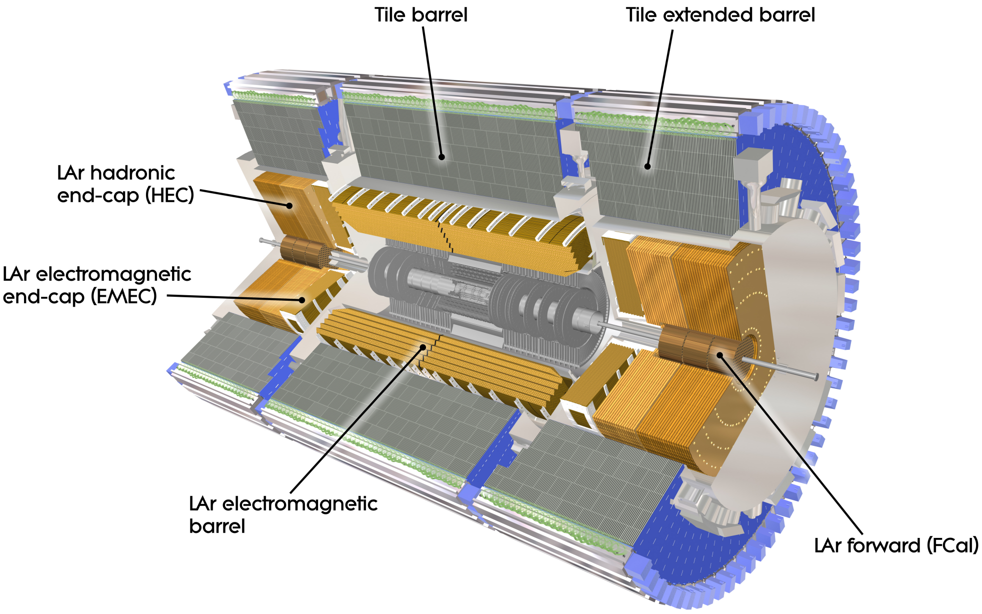

Tile barrel
Tile extended barrel
LAr hadronic end-cap (HEC)
LAr electromagnetic end-cap (EMEC)
LAr electromagnetic barrel
LAr forward (FCal)

Figure 2.6.: The layout of the electromagnetic and hadronic calorimeters in the ATLAS detector [109].

The hadronic calorimeter sits behind the EM calorimeters and includes the central barrel part and extended barrel part (the Tile calorimeter). Liquid Argon (LAr) hadronic end-cap calorimeters (HEC) and LAr forward calorimeters (FCal) also form part of the overall hadronic calorimeter, which is most suited to $E_{\mathrm{T}}^{\mathrm{miss}}$ measurements and jet reconstruction. The EMB has a total coverage of $|\eta| < 1.475$, while the EMECs have a coverage of $1.375 < |\eta| < 3.2$. The EM calorimeter is a lead-liquid argon detector where lead plays the role of the absorber and LAr the active material. The accordion shape is motivated to avoid cracks in coverage in the azimuthal direction. The EMEC is split into two separate wheels that cover the range $1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$. The first and most finely grained layer of the EMB (and the first wheel of the EMECs) is the LAr strip layer which gives high precision measurements of EM energy clusters in the η-direction. The second layer collects the majority of the energy deposition of the EM shower and so has a coarser granularity. The third layer collects the tail of the energy distribution and so is the least granulated layer. Figure 2.7 shows a slice of the EM calorimeter at $\eta = 0$ (i.e. perpendicular to the beam axis and therefore a barrel component where high precision physics is carried out), specifically the three layers and the accordion-like geometry.

The resolution in the η-ϕ plane is also given for each layer, as well as the radiation lengths.

The Tile calorimeter has coverage of $|\eta| < 1.7$ and is made of steel (the absorbing material) and silicon as the scintillating material. The HEC has coverage of $1.5 < |\eta| < 3.2$ and is a copper/LAr sampling calorimeter. The FCal has coverage of $3.1 < |\eta| < 4.9$ and is split into three modules. The first, an EM calorimeter makes use of a copper absorber and a LAr active material. The second two modules cater to hadronic needs and use mainly a tungsten absorber (to cater for high absorption lengths) with LAr active material.

The energy resolution for the calorimeters is parameterised by $\sigma(E)/E = (a/\sqrt{E[\mathrm{GeV}]}) \oplus b$. For the EM calorimeters $a \approx 10\%$ and $b \approx 0.7\%$, while for the hadronic calorimeters $a \approx 50\%$ and $b \approx 3\%$ in the barrel and end-cap, and $a \approx 100\%$ and $b \approx 10\%$ in the forward region.

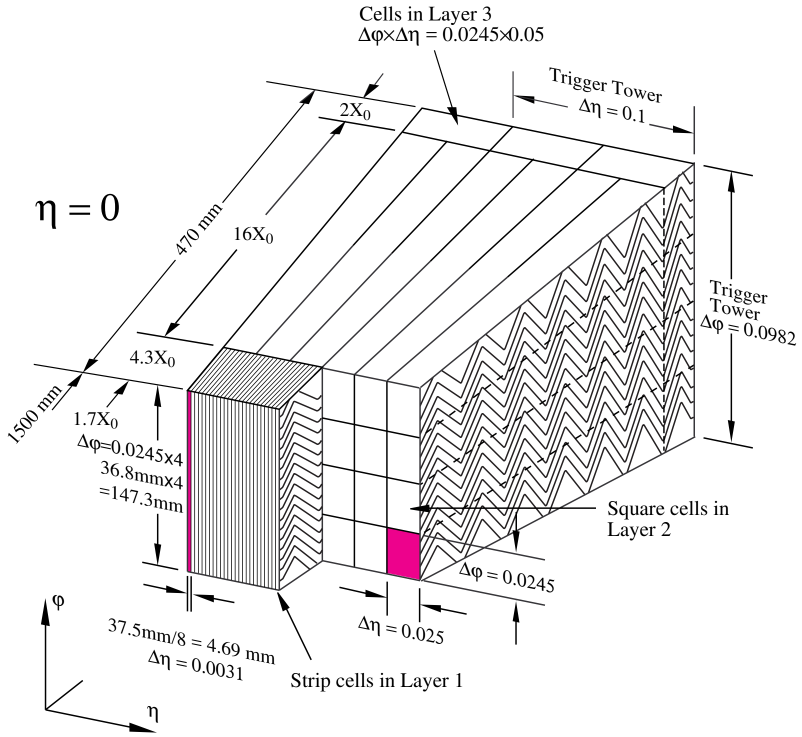
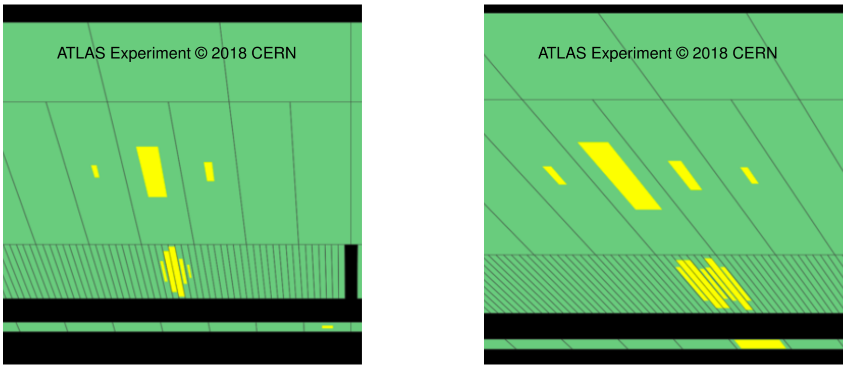


Figure 2.7.: A sketch of a barrel module for the electromagnetic calorimeter. Shown are the three layers with the η-ϕ resolution and radiation lengths [114].

A relevant example detailing the effectiveness of the different layers in the EM calorimeter for this thesis is shown in Figure 2.8. Two event displays depict the presampler[4], then a layer of lead, followed by the finely grained LAr strip layer and the two coarser layers. A prompt photon candidate (left image) and a π^0 candidate (right image) in the three layers of the EM calorimeter are shown. A π^0 decays into two photons and so for many analyses looking at prompt photons, these are considered fake photons from hadron decays and thus background. The prompt photon candidate shows a more collimated energy deposition in the finely grained first layer of the EM calorimeter, whereas the π^0 decay shows a more spread out energy deposit due to the presence of two photons. In the second layer of the calorimeter the two energy deposits look similar. Thus, the LAr strips are essential for prompt versus fake photon identification.

---

[4] The presampler is a thin layer of LAr that sits in front of the EM calorimeter. It is designed to correct for energy loss from particles in the ID and solenoid.



(a) Energy deposit signatures from a photon candidate.

(b) Energy deposit signatures from a π^0 candidate.

Figure 2.8.: Two event displays showing a a) photon candidate and a b) π^0 candidate decaying to two photons in the presampler and three different layers of the EM calorimeter. Lead separates the presampler from the three layers of the EM calorimeter.

2.2.4. Muon spectrometers

The components of the muon spectrometers have a coverage of $|\eta| < 2.7$ (but trigger on $|\eta| < 2.4$) and work together with the in-homogeneous field provided by the air-coil toroidal magnet (for $|\eta| < 1.4$) and the two end-cap magnets (for $1.6 < |\eta| < 2.7$). The region $1.4 < |\eta| < 2.6$ receives magnetic field contributions from both magnets. Thus, muons are curved due to the magnetic field lines that are mostly perpendicular to the particles, and the detectors can make accurate measurements of the particles' momenta. The momentum resolution for muons with p_{T} greater than 1 GeV is approximately $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}} = 10\%$. The layout of the muon detector systems is shown in Figure 2.9. It consists of four major types of sub-detectors. Two serve as tracking: monitored drift tubes (MDT) and cathode strip chambers (CSC); and two serve as muon triggers: resistive plate chambers (RPC) and thin gap chambers (TGC).

MDTs and CSCs

The MDTs are made up of drift tubes filled with an Ar/CO_2 gas mixture with a diameter of 30 mm. A gold-plated tungsten-rhenium wire of diameter 50 µm at a potential of 3080 V collects charges due to the ionisation of gas when a particle traverses the chamber. This allows for an average resolution of 80 µm per tube. They cover up to $|\eta| < 2.7$, except in the innermost end-cap layer where $|\eta| < 2.0$. The CSCs cover the innermost layer and have a coverage of $2.0 < |\eta| < 2.7$. To account for higher than expected backgrounds in this forward region, multi-wire proportional chambers with cathodes segmented into strips are used creating a higher granularity. A potential of 1900 V is applied. The ionising gas used is a mixture of CO_2 and CF_4. The resolution of a CSC chamber is 40 µm in the z-direction and 5 mm in the ϕ-direction. The locations of the MDT wires and CSC strips are aligned with an optical alignment system and are actively monitored. This ensures that the resolution of a muon trajectory is known to less than 30 µm.

RPCs and TGCs

The RPC and TGC form a muon triggering system in the pseudorapidity region of $|\eta| \leq 2.4$. They provide fast information of muons traversing the ATLAS detector. Some examples that they help with are muon transverse momentum measurements, bunch crossing identification and tracking information. RPCs are used in the barrel region. They consist of a parallel electrode-plate filled with a CO_2/C_5H_{10} mixture at a distance of 2 mm from each other. The electric field in the gap is 4.9 kV/mm. Three concentric cylindrical layers around the beam axis cover a pseudorapidity region of $|\eta| \leq 1.05$ in the barrel and provide good time and spatial resolution. The TGCs are used in the end-cap regions, $1.05 < |\eta| < 2.4$ and make use of multi-wire proportional chambers. The wire-to-cathode distance is always smaller than the wire-to-wire distance, and thus they provide a high rate capability and good time resolution. In addition to functioning as a trigger, they also provide the second ϕ coordinate which compliments the MDT measurements. The ionising gas is a mixture of CO_2 and n-C_5H_{12} and has an operation potential of 200 V.

2.2.5. Triggers

With a bunch crossing rate of 40 MHz at the ATLAS IP, a system in which only interesting events are saved is crucial. In Run 1 ATLAS used a three level trigger system (one level was hardware based and two were software) to reduce the number of data saved to the order of a few hundred Hertz. For Run 2, ATLAS revisited its trigger and data acquisition pipeline to be able to cope with the increased data and more challenging run conditions.

An upgraded custom-made electronic hardware trigger is used as the first trigger (L1). It uses information from a subset of detectors such as the triggering calorimeters and muon detectors to highlight candidate objects in regions of interest (RoI). Candidate objects are muons, EM clusters, jets, taus, missing transverse energy, and total energy. This reduces the event rate down to about 100 kHz. These RoIs then seed the second trigger. This is called the High Level Trigger (HLT) and replaces the older level 2 (L2) and event filter layers. A dedicated HLT computing farm makes use of shared algorithms to merge RoIs to allow for more efficient CPU processing, and thus a more efficient trigger. The HLT reduces the event rate from the L1 trigger of 100 kHz to a more manageable 1 kHz, which is then written to disk.

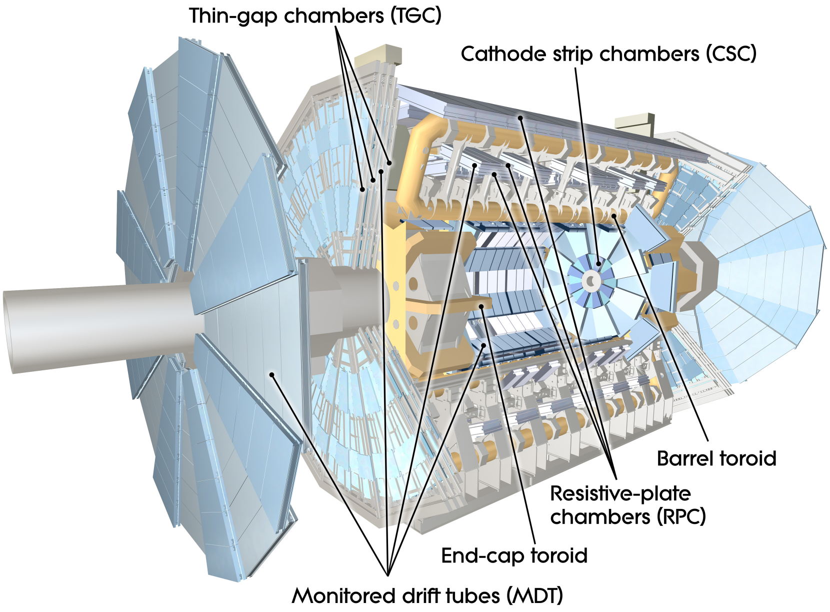

Thin-gap chambers (TGC)
Cathode strip chambers (CSC)
Barrel toroid
Resistive-plate chambers (RPC)
End-cap toroid
Monitored drift tubes (MDT)

Figure 2.9.: The layout of the four main muon detector components for the ATLAS detector. Also shown are the two toroid magnet components [109].

CHAPTER 3

Analysed data and Monte Carlo

3.1. Data

This thesis makes use of data collected with the ATLAS detector in 2015 and 2016 at $\sqrt{s} = 13$ TeV. This makes up only part of the LHC's Run 2 dataset[1]. Strict requirements are placed on the quality of the data, requiring only good luminosity blocks[2]. This corresponds to 3213 pb^{-1} and 32885 pb^{-1} of data collected in 2015 and 2016, respectively. Figure 3.1 shows the total integrated luminosity that the LHC delivered and what ATLAS recorded for the years 2015 and 2016. The difference in delivered and recorded luminosity can be due to a number of reasons ranging from problems in ATLAS data acquisition software and hardware, to the process of the ATLAS detector reaching data taking conditions once stable beams have been announced by the LHC.

Figure 3.2 shows the mean number of interactions per bunch crossing (*pileup*) for the years of 2015 and 2016. This corresponds to the mean of the Poisson distribution for the number of interactions calculated per bunch crossing. Higher pileup makes object identification more difficult and thus increases the systematic contributions that have to be taken into account. As the LHC increases its energy and instantaneous luminosity, the problem of larger pileup profiles becomes more prolific and needs to be addressed.

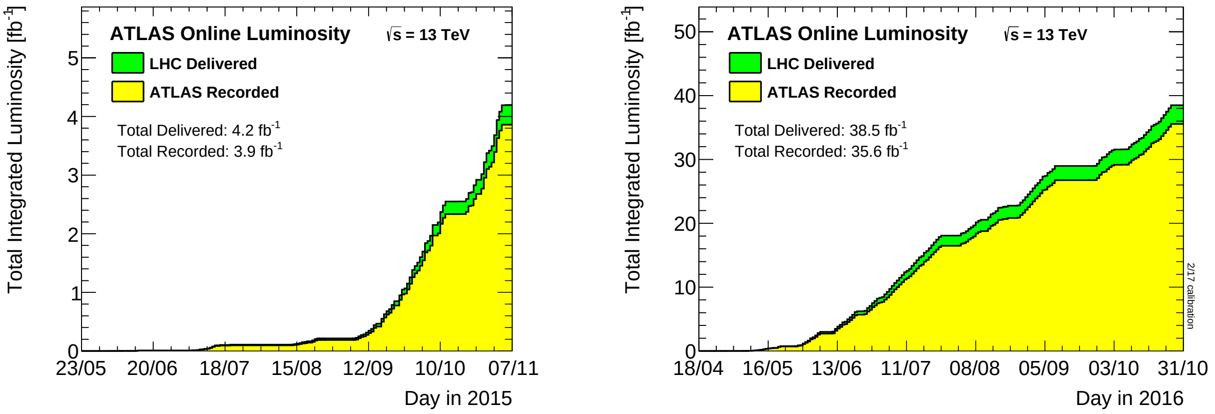
3.2. Monte Carlo datasets

Theoretical models are needed for which collected data can be compared against. The process of going from a theoretical prediction to a Monte Carlo (MC) dataset that is in the same format and has the same detector biases as ATLAS data is a complex process requiring three main steps (and many sub-steps). These are: event generation, detector simulation and object reconstruction.

---

[1] A total of around 120 fb^{-1} of data is expected to be collected by the end of Run 2, i.e. the end of 2018.

[2] A run in which ATLAS is collecting data is made up of luminosity blocks corresponding to around a minute of data-taking each. A good luminosity block defines whether those data are good for physics analyses. This requirement can be different if, for example, an analysis does not make use of a certain phase space of the detector that was not operating nominally.



(a) 2015

(b) 2016

Figure 3.1.: The total integrated luminosity delivered by the LHC and recorded by the ATLAS detector in 2015 and 2016 [115].

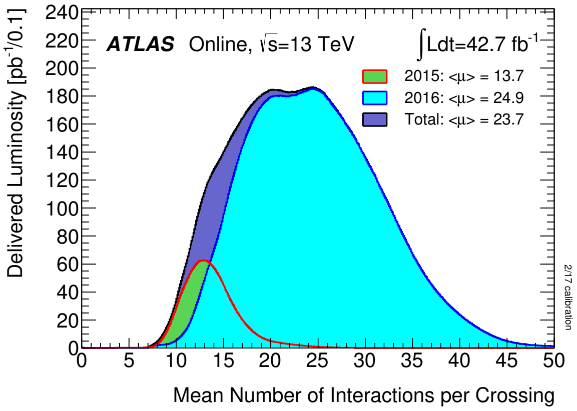


Figure 3.2.: The pileup profile for 2015 and 2016 data recorded with the ATLAS detector [115].

Event Generation

Event generation is the first step, which essentially has four "ingredients". Protons are composite particles made up of quarks and gluons, together forming partons. When protons collide Parton Distribution Functions (PDF) are needed to determine the fraction of momenta that these partons have. These PDFs are experimentally determined and are dependent on the CME. The PDF sets that are used in this analysis are: `NNPDF2.3LO` [116], `NNPDF3.0NLO` [117] and `CT10` [118]. PDFs are used as parameterisations to the next step, the hard scattering process, which is the Matrix Element (ME) calculation. The ME is the perturbative approach to modelling the process of interest, which can be represented by Feynman Diagrams. Different orders of precision can be modelled, with higher precisions being computationally more expensive and sometimes not practical. Amplitudes or probabilities for certain processes are a result of the ME calculation. Once the ME has been calculated and final state particles produced accordingly, ME particles are showered. Partons emit photons or gluons via the QCD or QED mechanisms due to bremsstrahlung, which can cause multiple cascades of particles. These then need to be taken into account when calculating the final kinematic distributions for an event. The process of adding QCD and QED bremsstrahlung is the showering of the ME. The last stage, called hadronisation, is when hadrons are formed from partons and are allowed to decay until stable particles are formed. This is due to QCD confinement which says quarks and gluons can not exist as individual particles, but need to form hadrons.

The steps above can in full or in part be carried out by what is called a MC generator. They can be interfaced to different PDFs or showering algorithms if necessary. The main MC generators used in this analysis are: `MadGraph5_aMC@NLO` (`MG5_aMC`) [119], `SHERPA` [120], `POWHEG-BOX v1` and `POWHEG-BOX v2` [121–125], and `PYTHIA 8` [126]. `SHERPA` provides its own showering and hadronisation model, while other generators need to be interfaced to a generator like `PYTHIA 8` or `HERWIG 7` [127, 128].

There are many parameters that are not necessarily well modelled in the above steps. MC tuning is needed to account for unforeseen circumstances such as multiple parton interactions or an increase in the strong coupling constant at low energies. This is done using tunes provided through empirical studies to data. The tunes used in this analysis are `A14` [129] and `Perugia2012` [130].

Detector Simulation

The simulation of particles through the ATLAS detector is done using GEANT 4 [131]. This step is the most CPU intensive as the interaction of each particle and its decay products needs to be simulated with the respective parts of the detector. The simulation of a single event can take anywhere from around 2 minutes to about 8 minutes depending on the complexity of the event. Millions of events for a given process need to be simulated. As previously mentioned, pileup plays an important role in the performance of object reconstruction and needs to be simulated accurately. `PYTHIA 8` is responsible for simulating the pileup due to QCD processes. Tuneable parameters called `A2` [132] and the `MSTW2008LO` [133] parton distribution functions (PDF) set are used. Techniques such as using pre-simulated electromagnetic showers can be used with a small penalty on accuracy, and as mentioned in Chapter 1.2, further research using ML techniques is being done to significantly speed up the simulation process by several orders of magnitude.

Object reconstruction

Detector properties combined with software algorithms are able to reconstruct objects for real and simulated data that pass through the detector (as shown in Figure 2.3). In general this is done using the ATLAS software infrastructure [134]. Object reconstruction is explained in more detail in Chapter 4.

3.2.1. Nominal signal and background MC samples

The $t\bar{t}\gamma$ signal sample is modelled at LO using `MG5_aMC` and the `NNPDF2.3LO` pdf set. Parton showering is simulated with `Pythia 8` using the `A14` tune. The five flavour scheme is used where all quark masses except the top mass are set to 0. The top quark mass is set to 172.5 GeV with a decay width of 1.320 GeV and the fine structure constant is set to 1/137. A set of kinematic cuts is needed to avoid infrared and collinear divergences. The photon and lepton transverse momenta are required to be larger than 15 GeV, while the absolute value of the pseudorapidity is required to be less than 5.0. A minimum ΔR distance of 0.2 is required between the photon and any charged particle in the final state. Higher order predictions of the $t\bar{t}\gamma$ process are incredibly CPU intensive and are not realistic to compute. NLO *k-factors* are used as a correction factor to account for higher orders in the QCD calculation (Section 3.2.2).

The nominal $t\bar{t}$ MC sample is generated with `PowHeg-Box v2` with `NNPDF3.0NLO`, and interfaced to `Pythia 8` using the `A14` tune and `NNPDF2.3LO` set.

The production of W/Z+jets is done with `Sherpa 2.2.1` while the dedicated W/Z+jets+γ ($W/Z\gamma$) samples are generated with `Sherpa 2.2.2`. Both use the `NNPDF3.0NLO` pdf set.

The single top-quark processes (t-, s- and tW-channels) are produced with `PowHeg-Box v1`. Each channel is generated using the `CT10` PDF set and interfaced to `Pythia 6`, with the t-channel also using the `Perugia2012` tune.

The diboson (WW, WZ and ZZ) samples are simulated using `Sherpa 2.1` with the `CT10` pdf set. The $t\bar{t}$+Z/W ($t\bar{t}V$) samples are produced with `MG5_aMC`, interfaced to `NNPDF3.0NLO` PDF set and showered with `Pythia 8`.

3.2.2. Signal sample NLO normalisation

The nominal $t\bar{t}\gamma$ MC sample at reconstruction level is LO as explained above. Thus, so called k-factors are used to correct the prediction to NLO precision. This calculation has been done in the approximation of a stable top quark [135], and extended to allow the decay of top quarks at $\sqrt{s} = 14$ TeV [136]. The authors from [136] performed dedicated calculations for $\sqrt{s} = 13$ TeV in the single-lepton and dilepton channels, which are used in this analysis. The NLO calculation was performed in the fiducial regions defined in Chapter 5.1.1 at parton level, separately for the single-lepton and dilepton channels. The LO calculation was performed at particle level[3] and care is taken so that objects at particle level correspond to those at parton level. For example, photons and leptons should be produced from the matrix element and not the parton shower. The k-factor is calculated as

$$k\text{-factor} = \frac{\sigma_{\text{NLO}}^{\text{parton}}}{\sigma_{\text{LO}}^{\text{particle}}}. \tag{3.1}$$

Systematic uncertainties in this theoretical calculation enter from sources such as the renormalisation and factorisation scales, PDFs, jet cone sizes (to evaluate the impact of additional radiation) and multiple parton interactions. This yields different systematic uncertainties in the single-lepton and dilepton channels, with the total uncertainty equating to 20% and 15%, respectively. Future versions of this analysis should place high priority on reducing these theoretical uncertainties. For more information on the k-factors and NLO cross-section calculation, refer to [1].

---

[3] Particle level is defined as events generated in the hard-process before detector simulation and reconstruction occurs.

CHAPTER 4

Object reconstruction and identification

Crucial to any successful analysis in HEP is the localisation, reconstruction, and identification of the physics objects within the ATLAS detector. A combination of the detector's geometry and material composition as well as advanced algorithms allow physicists to determine the object type, be it a photon, electron, muon, jet, or missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$). The different elements in the detector that play crucial roles in determining each object have been shown in Chapter 2.2. This Chapter presents an overview of the different algorithms used in ATLAS.

4.1. Electron and photon objects

Due to the similarity between electrons and photons and the comparable signatures they leave in the calorimeters, the reconstruction of these objects happens simultaneously. Information from the ID, presented in Chapter 2.2.2, and both the electromagnetic calorimeter (EMC) and the hadronic calorimeter (HCAL), presented in Chapter 2.2.3, are used. An overview of the reconstruction of electrons and photons follows.

- First, a *seed cluster* in the EMC is formed. A grid of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ (corresponding to the granularity in the barrel region of the EMC middle layer) is formed in the η-ϕ plane of the EMC. A *tower* for each element in the grid is formed by integrating the total energy over the presampler, first, second and third layers of the EMC. Then, for every element in the defined EMC space, a sliding window algorithm [137] is used to form EM energy cluster candidates. A defined region of 3×5 cells[1] in the η-ϕ plane is moved 0.025 steps in either the η or ϕ direction and searches for energy deposits whose summed transverse energy is greater than 2.5 GeV. These energy deposits form a seed cluster. Often candidate seeds can overlap and so an overlap removal is defined. If two candidate seeds overlap within an area of $\eta \times \phi = 5 \times 9$ cells, the candidate with the larger transverse energy is taken. If the transverse energy of the clusters is within 10% of each other, then the cluster with the largest individual tower is taken.

- Track reconstruction in the ID proceeds in two steps. The first is a pattern recognition technique which is based on energy loss due to interactions with detector material [138]. The second step is a track fit. The ATLAS Global χ^2 Track Fitter [139] is used with either a pion or an electron hypothesis. These tracks are then loosely matched to the seed clusters in the EMC.

---

[1] This cluster size changes slightly for barrel versus endcap parts of the EMC.

- These loosely-matched tracks can have hits in different parts of the detectors. From these hits, vertices are used for conversion reconstruction. For example, a converted photon can be associated with two-track conversion vertices. An unconverted photon would have no associated tracks.
- Finally, based on track information, conversion vertices and the seed cluster, an algorithm decides whether an object will be reconstructed as an electron, photon or both.

Central to electrons and photons is the use of "shower shape" variables. These are various properties of the electromagnetic showers that are exploited to allow for effective object differentiation. A list with explanations of these variables is shown in Table 4.1. References to these shower shape variables will be made in this, and later chapters.

Name	**Description**
Hadronic leakage	
R_{had} or R_{had1}	Transverse energy leakage in the HCAL normalised to transverse energy of the photon candidate in the EMC. In the region $0.8 \leq \lvert\eta\rvert \leq 1.37$, the entire energy of the photon candidate in the HCAL is used (R_{had}), while in the region $\lvert\eta\rvert < 0.8$ and $\lvert\eta\rvert > 1.37$ the energy of the first layer of the HCAL is used (R_{had1})
Energy ratios and width in the second layer of EMC	
R_η	Energy ratio of 3×7 to 7×7 cells in the $\eta \times \phi$ plane
R_ϕ	Energy ratio of 3×3 to 3×7 cells in the $\eta \times \phi$ plane
$w_{\eta 2}$	Lateral width of cluster in $\eta \times \phi = 3 \times 5$: $\sqrt{\frac{\sum_i E_i \eta_i^2}{\sum_i E_i} - \left(\frac{\sum_i E_i \eta_i}{\sum_i E_i}\right)^2}$
Energy ratios and widths in the first (strip) layer of EMC	
$w_{\eta 1}(w_{s3})$	Energy weighted width using 3 strips around the maximum: $\sqrt{\frac{\sum_i E_i (i - i_{\text{max}})^2}{\sum_i E_i}}$
$w_{\text{tot,s1}}(w_s)$	Energy weighted width using 20 strips around the maximum, see $w_{\eta,1}$
f_{side}	Energy within 7 strips without 3 central strips normalised to energy in 3 central strips
E_{ratio}	Ratio between difference of first 2 energy maxima divided by their sum ($E_{\text{ratio}} = 1$ if there is no second maximum)
ΔE	Difference between the second energy maximum and the minimum between first and second maximum ($\Delta E = 1$ if there is no second maximum)

Table 4.1.: Summary of the shower shape variables used for electron and photon reconstruction and identification. Summations over i run over the cluster constituents.

From here, different requirements on photon and electron identification and isolation determine the purities and efficiencies in which these objects are selected.

4.1.1. Photons

Photon identification [140] is performed by using rectangular cuts on the shower shape variables shown in Table 4.1. As shown in Figure 2.8, photons originating from different processes are characterised by the shapes of the electromagnetic shower developments through the EMC and HCAL. The variables in Table 4.1 characterise this. Two identification reference points are calibrated within ATLAS; *Loose* and *Tight*. The *Loose* selection is based on information provided by the second layer of the EMC as well as the HCAL. This corresponds to the variables R_{had}, R_{had1}, R_η and ω_{η_2}. This analysis mostly uses the *Tight* selection, which makes use of all shower shape variables in Table 4.1[2]. Information from the strip layer in the EMC is included, thus *Tight* events are a subset of the *Loose* events. To account for the slightly different shower shape energy deposits between converted and unconverted photons, the working points are optimised separately. To account for the calorimeter geometry and the effect of slightly different materials, the working points are also calibrated as a function of η.

To obtain a purer sample of prompt photons, isolation plays a crucial role. It significantly reduces the contribution from photons both from hadronic decays or hadrons misidentified as photons. Isolation for photons is based on the transverse energy in a ΔR size cone around the photon (excluding the photon's transverse energy). Transverse energy can originate either from tracks or from calorimeter deposits.

- $E_{\text{T}}^{\text{topoetcone}x}$ is defined as the calorimeter isolation and is obtained from summing the transverse energy of the topological clusters around a cone of $\Delta R < \frac{x}{100}$. Contributions from the photon candidate itself and pileup are subtracted.

- $p_{\text{T}}^{\text{cone20}}$ is defined as the track isolation and is obtained by summing the transverse momenta of all tracks (excluding those from photon conversions) with $p_{\text{T}} > 1$ GeV around of cone of $\Delta R < 0.2$. A distance requirement to the primary vertex along the beam axis, $|z_0 \text{sin}\theta| < 3$ mm, must also be fulfilled.

There are three provided working points within ATLAS, which are shown along with their definitions in Table 4.2. This analysis generally makes use of the *FixedCutTight* isolation working point.

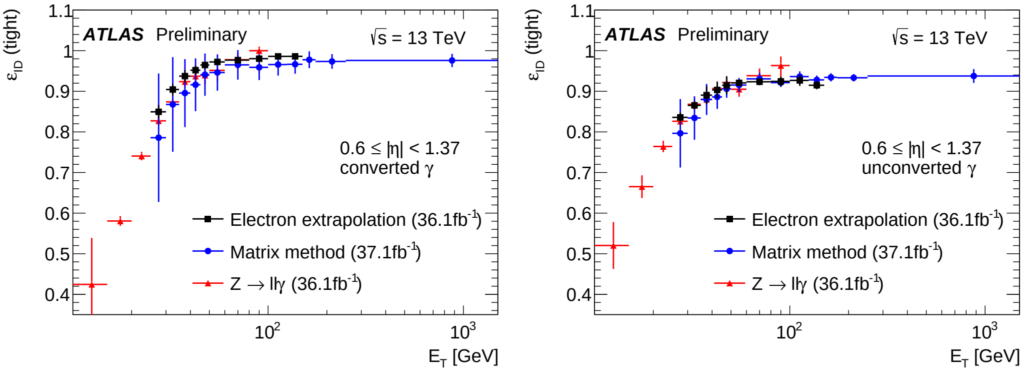
Working point	**Calorimeter isolation**	**Track isolation**
FixedCutTightCaloOnly	$E_{\text{T}}^{\text{topoetcone40}} < 0.022 p_{\text{T}} + 2.45$ [GeV]	none
FixedCutTight	$E_{\text{T}}^{\text{topoetcone40}} < 0.022 p_{\text{T}} + 2.45$ [GeV]	$p_{\text{T}}^{\text{cone20}}/p_{\text{T}} < 0.05$
FixedCutLoose	$E_{\text{T}}^{\text{topoetcone20}} < 0.065 p_{\text{T}}$	$p_{\text{T}}^{\text{cone20}}/p_{\text{T}} < 0.05$

Table 4.2.: The 3 defined isolation working points for photons.

The efficiencies of *Tight* photons are measured in data using three methods. This enables coverage of a wide E_{T} that is not possible otherwise. The cleanest source of prompt photons is measured in $Z \to ll\gamma$ decays for the efficiencies of $E_{\text{T}} = 10$ GeV up to approximately 100 GeV. Event yields after this are too low. For photons of about $25 < E_{\text{T}} < 150$ GeV the *electron extrapolation* method is used from $Z \to ee$ decays. Here, the showers that electrons make in the calorimeters are slightly altered to reflect those of photons. The last method in which efficiencies are measured applies to photons in the range 25 GeV $< E_{\text{T}} <$ 1.5 TeV. A photon sample is collected using single-photon triggers. The efficiency is essentially derived by comparing those events that pass the *Tight* track-based isolation criteria to the full *Tight* photon sample. In all the above methods various isolation criteria for the photons are applied. Figure 4.1 shows the efficiencies for the range $0.6 \leq |\eta| < 1.37$ for converted and unconverted *Tight* photons. The three methods used to obtain the efficiencies complement each other at different energy scales. The error bars include the statistical and systematic uncertainties associated with each method. The large spikes are from fewer statistics. These efficiency measurements are combined to provide scale factors to correct for data/MC mis-modelling.

---

[2] In some studies, such as for the hadronic fake background in Chapter 5.4 this requirement is reversed.



(a) Converted photons (b) Unconverted photons

Figure 4.1.: The efficiencies for *Tight* photons shown for the three complementary methods [141].

An important aspect to the shower shape variables is that of the correction factors that are applied to simulated samples. Samples used in photon identification are simulated with GEANT4 and the ATLAS reconstruction pipeline. However, GEANT4 reproduces the shower shapes of photons imperfectly and so an offset occurs when comparing these MC samples to real data. Corrections are derived from $Z \to ll\gamma$ and γ+jet MC samples which take into account systematic and statistical errors. An example is shown in Figure 4.2. This shows the MC simulation before and after corrections have been applied, compared to the measured data for a given bin of η. Corrections to shower shape variables are essential for photon identification in ATLAS. All above efficiencies are performed using the corrected shapes, and from hereon, this analysis only makes use of the corrected variables.

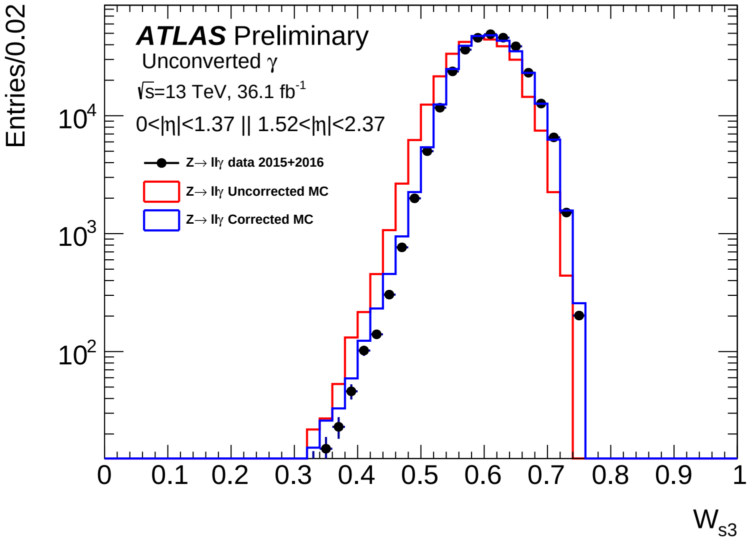


Figure 4.2.: The ω_{s3} shower shape variable. This compares the original simulated events to real data, as well as the final corrected simulated events [141].

4.1.2. Electrons

Electron identification [142] follows an MVA likelihood based approach in which variables from Table 4.1 are used as inputs. An overall probability that the object is a signal electron is calculated. These probabilities form a discriminant, which at different values or "working points" define different purities of electrons. Three working points are provided in ATLAS, which in increasing order of background rejection are *Loose*, *Medium* and *Tight*. This analysis makes use of the *Tight* working point. This equates to efficiencies (measured in $Z \rightarrow ee$ and $J/\psi \rightarrow ee$ events) of around 80% for $E_{\mathrm{T}} = 30$ GeV and increases with higher E_{T}.

Electrons are often required to fulfil isolation requirements. Two variables are defined; $E_{\mathrm{T}}^{\mathrm{cone}0.2}$ and $p_{\mathrm{T}}^{\mathrm{varcone}0.2}$.

- $E_{\mathrm{T}}^{\mathrm{cone}0.2}$ is the sum of the transverse energy of clusters formed in the calorimeters, within a cone of $\Delta R = 0.2$ around the electron cluster.

- $p_{\mathrm{T}}^{\mathrm{varcone}0.2}$ is the sum of the transverse momenta of all tracks within a cone of $\Delta R = \min(0.2, 10\ \mathrm{GeV}/E_{\mathrm{T}})$ around the candidate electron track.

This analysis makes use of the *Gradient* isolation working points, which uses cuts on the two above variables to target specific efficiencies. For *Gradient* isolation, this targets a total isolation efficiency of electrons (between the combination of calorimeter and track based isolation) of 90% (99%) at 25 GeV (60 GeV).

4.2. Muons

Muon reconstruction [143] in ATLAS relies on information from the ID (Chapter 2.2.2), the muon spectrometer (MS, Chapter 2.2.4) and calorimeters (Chapter 2.2.3). The ID provides track measurements and information related to distances from the IP. The MS provides accurate momentum measurements with high resolutions for muons. The calorimeters supplement the MS and ID with information regarding energy loss. From combining information from the sub-detectors, four main types of muons are identified:

- Combined muons: Tracks are present in both the MS and the ID for which reconstruction is performed separately. A combined track is formed with a global refit to the hits of both sub-detectors.

- Segment-tagged muons: A track in the ID can be extrapolated to the MS, but is only associated with one track segment in the MDT or CSC chambers. These muons are often low p_{T} tracks, or muons that fall into the reduced MS acceptance regions.

- Calorimeter-tagged muons: Tracks in the ID and a minimum-ionisation energy deposit in the calorimeter can be identified as a muon. The purity of this type of muon is very low, but it recovers acceptance in areas of the ATLAS detector in which the MS has poor coverage.

- Extrapolated muons: Tracks in the MS that are not matched to tracks in the ID are extrapolated back to the IP taking into account the energy loss in the calorimeter. These muons are predominantly used in the region $2.5 < |\eta| < 2.7$, for which the ID does not cover.

This thesis makes use of *Medium* identification criteria. This default selection in ATLAS is one that minimises the systematic uncertainties during reconstruction and calibration. Combined muon tracks are required to have at least three hits in at least two MDT layers. Extrapolated muon tracks are required to have at least three hits in MDR/CSC layers and are used only in the region $2.5 < |\eta| < 2.7$. Contamination from hadrons misidentified as muons is reduced by a requirement on the ratio of the muon's p_{T} in the MS and ID. Efficiencies are measured using a tag-and-probe method in almost pure $J/\psi \to \mu\mu$ or $Z \to \mu\mu$ events with at least one *Medium* muon. In the region $|\eta| < 2.5$, the efficiency is measured to be close to 99% for $p_{\mathrm{T}} > 5\ \ \mathrm{GeV}$.

Prompt muons from W- and Z-boson decays are more isolated then muons originating from semileptonic decays in jets. Thus, isolated muons are important when selecting final channels as well as suppressing background. Isolation makes use of track and calorimeter based isolation variables. The track based variable, $p_{\mathrm{T}}^{\mathrm{varcone30}}$, is defined as the scalar sum of all transverse momenta of the tracks with $p_{\mathrm{T}} > 1\ \ \mathrm{GeV}$ (excluding the muon track itself) in a cone of $\Delta R = \min(10\ \ \mathrm{GeV}/p_{\mathrm{T}}^{\mu}, 0.3)$ around the transverse momentum of the muon, p_{T}^{μ}. The calorimeter based isolation variable, $E_{\mathrm{T}}^{\mathrm{topoetcone20}}$, is the sum of the transverse energy from topological clusters in a cone of $\Delta R = 0.2$ around the muon (excluding the muon itself).

Relative isolation variables are used to define various working points that are used in different analyses. The *Gradient* based isolation working point is used in this thesis, described as cuts placed on $(p_{\mathrm{T}}^{\mathrm{varcone30}}/p_{\mathrm{T}}^{\mu}, E_{\mathrm{T}}^{\mathrm{topocone20}}/p_{\mathrm{T}}^{\mu})$. The efficiency for this working point (measured in $Z \to \mu\mu$ events) achieves greater than 90% (99%) efficiency for muons at 25 GeV (60 GeV).

4.3. Jets

Due to colour confinement and the requirement for particles to have zero net colour charge, quarks and gluons can not exist by themselves, but rather form hadrons through the process of hadronisation (Chapter 1). These hadrons then decay into a collimated spray of particles termed *jets*. The reconstruction and identification of jets [144] at the LHC only increases in importance as the pileup due to higher luminosities increases. They can form dominant backgrounds, or be flags for topological signatures in various processes. Jets are reconstructed using the anti-k_t algorithm [145] using a radius of $R = 0.4$. The anti-k_t algorithm is a sequential clustering algorithm based on a distance parameter for topological calorimeter clusters (topo-clusters) [146] i and j, and another distance parameter between topo-cluster i and the beam (B):

$$d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2})\frac{\Delta_{ij}^2}{R^2} \tag{4.1}$$

and

$$d_{iB} = k_{ti}^{-2}. \tag{4.2}$$

Here, R is the radius and $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, where k_{ti}, y_i and ϕ_i are the transverse momentum, rapidity and azimuthal angle of particle i. The minimum between d_{ij} and d_{iB} is taken. If the minimum distance is d_{ij}, the two topo-cluster entities are combined. If the minimum distance is d_{iB} then i is called a jet and removed from the list of entities. The process iterates over all topo-clusters until no entities are left. From the $1/k_{ti,j}^2$ terms, this tells us that harder particles are clustered with a higher priority. If no other hard jets are within a distance $2R$, it accumulates the softer particles resulting in a conical shape. If two hard particles are present, depending on the distribution of momentum, one may be more conical than the other. Essentially, soft particles do not alter the shape of the jet.

Corrections are applied to the jet energy scale (JES) [147] and the jet energy resolution (JER) [148] in the form of p_{T} and η dependent scale factors. The JES corrects the jet energy to that of the truth jet energy at the particle level. This essentially corrects the reconstructed jet energy and the orientation in the calorimeter.

Essential to the reconstruction of jets is the ability to reduce the contribution from pileup and enhance the jets that result from the hard process. For this, a multivariate tool called the Jet Vertex Tagger (JVT) is used [149]. The JVT defines the fraction of total tracks in the jet that are associated to the primary vertex.

4.3.1. *b*-jets

In top quark analyses b-jets are crucial since the branching ratio of $t \rightarrow W + b$ is completely dominant. However, tagging b-jets (b-tagging) is difficult due to busy environments. Specifically, distinguishing *light*-jets (u, d, s or gluons) and c-jets from b-jets uses MVA techniques using inputs containing many different sources of information.

b-jets originate from hadrons containing b-quarks. Compared to c- and *light*-quarks, b-quarks have unique properties that can be exploited. The b-hadron lifetime of around 1.5 ps allows the particle to travel further from the beam pipe before decaying, thus creating a secondary vertex. A second useful property of the b-quark is its large mass. Thus, when it decays it creates more charged particles.

A BDT algorithm called MV2c10 [150] is used for this purpose. The most important inputs are the tracks obtained from the ID. Additionally, there are outputs from three algorithms [151, 152] that get fed into MV2c10.

- Impact Parameter based algorithms (IP2D and IP3D): These algorithms make use of impact parameter information in the longitudinal (z_0) and transverse (d_0) planes. A likelihood is built using the signed significance of each (z_0/σ_{z_0}, d_0/σ_{d_0}). IP3D makes use of transverse and longitudinal impact parameters (thus making it 3D), while IP2D only uses the transverse impact parameters.

- Secondary Vertex finding algorithm (SV): This algorithm exploits the secondary vertex due to the longer decay lifetimes of b-quarks.

- Decay chain multi-vertex algorithm (JetFitter): This algorithm attempts to reconstruct the full b- and c- hadron decay chains. A Kalman filter algorithm tries to find the different vertices from which the b- and c-hadrons originate.

The jet p_{T} and η are also included in the training of MV2c10.

Several working points are defined in which analyses can specify the required b-jet efficiency. This is defined as $\epsilon_b = N_{b\text{-tagged}}/N_{b\text{-true}}$. In this analysis, the working point $\epsilon_b = 77\%$ is used in the main event selections.

Another way to use information from b-tagging is in the form of the *pseudo-continuous weights*. For each jet in an event, a b-tagging score from the MV2c10 algorithm is assigned. This creates a histogram of fixed bins corresponding to different efficiencies. *Pseudo-continuous* refers to the fact that there are only five bins that are calibrated and the distribution is not truly continuous. These are shown in Table 4.3. These distributions for leading, sub-leading and subsub-leading jets are used extensively in the Event-level Discriminator (Chapter 6.2).

4.4. Missing transverse energy

The ATLAS detector allows for some energy in collisions to go undetected. This can be due to technological imperfections or limited detector coverage, but also the nature of particles such as neutrinos. Often perceived as massless (a good approximation), neutrinos are not charged and do not leave any tracks in the ID or energy deposits in the calorimeters. They pass undetected through the detector. However, due to momentum conservation in the plane transverse to the

Bin #	ϵ_b
1	$> 85\%$
2	$77\% - 85\%$
3	$70\% - 77\%$
4	$60\% - 70\%$
5	$< 60\%$

Table 4.3.: The definition of the pseudo-continuous b-tagging weights by bin number.

beam axis of collider experiments, the overall transverse momentum of a collision should sum to 0. The residual energy is known as *missing transverse energy* ($E_{\mathrm{T}}^{\mathrm{miss}}$) [153]. The approximation of $E_{\mathrm{T}}^{\mathrm{miss}}$ is given by the negative vectorial sum of the momenta for each of the calibrated objects; electrons, photons, taus, muons, jets, and a soft term. The soft term enters as the sum of all transverse momenta for tracks that have not already been matched to a physics object, but are still associated to the primary vertex. Thus,

$$E_{x(y)}^{\mathrm{miss}} = -1 \cdot (p_{x(y)}^{e} + p_{x(y)}^{\gamma} + p_{x(y)}^{\tau} + p_{x(y)}^{\mu} + p_{x(y)}^{\mathrm{jets}} + p_{x(y)}^{\mathrm{soft}}) \tag{4.3}$$

$$E_{\mathrm{T}}^{\mathrm{miss}} = \sqrt{(E_{x}^{\mathrm{miss}})^2 + (E_{y}^{\mathrm{miss}})^2}. \tag{4.4}$$

The $E_{\mathrm{T}}^{\mathrm{miss}}$ can essentially be thought of as a particle originating from the hard scattering. This is useful for some BSM searches, which predict other minimally-interacting particles.

4.5. Object overlap removal procedure

Particle tracks and energy deposits can in some cases be used to reconstruct multiple objects. To prevent double counting a standard overlap removal procedure is performed. First, electrons that share tracks with any other electrons are removed. Any electron found sharing a track with a muon is then removed. Any jet found within $\Delta R < 0.2$ of an electron is removed. Then, if any electron is found within $\Delta R < 0.4$ of a jet it is removed. Any jet with less than 3 tracks associated to it found within $\Delta R < 0.2$ of a muon is removed. Then, for any muon subsequently found within $\Delta R < 0.4$ of a jet is removed. Any photon found within a $\Delta R < 0.4$ of an electron or a muon is removed. Any jet found within $\Delta R < 0.4$ of a *Loose* photon is removed.

4.6. Event overlap removal and categorisation

Double counting of prompt photons may occur due to the overlap of MC samples (Chapter 3.2), for example, in $t\bar{t}\gamma$ and $t\bar{t}$, and $W/Z\gamma$ and W/Z+jets processes. Thus, a truth based matching scheme is used to remove those types of events. The *truth origin* defines the mother type of the particle of interest while *truth type* indicates the actual particle type. It is possible to classify the type of photon into four general categories based on these truth labels. These will be used extensively and are called: *signal*, $e \to \gamma$ *fakes*, *hadronic fakes* and *Other prompt*.

$e \to \gamma$ fakes

Photons are classified as “$e \to \gamma$ fakes” when a reconstructed photon is truth matched to an electron or if there is a truth electron in a cone of radius 0.2 around the reconstructed photon.

Hadronic fakes

Photons are classified as "hadronic fakes" when the particle's truth origin corresponds to a hadron. The truth type demands that this photon originates from a boson or a lepton and does not originate from initial or final state radiation. A further requirement is that the reconstructed photon is not an $e \rightarrow \gamma$ fake based on the above mentioned criteria.

Signal

Signal events are categorised by reversing the requirements placed on the $e \rightarrow \gamma$ fake and hadronic fake categories. Specifically, the reconstructed photon should not be classified as originating from a hadron, should not be a "background photon" and should not be truth matched as or to an electron. These requirements are met by the nominal signal MC.

Other prompt

Photons classified as "Other prompt" are defined in the same way as those in signal. The difference being that this category does not include signal MC but rather diboson, single-top and $t\bar{t}V$ MC samples. Depending on the channel, this category can also contain $Z\gamma$ for single-lepton and $W\gamma$ for dilepton channels. This is to account for when the $Z\gamma$ or $W\gamma$ backgrounds are negligible or dominant in the respective channel.

The signal and background processes introduced above will be presented in more detail in Chapter 5.

CHAPTER 5

Top-quark pairs and photons at $\sqrt{s} = 13$ TeV

This chapter focuses on the general analysis strategy and the signal and backgrounds for the $t\bar{t}\gamma$ process at $\sqrt{s} = 13$ TeV. Section 5.1 presents the method of extracting the cross-sections using a maximum likelihood fit. Section 5.2 presents the triggers and event selections used to obtain a signal region. This is followed by the $e \to \gamma$ fake (Section 5.3) and hadronic fake (Section 5.4) background data-driven scale factor extractions using the tag-and-probe and ABCD method, respectively. Section 5.5 shows the data-driven estimate of the fake lepton background using the matrix method. Section 5.6 presents the prompt photon background contributions. Finally, Section 5.7 presents kinematic distributions in the signal region with all scale factors derived in the previous sections applied.

5.1. Analysis strategy

This section discusses the strategy and methods to obtain fiducial cross-section results. The fiducial region and theoretical NLO predictions are presented in Section 5.1.1. Then, the method to extract the cross-section via a maximum likelihood fit is discussed in Section 5.1.2 Lastly, the different types of fitting scenarios are presented in Section 5.1.3.

Important to note is that in addition to fiducial cross-section measurements, the analysis team also carried out *differential* cross-section measurements. These are not presented in this thesis, and more information can be seen in [1].

5.1.1. Fiducial cross-section measurements

The final cross-section measurements are carried out in a subset of the phase space defined by the event selections presented further in this chapter. This is labelled the *fiducial region*. This new region is defined at particle level and is designed to be as close as possible to the region defined at reconstruction level (when detector effects and sensitivities come into play), thus, it is used to reduce the impact that detector effects may have on the measurement.

The object requirements at particle level are defined as follows: Leptons do not originate from hadron decays and have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$. Photons have $E_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.37$ and must not be from hadron decays or dressed to a lepton. The photon is dressed to the lepton if it does not originate from a hadron and is within $\Delta R < 0.1$ around the lepton. They must also be isolated in that $p_{\mathrm{T}}^{\mathrm{cone}} < 0.1$. Jets are re-clustered with the anti-k_t algorithm with a radius of $\Delta R = 0.4$ and have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$. A ghost matching scheme [154] is used to determine the flavour of the jets.

The event requirements are as follows: The single-lepton (dilepton) channels require at least four (two) jets with at least one of them being tagged as coming from a b-hadron. Exactly one (two) leptons are required. Events are removed if $\Delta R(\gamma, jet) < 0.4$ and $\Delta R(\gamma, lep) < 1.0$.

We can define two quantities; the acceptance (A),

$$A = \frac{N_{\mathrm{gen}}^{\mathrm{fid}}}{N_{\mathrm{gen}}^{\mathrm{all}}}, \tag{5.1}$$

and the correction (C),

$$C = \frac{N_{\mathrm{reco}}}{N_{\mathrm{gen}}^{\mathrm{fid}}}. \tag{5.2}$$

The acceptance factor (A) gives the ratio of the generated signal particle level events that pass the fiducial selections ($N_{\mathrm{gen}}^{\mathrm{fid}}$) to the total number of generated signal events ($N_{\mathrm{gen}}^{\mathrm{all}}$). It is used to extrapolate the theoretical cross section to a smaller phase space. The correction factor gives the ratio of the signal events that pass the full event selection at reconstruction level (N_{reco}) to the generated particle level signal events that pass the fiducial selections. It essentially turns the final cross section into a fiducial cross section.

Experimentally, a fiducial cross section for a given channel is determined from the equation

$$\sigma = \frac{N_{\mathrm{data}} - \sum_i N_{\mathrm{bkg},i}}{\mathcal{L} \times C}, \tag{5.3}$$

where N_{data} is the number of data, $N_{\mathrm{bkg},i}$ is the sum of all background contributions, $\mathcal{L}$ is the luminosity, and C is the correction factor from Equation 5.2.

In practice, a result often quoted is the scaling of the measured cross section to the (NLO) theoretical prediction. The so called *signal strength*, μ, is defined as

$$\mu = \frac{\sigma_{\mathrm{fid}}}{\sigma_{\mathrm{fid}}^{\mathrm{NLO}}}, \tag{5.4}$$

where the fiducial NLO theory prediction is

$$\sigma_{\mathrm{fid}}^{\mathrm{NLO}} = \sigma_{\mathrm{total}}^{\mathrm{LO}} \times A \times k. \tag{5.5}$$

From the k-factors and the acceptances defined earlier, we can calculate the theoretically predicted NLO cross sections for each channel with their total uncertainties. These values are summarised in Table 5.1. It is important to note that tau leptons decay hadronically ($\tau^- \to \nu_\tau d\bar{u}$) and leptonically ($\tau^- \to \nu_\tau l^- \bar{\nu}_l$, $l = e, \mu$) and so a dedicated measurement in the "τ-channel" is difficult. In this analysis, in addition to the single-lepton and dilepton channels, five further channels are defined (e+jets, μ+jets, ee, $\mu\mu$ and $e\mu$) in which cross-section measurements are performed. These channels do not distinguish electrons, muons and jets from tau leptons. The theoretical cross section calculations are consistent and follow this same strategy. Thus, the e+jets, μ+jets, ee, $\mu\mu$ and $e\mu$ measurements should *not* be thought of as measurements made in the $t\bar{t}$ channels presented in Figure 1.2.

	single-lepton		dilepton		
Cross section [fb]	495 ± 99		63 ± 9		
	e+jets	μ+jets	ee	$e\mu$	$\mu\mu$
Cross section [fb]	247 ± 49	248 ± 50	16 ± 2	31 ± 5	16 ± 2

Table 5.1.: The theoretical fiducial NLO cross sections for each channel with their total uncertainties.

5.1.2. Maximum likelihood fit

To extract the signal strength described in Equation 5.4 a binned maximum likelihood fit is performed on a distribution believed to be able to effectively separate signal and background processes. The Event-level Discriminator (ELD) serves this purpose and will be presented in Chapter 6.2. A general likelihood can be described as

$$\mathcal{L} = \mathcal{L}(\mu; \mathrm{ELD}, \theta) = P \times G. \tag{5.6}$$

This gives the probability of observing a certain signal signal strength for the ELD distribution given a set of systematic uncertainties. The first term on the right refers to a Poisson distribution modelling the signal region. The second term refers to a Gaussian distribution which constrains the nuisance parameters (NPs). This can be expanded:

$$\mathcal{L} = \prod_{j \in \mathrm{bins}} P(N_j^{\mathrm{data}} | \mu \cdot N_j^{\mathrm{sig.}}(\vec{\theta}) + \sum_{b \in \mathrm{bkgs.}} N_j^b(\vec{\theta})) \times \prod_t G(0|\theta_t, 1). \tag{5.7}$$

The first Poisson distribution gives the probability of observing a number of data events (N_j^{data}) for a given number of signal ($\mu \cdot N_j^{\mathrm{sig.}}$) and total background ($N_j^b$) events in bin j of the ELD. The number of signal and background events in each bin are parameterised by all NPs ($\vec{\theta}$), where each NP (θ_t) is constrained by a Gaussian distribution. Each NP is constrained with an “up” and “down” variation[1], where each variation is described as $\pm 1\sigma$ away from the nominal value.

A NP can also enter as an un-constrained or free floating parameter where no prior is assumed. A typical example for a free floating NP is a dominant background’s normalisation. The NPs give the signal and background the freedom to adjust their predictions to observed data. The likelihood function is then maximised where the global maximum is the point at which the parameters are estimated most accurately. In practice the negative log-likelihood is minimised, with $\pm 1\sigma$ taken as the uncertainties. The likelihood and fits were built and performed using custom wrappers on top of the RooFit framework [155].

5.1.3. Fit scenarios

There are multiple ways to define the channel that enters the maximum likelihood fit. An overview of the scenarios is shown in Figure 5.1 and explained below.

For the individual channels this is straightforward and separate likelihood fits in each of the five channels (e+jets, μ+jets, ee, $e\mu$ and $\mu\mu$) are performed. When it comes to carrying out single-lepton and dilepton channel fits there are two options considered.

- Merged: This is when histograms from the e+jets and μ+jets channels are essentially added together to form the single-lepton channel, and histograms from ee, $e\mu$ and $\mu\mu$ are added to form the dilepton channel. The advantage (and ultimately deciding factor) to using this method is that only one signal region is used in the fit, meaning there is only one final signal strength and set of histograms for each single-lepton and dilepton channel. This makes differential measurements more straightforward. For some background studies and the training of the ELD, the merged channels were used to take advantage of the increased statistics.

---

[1] The up and down choice of names is arbitrary and do not necessarily reflect a positive and negative variation. At times both variations can be positive or negative.

- Combined: This is when multiple signal regions are fit simultaneously, which allows the possibility for nuisance parameters to be more effectively constrained. Essentially, more terms are added to the likelihood in Equation 5.7. For example, a "combined" fit could be carried out on the e+jets, μ+jets, ee, $e\mu$ and $\mu\mu$ histograms, which outputs one signal strength but five sets of new histograms. This strategy is used only for a 5-channel combined fit to obtain an "inclusive" cross-section.

Figure 5.1 shows the chosen scenarios used for the final fits, which have red borders. This includes the five individual channels, the single-lepton and dilepton merged fits, and the "5-channel combined fit".

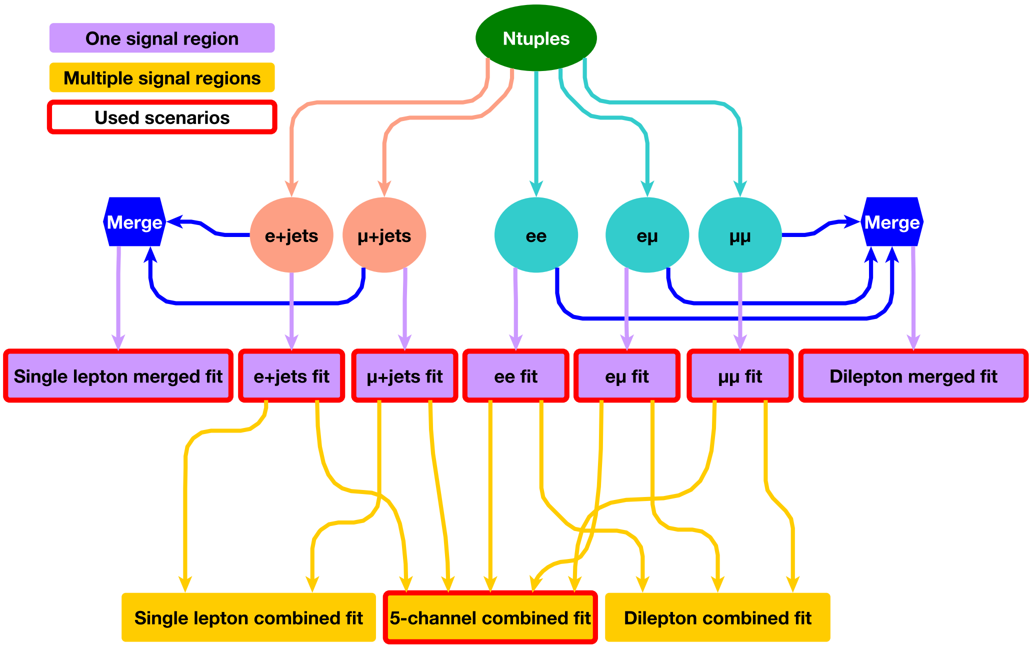

One signal region
Multiple signal regions
Used scenarios
Ntuples
Merge
e+jets
μ+jets
ee
eμ
μμ
Merge
Single lepton merged fit
e+jets fit
μ+jets fit
ee fit
eμ fit
μμ fit
Dilepton merged fit
Single lepton combined fit
5-channel combined fit
Dilepton combined fit

Figure 5.1.: An overview of the different scenarios for how fits can be carried out. Purple (second to last row) indicates a merged or single region fit, yellow (last row) indicates a combined or multiple region fit. The boxes with borders indicate the final chosen strategy.

5.2. Signal region event selections

Selections are placed on the objects defined in Chapter 4 to suppress background contributions and thus enrich the $t\bar{t}\gamma$ purity. The strategy is to follow the event selections required for $t\bar{t}$, but with the additional requirements on the photon. The triggers used are presented first, followed by the selection cuts.

Triggers

Different HLT triggers are used for data taking periods in 2015 and 2016 [156–158]. For this analysis single lepton triggers for electrons and muons are used to define parameter spaces of interest. These are described below.

For the 2015 electron candidates, the first trigger requires a minimum E_{T} threshold of 24 GeV and a *Medium* identification requirement. It is seeded by a L1 trigger requiring an E_{T} of 20 GeV and applies a veto against any energy deposited in the hadronic calorimeter (L1EM20VH). At $E_{\mathrm{T}} > 60$ GeV, this is superseded by a trigger without the L1 seed, and at $E_{\mathrm{T}} > 120$ GeV a trigger with *Loose* identification is used. For 2016 electron candidates of threshold $E_{\mathrm{T}} >$ 26 GeV, a *Tight* identification, and a *Loose* isolation is required. At $E_{\mathrm{T}} > 60$ GeV the isolation requirement is dropped and the identification relaxed to *Medium*. At $E_{\mathrm{T}} > 140$ GeV the identification is further relaxed to a *Loose* identification requirement. In all cases the 2016 triggers no longer require transverse impact parameter cuts.

For the 2015 muon candidates, the trigger requires a minimum of $p_{\mathrm{T}} \geq 20$ GeV with a *Loose* isolation criterion. This trigger is seeded by the L1 trigger where the muon candidate has a minimum p_{T} of 15 GeV (L1MU15). Once the p_{T} is greater than 50 GeV, the trigger is superseded by another trigger that does not require isolation. This compensates for a decrease in efficiency due to higher p_{T} muons. The 2016 muon trigger requires candidates to have transverse momentum greater than 26 GeV. A *Medium* isolation is also required. Again, at p_{T} greater than 50 GeV the trigger is superseded by another trigger that does not require isolation. A summary of all triggers used is shown in Table 5.2.

Year	electron triggers			
	p_{T} threshold [GeV]	Identification menu	Isolation menu	L1 seed
2015	$\geq$24	*Medium*	None	L1EM20VH
	$\geq$60	*Medium*	None	-
	$\geq$120	*Loose*	None	-
2016	$\geq$26	*Tight*	Gradient(*Loose*)	-
	$\geq$60	*Medium*	None	-
	$\geq$140	*Loose*	None	-
Year	muon triggers			
	E_{T} threshold [GeV]	Identification menu	Isolation menu	L1 seed
2015	$\geq$20	None	Gradient(*Loose*)	L1MU15
	$\geq$50	None	None	-
2016	$\geq$26	None	Gradient(*Medium*)	-
	$\geq$50	None	None	-

Table 5.2.: The requirements of the lepton triggers considered in the event selections.

Signal region cuts

After potentially interesting data has been selected using single lepton triggers, using objects reconstruction methods as described in Chapter 4, further selection cuts are made to enhance the purity of the $t\bar{t}\gamma$ signal.

The presence of electrons and muons in an event is characterised by the triggers mentioned above. Additional cuts are applied. Electrons are required to have $p_{\mathrm{T}} > 25$ GeV ($p_{\mathrm{T}} >$ 27.5 GeV) for the 2015 (2016) data, while muons are required to have $p_{\mathrm{T}} > 27.5$ GeV for the 2015 and 2016 data. Only one reconstructed lepton is required for the single-lepton channels, while exactly two leptons of opposite sign are required for the dilepton channels. All electrons are required to have *Tight* identification, while muons require *Medium* identification. Both are required to be isolated based on calorimeter and track based information to obtain high efficiencies.

At least four (two) jets are required in the e+jets and μ+jets (ee, $e\mu$ and $\mu\mu$) channels. At least one of these reconstructed jets should be tagged as a b-jet with a working point efficiency of 77%.

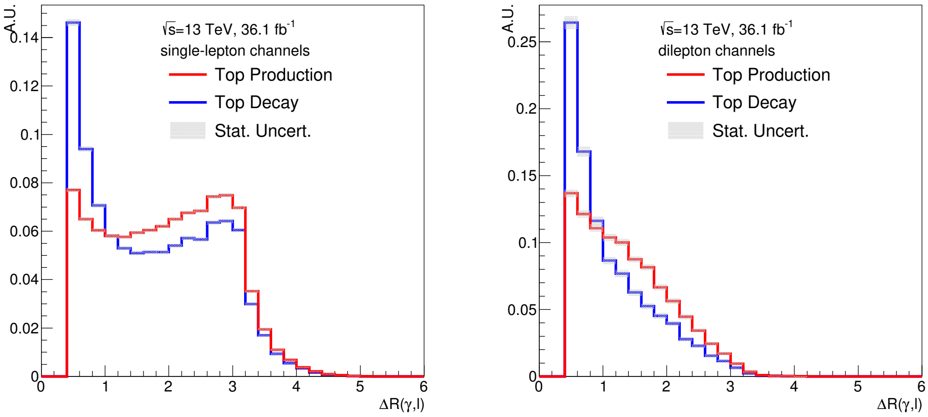
In all the dilepton channels a minimum invariant mass between the two leptons is such that $m(l,l) > 15$ GeV. This suppresses low mass Drell-Yan events where a quark and an anti-quark form a lepton pair through the propagation of a virtual photon. Low mass resonances such as the J/ψ ($c\bar{c}$) and the Υ ($b\bar{b}$) are also suppressed. In the ee and $\mu\mu$ channels an invariant mass veto of the two leptons is placed around the mass of the Z-boson, such that $m(l,l) \notin [85{,}95]$ GeV. Similarly, an invariant mass veto between the two leptons and the photon ($m(l,l,\gamma)$) is placed in the same way around the mass of the Z-boson for the same channels.

In the ee and $\mu\mu$ channels a cut on the $E_{\mathrm{T}}^{\mathrm{miss}}$ is used such that $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV.

In the e+jets channel a veto is placed on the invariant mass of the photon and the electron such that $m(\gamma, e) \notin [85{,}95]$ GeV. This suppresses $e \to \gamma$ fakes from Z-boson decays.

To enhance the $t\bar{t}\gamma$ signal region, cuts are made on photons. In all channels, exactly one photon of $p_{\mathrm{T}} > 20$ GeV is required. This photon needs to be *Tight* and isolated. A final cut is placed on the ΔR distance between the photon and the lepton, $\Delta R(\gamma, l) > 1.0$. This cut reduces the contribution of photons from radiative decay, as shown in Figure 1.6.

The $\Delta R(\gamma, l)$ distribution with all cuts mentioned above applied (except the final ΔR cut) is shown in Figure 5.2. Top production includes all types of processes shown in Figure 1.5, while top decay ideally contains events only from top quark decay products. There is some ambiguity with truth records due to the matrix element calculation and recording off-shell particles. Even after this cut there is still a considerable contribution of photons from radiative decay. The second peak at $\Delta R(\gamma, l) \approx \pi$ for the single-lepton channels occurs when the photon is radiated from the top quark or top quark decay products which lies opposite (in the η-ϕ plane) to the W-boson decaying into a lepton and neutrino. Future analyses can benefit from MVA techniques to further reduce this background contribution.



(a) single-lepton

(b) dilepton

Figure 5.2.: The ΔR distance between the lepton and the photon. A cut of $\Delta R > 1.0$ was chosen as the ideal value.

Validation plots for the region defined above but before photon selections can be seen in Appendix B.1, where good agreement is seen between data and MC. This is essentially a $t\bar{t}$ validation region.

A summary of all the cuts mentioned above is shown in Table 5.3. Table 5.4 shows the yields for each process before any data-driven scale factors have been applied. Only statistical uncertainties are included. In the e+jets (μ+jets) channel the $e \to \gamma$ fake contributes is the largest background and makes up around 14% (12%) of the signal + background MC. The $W\gamma$ background contribution is second largest with approximately 10% (11%). This is followed by the hadronic fake background of around 8% (9%). The biggest difference between the two channels is the fake lepton contribution which makes up approximately 5% (1%). For the ee ($\mu\mu$) channel the largest background contribution comes from the $Z\gamma$ background at around 12% (23%), with the other background processes essentially negligible. Because of the much smaller contribution of $Z\gamma$ in the $e\mu$ channel, this is grouped together as "Other prompt". For the $e\mu$ channel, all backgrounds combined make up about 6% of the signal+background contribution making it the purest channel.

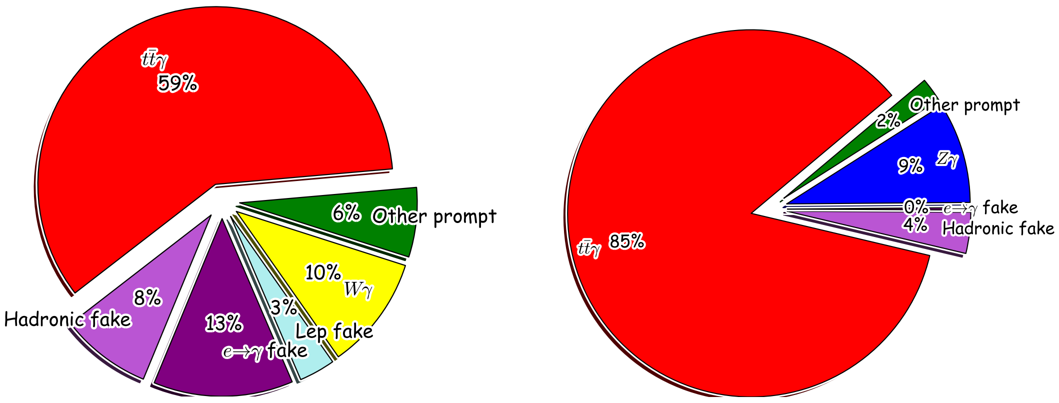
Thus, it is deemed that the composition of backgrounds for the e+jets and μ+jets channels are similar enough that these channels are often merged to form the single-lepton channel. Similarly, for the ee, $\mu\mu$ and $e\mu$ channels, these are combined to form the dilepton channel. This is often due to simplicity but also because the merged channels benefit from the increased statistics in many studies. Figure 5.3 summarises the processes that make up the combined single-lepton and dilepton channels by percentage, before data-driven scale factors have been applied. The different sources of the background processes as well as how scale factors are derived for various backgrounds will be discussed in upcoming sections.

Channel	e+jets	μ+jets	ee	$\mu\mu$	$e\mu$
Common	Primary vertex, Event cleaning, Run number, etc.				
	1 e	1 μ	2 e, OS	2 μ, OS	1 e + 1 μ, OS
Photon	1 γ with $p_{\mathrm{T}} > 20$ GeV				
Jets	≥ 4		≥ 2		
b-jets	≥ 1				
$m(\ell, \ell)$	-		$\notin [85, 95]$ GeV		-
	-		> 15 GeV		
$m(\ell, \ell, \gamma)$	-		$\notin [85, 95]$ GeV		-
$E_{\mathrm{T}}^{\mathrm{miss}}$	-		> 30 GeV		-
$m(\gamma, e)$	$\notin [85, 95]$ GeV	-			
$\Delta R(\gamma, \ell)$	> 1.0				

Table 5.3.: A summary of the event selections for each channel.

	e+jets	μ+jets	ee	$\mu\mu$	$e\mu$
$t\bar{t}\gamma$	3207 ± 14	3200 ± 13	145 ± 3.0	174 ± 3.1	401 ± 5.2
hadronic fake	441 ± 13	457 ± 14	7 ± 2	6 ± 2	18± 3
$e \to \gamma$ fake	774 ± 21	604 ± 16	0.5 ± 0.3	0.2 ± 0.5	1 ± 1
fake lepton	293 ± 35	63 ± 22	-	-	-
$W\gamma$	541 ± 29	579 ± 34	-	-	-
$Z\gamma$	-	-	21 ± 4	54 ± 14	-
Other prompt	411 ± 47	275 ± 15	5 ± 1	4 ± 1	8 ± 1
Total	5665 ± 71	5178 ± 50	179 ± 5	238 ± 14	428 ± 5.9
Data	6002	5660	196	233	473

Table 5.4.: Yields for signal and backgrounds for each channel. Only statistical uncertainties are included. Scale factors for the different backgrounds are not included.



(a) Merged single-lepton channels

(b) Merged dilepton channels

Figure 5.3.: Relative contributions of each process for the single-lepton and dilepton channels after merging. No data-driven scale factors have been applied.

5.3. $e \to \gamma$ fake background

An important background to the $t\bar{t}\gamma$ single-lepton channels arises when electrons are misidentified as photons. This makes up around 13% of total events in the signal region, and thus is the largest background contribution. This occurs mainly from the $t\bar{t}$ dileptonic decays, specifically from the $e\mu$ and ee channels. Another source is from $Z \to ee$ decays, hence the invariant mass window cut in the e+jets channel. For this reason the $e \to \gamma$ fake background is negligible in the dilepton channels. The approach to estimating the $e \to \gamma$ fake background follows a data-driven *tag-and-probe* method. The derived scale factors are used to correct the number of fake photons predicted by the MC samples.

Two control regions are needed to carry out this study. The first is a region enriched with fake photons called the $e\gamma$ control region. Here, one electron is trigger matched and at least one photon is required. The azimuthal angle between the electron and the leading p_{T} photon is required to be larger than 150 degrees implying that they are back-to-back. An invariant mass requirement between the electron and the photon is placed so that $m(l, \gamma) \in [40, 140]$ GeV. In this CR the electron is called the *tag electron* while the photon is called the *probe photon*. This CR implies that an electron has either been missed in the reconstruction, or that the photon is fake. The second control region, called the ee CR has the same definitions as the $e\gamma$ CR. The only key difference is that the requirement on the photon is replaced with a requirement to be an electron of opposite charge to the initial tag electron. This second electron now acts as the probe with the implication that this CR only contains real electrons.

Truth studies can be performed on the probe photon in the $e\gamma$ CR to determine the source of the $e \to \gamma$ fakes. The study is done by looking at the seed cluster in the EM calorimeter and calculating the angular distance to the truth particle before simulation. In addition the probe photon's kinematics can also be compared to those of the probe electron (not shown in this thesis). From these studies, four cases have been identified:

- Approximately 60% of the photons are simply mis-reconstructed as an electron.
- Approximately 25% of the photons are mis-matched to a true photon. The reconstructed photon's p_{T} is larger than the truth photon's p_{T} by more than 10%. In addition there is a truth electron which satisfies $\Delta R < 0.05$ to the seed cluster.
- Approximately 2% of the photons are truth matched to a reconstructed photon with no electrons in the nearby vicinity. This indicates that these are actual prompt photons

possibly from the decay $Z \to ee + \gamma$. These photons are not counted as fake.

- Approximately 3% of photons are matched to a true photon, have a p_{T} difference of less than 10%, and true electron is within $\Delta R < 0.05$ of the photon. This indicates a real non-prompt photon.

We can infer a fake rate from MC, which is defined as

$$\mathrm{FR}^{\text{e-fake}}_{\mathrm{MC}} = \frac{N_{e,\gamma}}{N_{e,e}}, \tag{5.8}$$

where $N_{e,\gamma}$ ($N_{e,e}$) is the number of events observed in the $Z \to e\gamma$ ($Z \to ee$) CR. The MC sub index means that this estimate is derived using truth information and the number of events described by the MC. To define the data-driven fake rate, in both CRs the backgrounds that do not contribute to the Z-boson peak are subtracted using a sideband fit of the $m(l, \gamma)$ distribution. Equation 5.8 can be adjusted so that

$$\mathrm{FR}^{\text{e-fake}}_{\text{d.d.}} = \frac{N^{\text{data}}_{e,\gamma} - N^{\text{non-Z}}_{e,\gamma}}{N^{\text{data}}_{e,e} - N^{\text{non-Z}}_{e,e}}, \tag{5.9}$$

where $N^{\text{non-Z}}_{e,\gamma/e}$ indicates the tails of the Z-boson peak in the respective CRs.

Finally, the ratio of the data-driven fake rate to the MC fake rates gives scale factors that can be applied as corrections to the $e \to \gamma$ fake MC samples,

$$\mathrm{SF}^{\text{e-fake}}_{\mathrm{FR}} = \frac{\mathrm{FR}^{\text{e-fake}}_{\text{d.d.}}}{\mathrm{FR}^{\text{e-fake}}_{\mathrm{MC}}}. \tag{5.10}$$

To give a sense of the size, an overall SF is derived to be 0.97 ± 0.02(stat). However, in practice 2D fake rates and therefore SFs are derived for bins of p_{T} and η.

A number of variations to the above procedure are considered to account for systematic uncertainties. For the data-driven fake rate calculation the signal (background) modelling is varied from a Crystal-ball[2] (Bernstein 4th order polynomial) function to the MC predicted template (a Gaussian function). The MC model considered to subtract the prompt photon contribution from the $Z \to ee$ sample is switched to a dedicated $Z \to ee\gamma$ sample. Finally, the range of the fit in the low and high ends of the tail are shrunk between 5 and 10%. Each of these contributions is summed in quadrature. To give a sense of the size, the overall relative systematic uncertainty on the $e \to \gamma$ fake scale factor is 22%. As with the scale factors, in practice, the systematic uncertainties are calculated in p_{T} and η bins. Further theoretical uncertainties taken into consideration are explained in Chapter 7.

The final component of the $e \to \gamma$ fake background is a closure test showing the extrapolation from the two CRs used above to essentially the signal region defined in Section 5.2. The difference being that the required photon is replaced by an electron. The e+jets channel is now dominated by $t\bar{t}$ and Z+jets events. A ratio of data over prediction is calculated, which yields 0.98 ± 0.01(stat.). This is applied on top of the 3D scale factors in the single-lepton channels. For the dilepton channels, due to lack of $e \to \gamma$ fake events, an extra 50% normalisation uncertainty is assumed instead.

5.4. Hadronic fake background

The hadronic fake background plays an important role in many analyses when photons are required. It is for this reason that the Prompt Photon Tagger (Chapter 6.1) was developed.

---

[2] A convolution of a Gaussian function as the main contribution and below some threshold in the low end part of the tail, a power law function.

The majority of hadronic fake photons in the single-lepton channels comes from $t\bar{t}$ events where a final state jet radiates a photon. Small contributions also arise from W+jets and single top processes. In the dilepton channels there are small contributions from the Z+jets events. Before requiring an isolated photon this background is the dominant process in the single-lepton channels. The new strategy for this analysis (compared to the 8 TeV result) significantly reduces the contribution of hadronic fake photons by requiring isolated photons. In the single-lepton (dilepton) channel it contributes around 8% (4%) to the overall number of events. This is a non-negligible contribution (specifically in the single-lepton channel) and so correct modelling needs to be ensured. A more accurate estimation of this background that does not rely on MC modelling follows a data driven ABCD method. Scale factors to match MC prediction to data are derived in control regions where the isolation and identification of the photon have been reversed, thus enriching the hadronic fake contribution.

The ABCD method relies on having four orthogonal and non-correlated regions. Using the same event selections defined for the single-lepton channels in Section 5.2, three new regions are defined by tweaking or inverting a photon related requirement. Region A requires that the photon fail at least two out of four shower shape variables used for tight photon identification ($\omega_{\mathrm{s3}}, F_{\mathrm{side}}, \Delta E, E_{\mathrm{ratio}}$)[3]. Region B requires photons to fail the same ID as A and additionally fail the isolation requirement. Region C requires that photons fail the same isolation requirement as B, but pass the tight ID. Region D is the signal region. An extra cut of $p_{\mathrm{T}}^{\mathrm{cone20}} > 3$ GeV is also required in region B and C to reduce the prompt photon contamination. This is illustrated by the schematic in Figure 5.4.

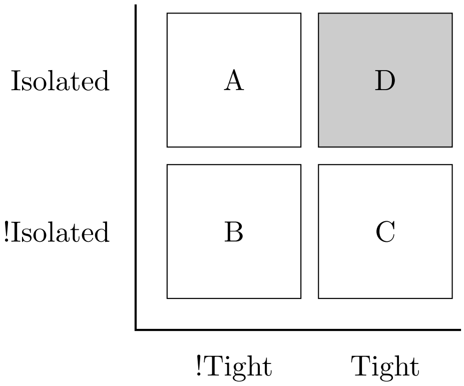


Figure 5.4.: The control regions with the orthogonal photon definitions defined for the hadronic fake ABCD method. Region D is the signal region. An exclamation mark indicates the reverse of the cut.

If the photon isolation and ID are uncorrelated, then the assumption holds that the ratio of tight vs non-tight photons should not change depending on if the photons are isolated or non-isolated. Similarly, the ratio of isolated vs non-isolated photons should not change if the photon is tight or non-tight. Thus,

$$\frac{N_{\mathrm{D}}^{\text{h-fake}}}{N_{\mathrm{C}}^{\text{h-fake}}} = \frac{N_{\mathrm{A}}^{\text{h-fake}}}{N_{\mathrm{B}}^{\text{h-fake}}} \tag{5.11}$$

and

$$\frac{N_{\mathrm{D}}^{\text{h-fake}}}{N_{\mathrm{A}}^{\text{h-fake}}} = \frac{N_{\mathrm{C}}^{\text{h-fake}}}{N_{\mathrm{B}}^{\text{h-fake}}}. \tag{5.12}$$

---

[3] The choice of failing two shower shape variables is based on studies that maximise statistics while keeping prompt photon contamination low.

Apart from the two variables needing to be uncorrelated, the ABCD method requires that enough statistics are in each region (which is why this method can not be used in the dilepton channels), and that there is only a small signal contamination in region A, B and C. However, in practice photon isolation and identification are not uncorrelated, and so this needs to be taken into account. A double ratio, θ_{MC} is defined as

$$\theta_{\mathrm{MC}} = \frac{N^{\text{h-fake}}_{\text{D, MC}} \Big/ N^{\text{h-fake}}_{\text{C, MC}}}{N^{\text{h-fake}}_{\text{A, MC}} \Big/ N^{\text{h-fake}}_{\text{B, MC}}}. \tag{5.13}$$

This quantifies the correlation between isolation and identification. Assuming 0 correlation the term will be 1, which often is not the case. The MC sub index indicates that these are the predicted number of events for the hadronic fake contribution according to MC[4].

From Equation 5.12 and Equation 5.13, we can define the estimated hadronic fake contribution in the signal region:

$$N^{\text{h-fake}}_{\text{D, est.}} = \frac{N^{\text{h-fake}}_{\text{A, data}} \cdot N^{\text{h-fake}}_{\text{C, data}}}{N^{\text{h-fake}}_{\text{B, data}}} \times \frac{N^{\text{h-fake}}_{\text{D, MC}} \Big/ N^{\text{h-fake}}_{\text{C, MC}}}{N^{\text{h-fake}}_{\text{A, MC}} \Big/ N^{\text{h-fake}}_{\text{B, MC}}}, \tag{5.14}$$

where $N^{\text{h-fake}}_{\text{A/B/C, data}}$ indicates the number of hadronic fake events estimated by subtracting the sum of all background events (including the hadronic fake MC) and signal events from the total data events in the respective region. Finally, the hadronic fake scale factor is defined as

$$\mathrm{SF}^{\text{h-fake}} = \frac{N^{\text{h-fake}}_{\text{D, est.}}}{N^{\text{h-fake}}_{\text{D, MC}}}. \tag{5.15}$$

The scale factors are extracted in bins of p_{T}, η and converted/unconverted photon type. They range from 0.8 to 3.2 and are applied to the hadronic fake MC samples in both the single-lepton and dilepton channels. Several sources of systematic uncertainties are considered as a result of the ABCD method. The statistical uncertainty of the number of hadronic fake events estimated from data as well as from the θ_{MC} factor are considered. Uncertainties due to $t\bar{t}\gamma$ and the background subtraction is also included, where the $t\bar{t}\gamma$ signal is varied by 100%, the other prompt sources and the $e \to \gamma$ fake background are varied by 50%, and the fake lepton background is varied accordingly as explained in Section 5.5. Further theoretical background modelling uncertainties taken into consideration are explained in Chapter 7.

5.5. Non-prompt and fake lepton background

Leptons from W-boson decays categorise the channels in which this analysis is performed. These are called prompt leptons. However, non-prompt and fake leptons may satisfy the event selection criteria. From hereon, both of these types of events are categorised as "fake lepton" background and are predominantly in the single-lepton channels. The main contribution comes from the QCD driven multi-jet processes in association with a photon. Non-prompt or fake electrons and muons can arise from the semi-leptonic decay of b and c quarks. For electrons there can be additional contributions from photon conversions and jets in the electromagnetic calorimeter. Muons can be non-prompt or faked from energetic showers in the hadronic calorimeter, or from energetic hadrons that punch-through to the hadronic calorimeter. In the dilepton channels a fake and a real lepton are required, or two fake leptons. This contribution is negligible and so the fake lepton background for the dilepton channels is neglected. The estimation of the fake lepton background in the single-lepton channels follows the fully data-driven Matrix Method [159] approach explained in the following.

---

[4] A data-driven approach to the double ratio term can be done but it requires a more complicated region definition since $N^{\text{h-fake}}_{\text{D, data}}$ is the quantity sought after.

An additional loose sample is created that uses looser lepton identification and isolation requirements than those outlined in Section 5.2. For the loose sample the electrons are required to have *Medium* identification with no isolation requirement, while muons keep the same identification (*Medium*) but the requirement on isolation is dropped. At this point, no requirement on a photon is made on either the tight or the loose samples.

An assumption can be made that the tight sample will contain mostly real leptons, while the loose sample is enriched with events that contain fake leptons. Thus, we can write

$$N^{\text{loose}} = N^{\text{loose}}_{\text{real}} + N^{\text{loose}}_{\text{fake}} \tag{5.16}$$

$$N^{\text{tight}} = N^{\text{tight}}_{\text{real}} + N^{\text{tight}}_{\text{fake}}, \tag{5.17}$$

which says that the total number of loose (tight) leptons is a linear combination of the real and fake leptons in the loose (tight) sample. We define ϵ_{real} as the probability of a real lepton in the loose sample to pass the tighter selection. Similarly, ϵ_{fake} is the probability of a fake lepton in the loose sample to pass the tighter selection. The total leptons in the tight sample can be written as

$$N^{\text{tight}} = \epsilon_{\text{real}} N^{\text{loose}}_{\text{real}} + \epsilon_{\text{fake}} N^{\text{loose}}_{\text{fake}}. \tag{5.18}$$

If the real and fake efficiencies are known, then from the equations above, the number of events with fake leptons in the tight sample can be estimated using

$$N^{\text{tight}}_{\text{fake}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} (\epsilon_{\text{real}} N^{\text{loose}} - N^{\text{tight}}).$$

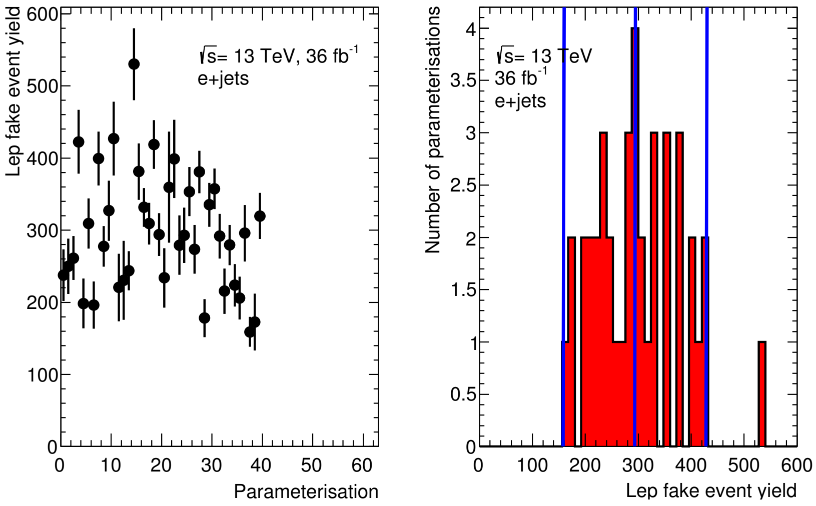
The real efficiencies are estimated using the tag-and-probe method in the $Z \to ee$ and $Z \to \mu\mu$ control regions. In the region of 80 to 100 GeV a pure sample is obtainable. The fake efficiencies are estimated in data samples dominated by non-prompt and fake leptons. Among other requirements, one *Loose* lepton, at least one jet and low $E_{\text{T}}^{\text{miss}}$ are required. These measurements can be done with respect to a number of variables that depend on the lepton or event kinematics, for example the lepton p_{T} and η, number of b-jets, and jet p_{T}. Thus, they are parametrised as a function of these variables. The efficiency measurements are provided centrally in ATLAS and are explained in more detail in [160].

Individual event weights are calculated as functions of the efficiencies and take the form

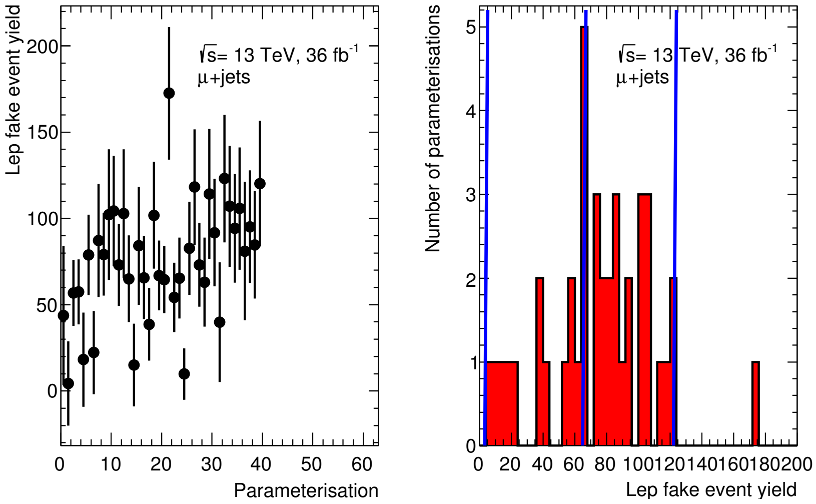
$$w_i = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} (\epsilon_{\text{real}} - \delta_i), \tag{5.19}$$

where δ_i is 1 if the event i passes the tight selection, otherwise 0. These event weights are applied to the loose data sample. Lastly, by adding the photon related cuts a fake lepton estimation in the full signal region is obtained.

Figure 5.5 shows the spread of different event yields for the e+jets and μ+jets channels. The left plots show the event yield as a function of each parameterised set of weights, while the right plots show the projection of the y-axis. The closest parametrisation to the mean of the new histogram is taken as the nominal weight. The chosen up and down systematic variations are chosen as the extreme edges of the histogram (neglecting the obvious outliers).



(a) e+jets



(b) μ+jets

Figure 5.5.: The fake lepton yields for each parametrisation with the y-axis projection displayed on the right. The blue lines indicate the mean (the nominal weight) and the up/down systematic variations which form the extremes of the histogram.

Both the mean and up/down systematic variations are indicated by the blue vertical lines. Further parameterisations are neglected due to being outliers and are not included in the plot. For the e+jets channel these are p_{T} related, while for μ+jets these are $\Delta\phi$ related.

5.6. Prompt photon backgrounds

Events in which a real prompt photon is radiated from anything except top quarks (or incoming quarks) make up about 16% (11%) of total events in the combined single-lepton (dilepton) channels. For the single-lepton channel this is largely due to the $W\gamma$ process, whereas for the dilepton channel the main contribution is from the $Z\gamma$ process. Other processes such as single-top, diboson and $t\bar{t}V$, where each have an additional prompt photon in the final state have small or even negligible contributions. Table 5.5 shows the prompt background composition for each of the five channels after event selections discussed in Section 5.2 have been applied.

In the single-lepton (dilepton) channel the $W\gamma$ ($Z\gamma$) contribution is largest and is therefore separated from the rest of the prompt photon backgrounds. These are then grouped together as “Other prompt”.

A validation region (VR) is defined for the $Z\gamma$ and $W\gamma$ backgrounds primarily to ensure that the modelling of these processes is adequate. A maximum likelihood fit to the jet p_{T} distribution is performed, of which the mechanics of the fit are explained in Chapter 5.1[5]. All systematics explained in Chapter 7 have been included. The fake lepton data-driven background estimate from the previous section as well as all previously derived scale factors for various backgrounds are also applied. While these should be derived specifically for the region in question, given the function of these VRs this approximation is sufficient. It is important to emphasise that these are validation regions and so none of the results in this section actually enter the final result. They are purely used as a cross-check and closure test.

Process	$e\mu$	$\mu\mu$	ee	e+jets	μ+jets
$W\gamma$	-	-	-	541 ± 29	579 ± 34
$Z\gamma$	-	54 ± 14	21 ± 4	295 ± 46	152 ± 13
Single-top	6 ± 1	3 ± 1	34 ± 1	85 ± 5	92 ± 6
Diboson	-	-	-	4 ± 1	6 ± 1
$t\bar{t}V$	3.0 ± 0.2	1.5 ± 0.1	1.3 ± 0.1	27 ± 1	25 ± 1

Table 5.5.: The yields of prompt photon background events from MC. The numbers are normalised to the total integrated luminosity. Only statistical uncertainties are included. A dash indicates the yield is < 1 and therefore negligible.

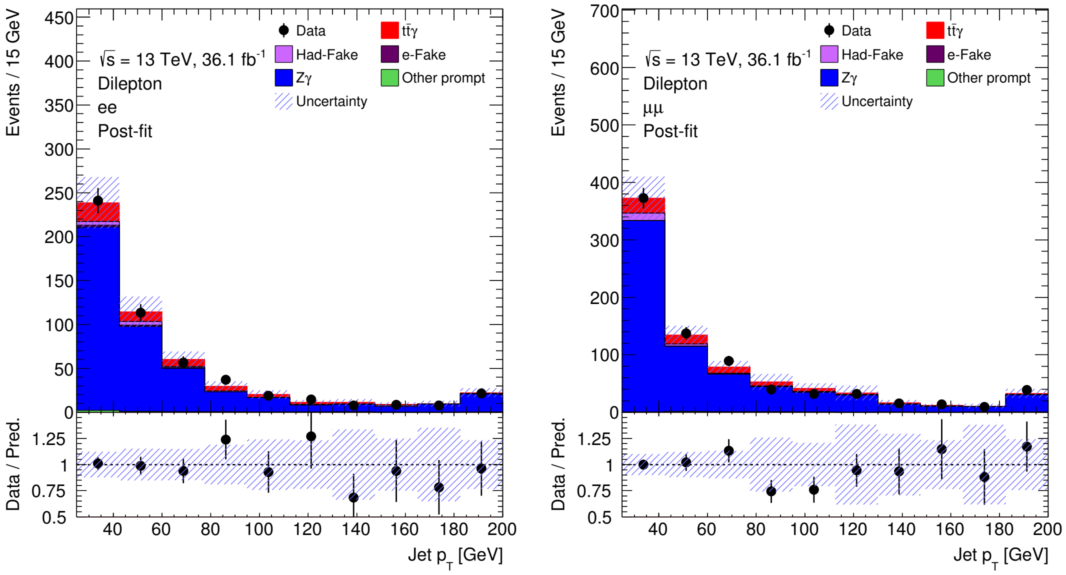
$Z\gamma$ Validation region

The $Z\gamma$ VR is comprised of the dilepton channels with ee and $\mu\mu$ final states. They each have similar object definitions and event selections to the signal region with the exception of the following: number of jets $\geq$ 0, number of b-tagged jets $= 1$ and finally the invariant mass of the two leptons is required to be in a mass window of [60, 100] GeV. With the requirement of b-tagged jets $= 1$, we enter the heavy flavour topology, similar to our signal. A second $Z\gamma$ VR is defined with the same cuts above, except there is a requirement to have 0 b-tagged jets, i.e., the light flavour topology.

The post-fit histograms for the jet p_{T} can be seen in Figure 5.6. The results from the fits are shown in Table 5.6. The table shows that there is a 3% and 10% overestimation of the $Z\gamma$ MC in the heavy flavour VR for the ee and $\mu\mu$ channels, respectively. The small contribution of $t\bar{t}\gamma$ could also contribute. However, this result is well within the given uncertainties. Similarly, the light flavour VR overestimates the MC by 19% and 22% for the ee and $\mu\mu$ channels, respectively. There is a small contribution from hadronic fake backgrounds which are modelled by MC and so could contribute to this overestimation. However, these results are within the total uncertainties.

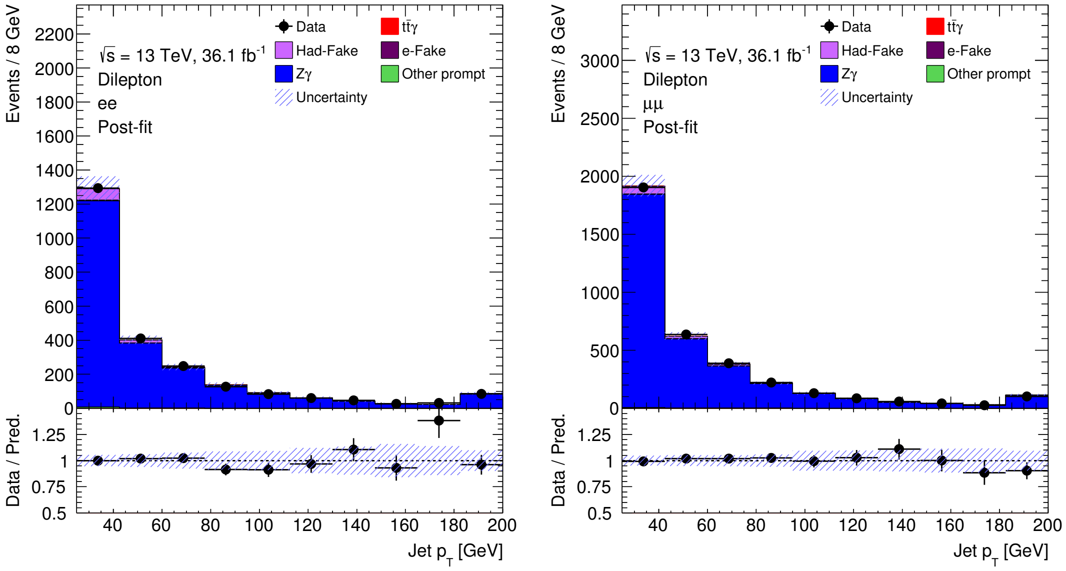
---

[5] The maximum likelihood fit to the signal region was repurposed for this validation region, and so will only be presented later. Essentially, a value of $\mu = 1$ indicates the theoretical SM prediction.



(a) Heavy flavour ee channel

(b) Heavy flavour $\mu\mu$ channel



(c) Light flavour ee channel

(d) Light flavour $\mu\mu$ channel

Figure 5.6.: Heavy flavour (top) and light flavour (bottom) post-fit validation regions for the dilepton channel. All systematic uncertainties are included.

channel	ee		$\mu\mu$	
	μ	$\pm$	μ	$\pm$
b-jets=0	1.19	$^{+0.26}_{-0.22}$	1.22	$^{+0.24}_{-0.20}$
b-jets=1	1.03	$^{+0.35}_{-0.24}$	1.10	$^{+0.29}_{-0.23}$

Table 5.6.: Fit results for the $Z\gamma$ validation region. Statistical and systematic uncertainties are included.

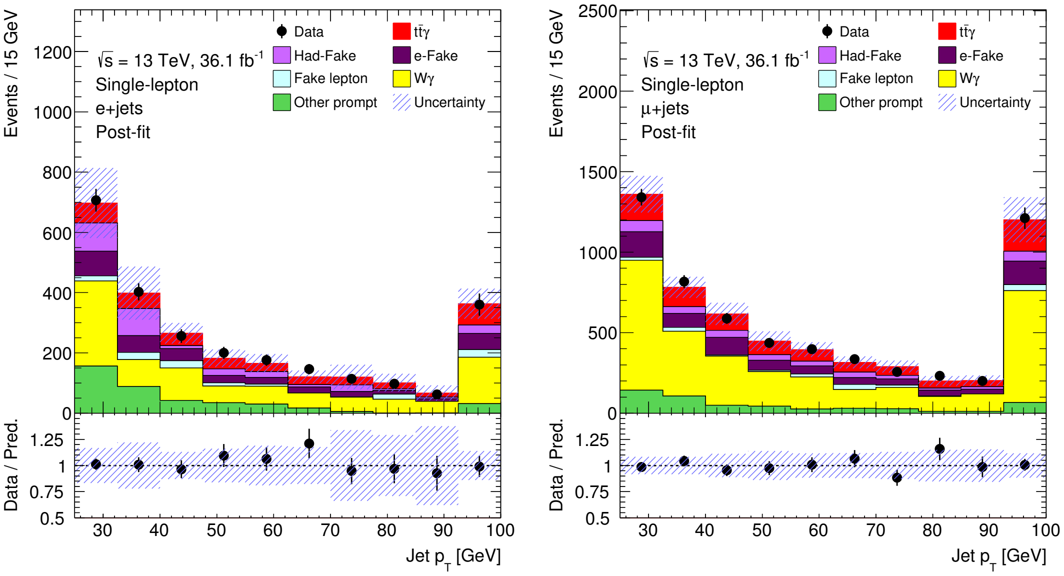
$W\gamma$ Validation region

The $W\gamma$ VR is comprised of the single-lepton channels with e+jets and μ+jets final states. They each have similar object definitions and event selections to the signal region with the exception of the following: $2 \leq$ number of jets ≤ 3, number of b-tagged jets $= 1$ and $E_{\mathrm{T}}^{\mathrm{miss}} >$ 40 GeV. In addition, there is a cut on the ELD distribution (Chapter 6.2) of < 0.04 to further reduce signal contamination. For the e+jets channel an invariant mass cut between the photon and the lepton of $m(\gamma, lep) < 80$ GeV helps reduce the $e \to \gamma$ fake contribution. With the requirement of b-tagged jets $= 1$, we enter the heavy flavour regime, similar to our signal. A second $W\gamma$ VR is defined with different cuts to the above. In addition to requiring 0 b-tagged jets (i.e., the light flavour regime) we ask for $1 \leq$ number of jets ≤ 3 and $\Delta R(\gamma, lep) < 2.8$. For the e+jets channel we exclude events where the mass between the photon and the lepton falls in a window of [60,100] GeV.

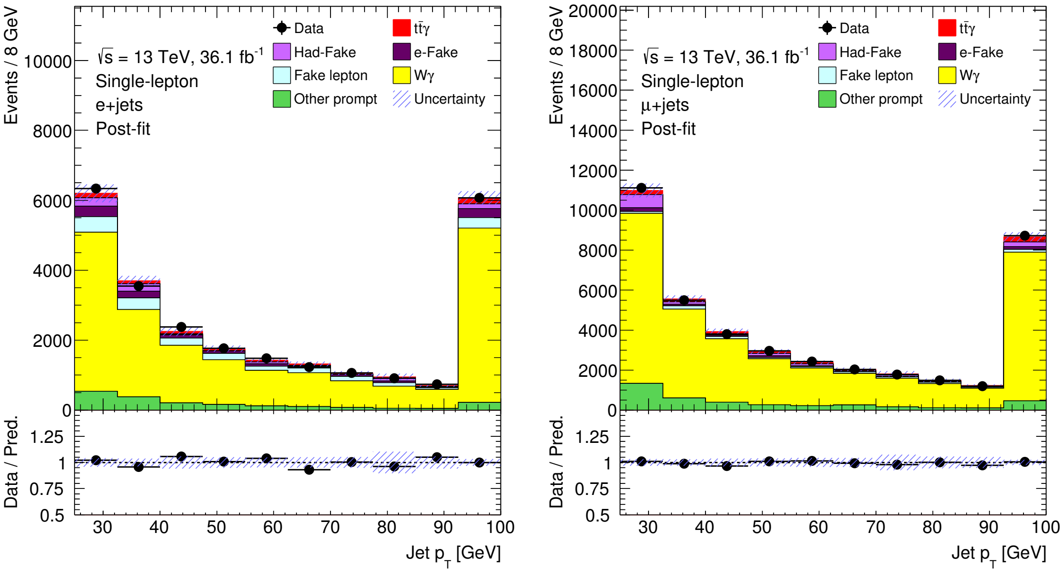
The post-fit histograms for the jet p_{T} can be seen in Figure 5.7. The results from the fits are shown in Table 5.7. The fit results show that the light flavour VR overestimates the MC contribution between 15% and 18%, of which the uncertainties do not fully cover. Similarly for the heavy flavour μ+jets VR, a 26% over estimation is seen of which uncertainties do not cover. Given that none of these regions are completely pure with the $W\gamma$ process, and no dedicated studies were performed on the backgrounds in this phase space, nothing conclusive can be said. For this reason, the $W\gamma$ background has no constraint on the normalisation in the final fit for single-lepton channels; it is a free parameter.

channel	e+jets		μ+jets	
b-jets=0	1.15	$^{+0.05}_{-0.05}$	1.18	$^{+0.06}_{-0.06}$
b-jets=1	1.14	$^{+0.33}_{-0.32}$	1.26	$^{+0.13}_{-0.14}$

Table 5.7.: Fit results for the $W\gamma$ validation region. Statistical and systematics uncertainties are included.



(a) Heavy flavour e+jets channel (b) Heavy flavour μ+jets channel



(c) Light flavour e+jets channel (d) Light flavour μ+jets channel

Figure 5.7.: Heavy flavour (top) and light flavour (bottom) post-fit validation regions for the single-lepton channel. All systematic uncertainties are included.

5.7. Pre-fit distributions

This section shows a selection of distributions in the respective SRs comparing data to predicted MC. All appropriate scale factors and scalings from the previous sections have been applied. For brevity, only the single-lepton and dilepton combined channels are shown in Figure 5.8 and 5.9, respectively. All systematic uncertainties (Chapter 7) have been included and are represented by the blue hashed band (which includes the statistical uncertainty). The sources of these uncertainties will be further discussed in Section 7. "Pre-fit" on the distributions indicates this is before any fitting has been performed. Table 5.8 shows the yields for the signal, backgrounds, and data.

The variables associated with "b-tagging score" relate to the weights presented in Table 4.3 for all jets. The transverse mass of the W-boson ($m_{\mathrm{T}}(W)$) is calculated as

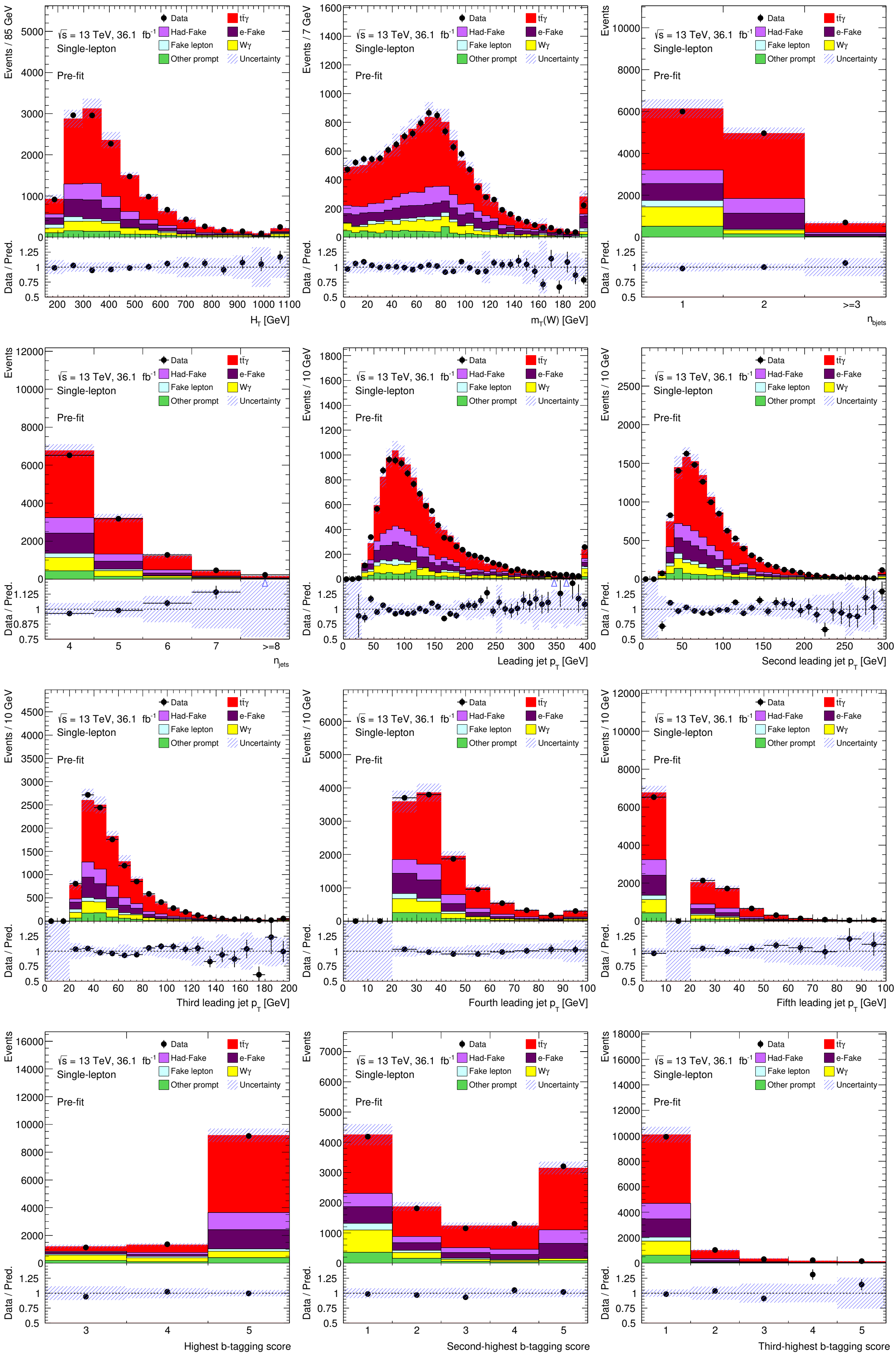
$$m_{\mathrm{T}}(W) = \sqrt{2 \times p_{\mathrm{T}}(l) \times E_{\mathrm{T}}^{\mathrm{miss}} \times (1 - \cos(\Delta\phi(l, E_{\mathrm{T}}^{\mathrm{miss}})))}. \tag{5.20}$$

The variable H_{T} is the scalar sum of the transverse momentum for final state jets and leptons. In the case where a jet does not exist (such as the fifth leading jet p_{T}), the variable is padded with zeros.

The selected plots shown are also the inputs for the Event-level Discriminator, which will be explained in Chapter 6.2. The same phase space is shown for the training of the single-lepton NN, while a slightly tighter phase space is shown compared to the training of the dilepton NN, as this represents the full signal region. In all figures, good data/MC can be seen with the projected uncertainties covering any deviations that might arise.

	single-lepton	dilepton
$t\bar{t}\gamma$	6488 ± 420	720 ± 34
hadronic fake	1435 ± 290	49 ± 27
$e \to \gamma$ fake	1653 ± 170	2 ± 1
fake lepton	356 ± 200	-
$W\gamma$	1126 ± 45	-
$Z\gamma$	-	75 ± 52
Other prompt	691 ± 260	18 ± 7
Total	11747 ± 710	863 ± 78
Data	11662	902

Table 5.8.: Pre-fit yields for signal and backgrounds for the single-lepton and dilepton channels. All scale factors and systematic uncertainties are included.


√s = 13 TeV, 36.1 fb⁻¹
Single-lepton
Pre-fit
Data
tt̄γ
Had-Fake
e-Fake
Fake lepton
Wγ
Other prompt
Uncertainty
Events / 85 GeV
Events / 7 GeV
Events
Events / 10 GeV
Data / Pred.
H_T [GeV]
m_T(W) [GeV]
n_bjets
n_jets
Leading jet p_T [GeV]
Second leading jet p_T [GeV]
Third leading jet p_T [GeV]
Fourth leading jet p_T [GeV]
Fifth leading jet p_T [GeV]
Highest b-tagging score
Second-highest b-tagging score
Third-highest b-tagging score
>=3
>=8

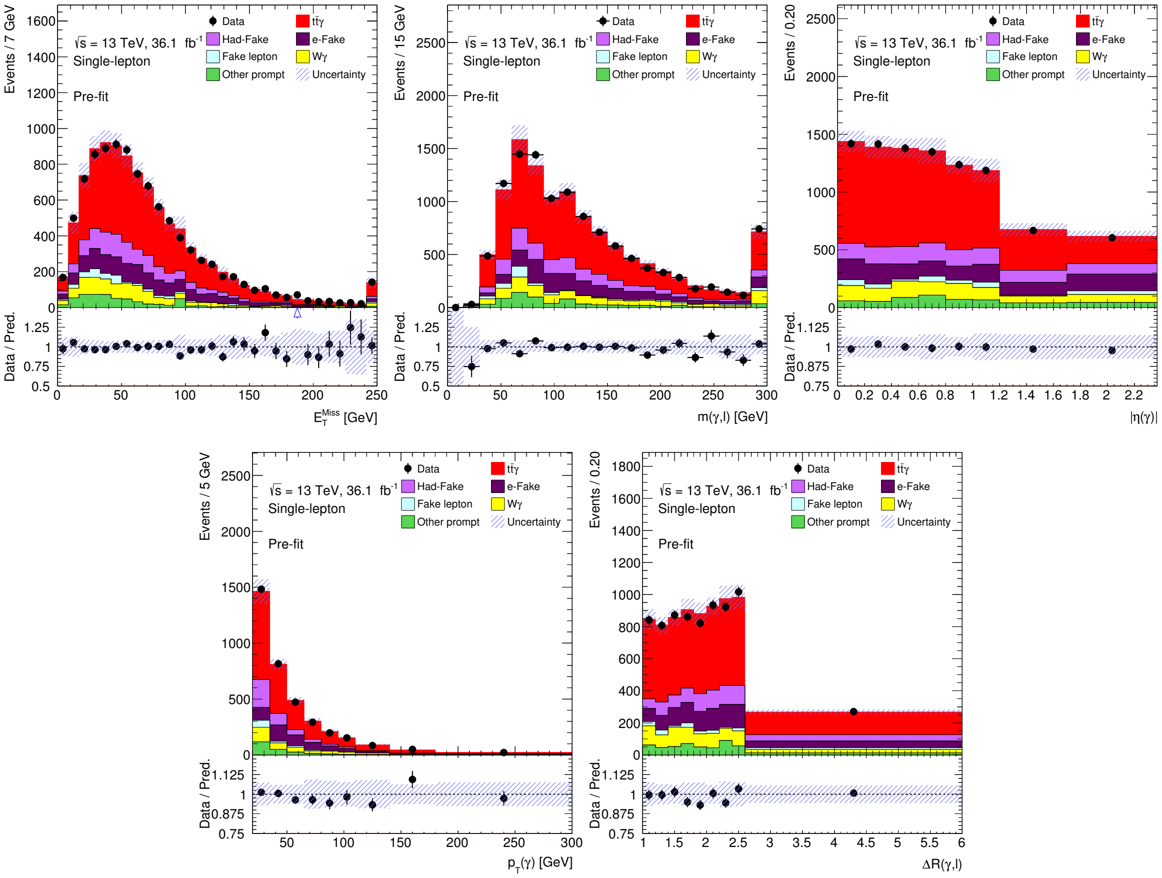


Figure 5.8.: A selection of pre-fit distributions for the single-lepton channel. All scale factors and systematic uncertainties are included.

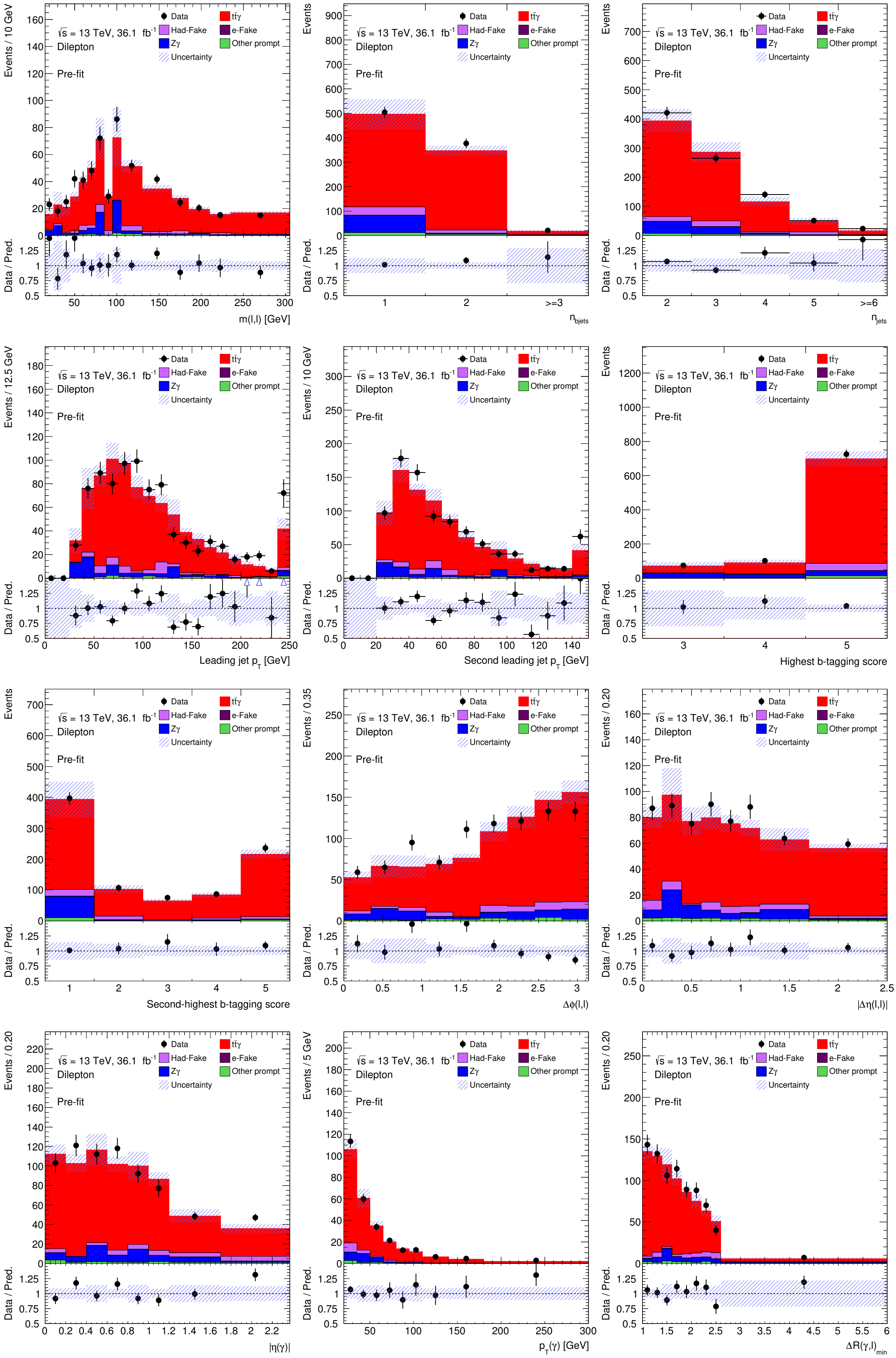


Figure 5.9.: A selection of pre-fit distributions for the dilepton channel. All scale factors and systematic uncertainties are included.

CHAPTER 6

Object- and event-level neural networks

An event has an intrinsic topology; the number of jets or leptons, the angular distance between different particles, the $E_{\mathrm{T}}^{\mathrm{miss}}$, all form the overall defining characteristics of an event for a certain process. The constituents of an event, the photons, electrons, individual jets etc., have kinematic properties and, in the case of jets, flavours. The topology of an event as well as the properties of its constituents can be exploited for better signal/background separation. NNs are used for this purpose.

This chapter presents two separate neural networks (NNs) developed in the period of this thesis. The first (Section 6.1) is an analysis independent, object-level NN designed to work on top of the photon reconstruction and identification presented in Chapter 4.1.1. Labelled the Prompt Photon Tagger (PPT), this NN was developed as a Göttingen effort with contributions largely from Benedikt Völkel during his Master's thesis [161]. The application and evaluation to $t\bar{t}\gamma$ events, as well as related studies include significant contributions from myself and Knut Zoch, who focused on crucial systematic uncertainties and scale factor studies.

The second NN developed and presented is the Event-level Discriminator (ELD) for the $t\bar{t}\gamma$ analysis (Section 6.2). The ELD makes use of the PPT as an input variable in the single-lepton channel, as will be explained later.

6.1. The Prompt Photon Tagger

In the previous measurements at 7 and 8 TeV, the dominant background process was from hadronic fakes. Thus, a multivariate analysis (MVA) tool in the form of a NN was developed to try mitigate the role this background would play in the final fit. Furthermore, since this background is common among many analyses within ATLAS, the goal was to keep this tool analysis independent. To accomplish this the PPT was trained purely on shower shape variables (Chapter 4.1) using the same prompt photon and hadronic fake enriched samples used in photon identification. Section 6.1.1 gives an overview of the procedure to train the PPT and Section 6.1.2 evaluates the PPT's performance on $t\bar{t}\gamma$ samples. Section 6.1.3 gives an overview of how scale factors and systematic uncertainties are derived for this novel tool.

6.1.1. Training

Training was carried out on two MC datasets provided by the photon identification group within ATLAS. The first, simulating the QCD-Compton process, contains prompt photons.

The second MC dataset contains di-jets events which contains hadronic fake candidates. Photon candidates are required to have a $p_{\mathrm{T}} > 25$ GeV as well as $|\eta| < 2.5$. The crack region ($1.37 < |\eta| < 1.52$) is also excluded. Photons are required to pass *Tight* identification[1]. Approximately 1 million prompt photons and 200k hadronic fakes are selected from the MC datasets with 80% of them used for training and 20% used for testing purposes.

A subset of shower shape variables was selected based on their separation information from Table 4.1. This is defined as

$$S = \frac{1}{2}\sum_{i}^{\text{bins}} \frac{(s_i - b_i)^2}{s_i + b_i}, \tag{6.1}$$

where for each bin, i, s_i and b_i indicate the number of signal and background events, respectively. Once ranked by separation, the top six variables were chosen for the PPT. These are detailed in Table 6.1 with their definitions explained in Chapter 4.1.1.

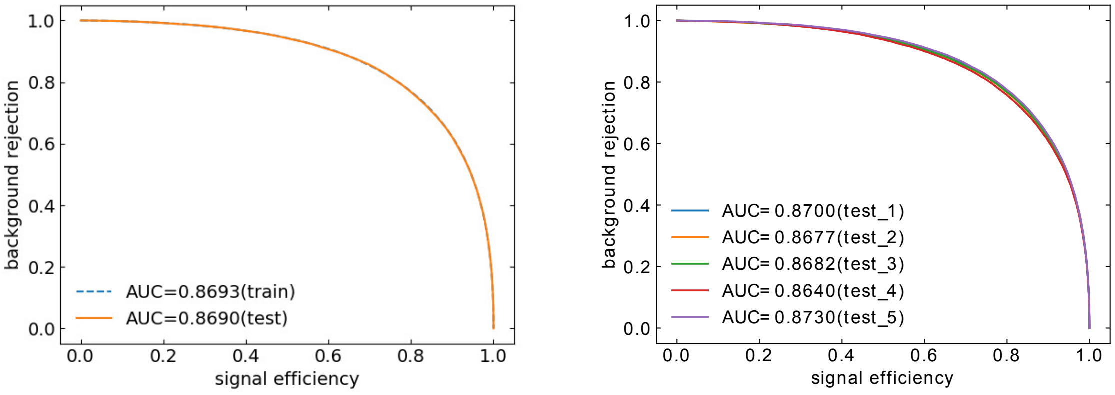
variable	separation
f_{side}	7.21 %
R_ϕ	7.01 %
R_η	4.83 %
$w_{\eta 1}$	4.14 %
R_{had}	3.33 %
$w_{\eta 2}$	2.01 %

Table 6.1.: The six variables chosen for the PPT based on separation calculated according to Equation 6.1.

Grid searches were performed to determine the best NN architecture. The chosen PPT consists of five hidden layers. The first, third and fifth layer contain 64, 40 and 52 neurons, respectively, of which the first uses a ReLu activation function while the other two use softmax activation functions. The second and fourth layers are batch normalisation layers. The final output layer comprises one single neuron with a sigmoid activation function making it a binary classifier. In total, there are 5441 trainable parameters. The performance of this network is shown in a Receiver Operating Characteristics (ROC) curve in Figure 6.1a. This shows the background rejection versus the signal acceptance for a given value of the PPT output. Crucially, the testing and training samples have very similar curves, meaning that naively it looks as if the network can generalise to unseen data. Further stability of the neural network architecture and model can be examined using k-fold cross validation. In this case, 5-fold validation was carried out where the datasets were split into five equal samples. Four are used to train while the fifth is used as test set. Each sample has a chance to be the test set and so five ROC curves are drawn. These are shown in Figure 6.1b for the test samples, where the agreement between each curve can be seen. Thus, the network generalises well to unseen data and is not overtrained. Further details on the training of the PPT can be seen in [161].

---

[1] In future versions of the PPT the *Tight* requirement will be dropped to cover a larger phase space and cater for a wider range of analyses.



(a) Training and test sets

(b) 5-fold cross validation

Figure 6.1.: ROC curves of training and test sets as well as those obtained in 5-fold cross validation for the chosen PPT architecture.

6.1.2. Evaluation

The PPT was one of the first neural networks in ATLAS to be trained with modern machine learning libraries (Keras) and be injected into analysis code using LWTNN. The implications of this are discussed in more detail in Section 1.2.

We can gauge the performance of the neural network on the $t\bar{t}\gamma$ MC and data samples (before event-level cuts). Figure 6.2 shows the categorisation for three types of photons; those classified as $e \to \gamma$ fakes (photon type = 20), those as hadronic fakes (photon type = 10) and those as prompt photons from various mother particles (photon type < 10). Prompt photons below 10 are categorised as such since there are ambiguities in the truth records and the exact origin of photons. Each row has been normalised separately. The first figure shows the behaviour of the PPT for photons in which there is no requirement on the isolation. For prompt photons there is a clear trend of more events congregating on the right of the distribution. For hadronic fakes there is a clear trend of being classified to the left. The distribution for $e \to \gamma$ fakes is more or less flat, with a slight trend towards being classified as prompt. This is expected as no information about $e \to \gamma$ fakes is provided in the training of the PPT. The second figure shows the response for the same samples, however *FixedCutTight* isolation is applied to all photons. This significantly reduces the hadronic fake contribution and the overall phase space. Prompt photons are still classified effectively, however the hadronic fake distribution is more smeared out. Interestingly, prompt photons with photon type = 6 have a flat distribution for the figure in which no isolation is required, but are classified more as prompt once *FixedCutTight* isolation is applied. These prompt photons are classified as “other”, where they are not from ISR, the top quark or the top quark decay products. It is a limitation when trying to directly match prompt truth photons. This indicates that there is a contribution of hadronic fake photons that are misidentified using the truth classifying scheme that are then removed once including isolation.

The correlation between isolated photons and the shape of the PPT can be further explored as different analyses can have different isolation requirements. For instance, the photon could be very isolated (as is the case with the $t\bar{t}\gamma$ analysis) or not isolated at all (such as busy environments involving jets). Figure 6.3 shows the effect of applying increasing “tightness” of isolation to the output of the PPT. The shape variation shows that there is a correlation between the PPT and photon isolation. To account for this in the $t\bar{t}\gamma$ analysis, a systematic uncertainty is derived and applied to account for the fact that training for the PPT was performed in a non-isolated phase space. This will be discussed further in the next section.

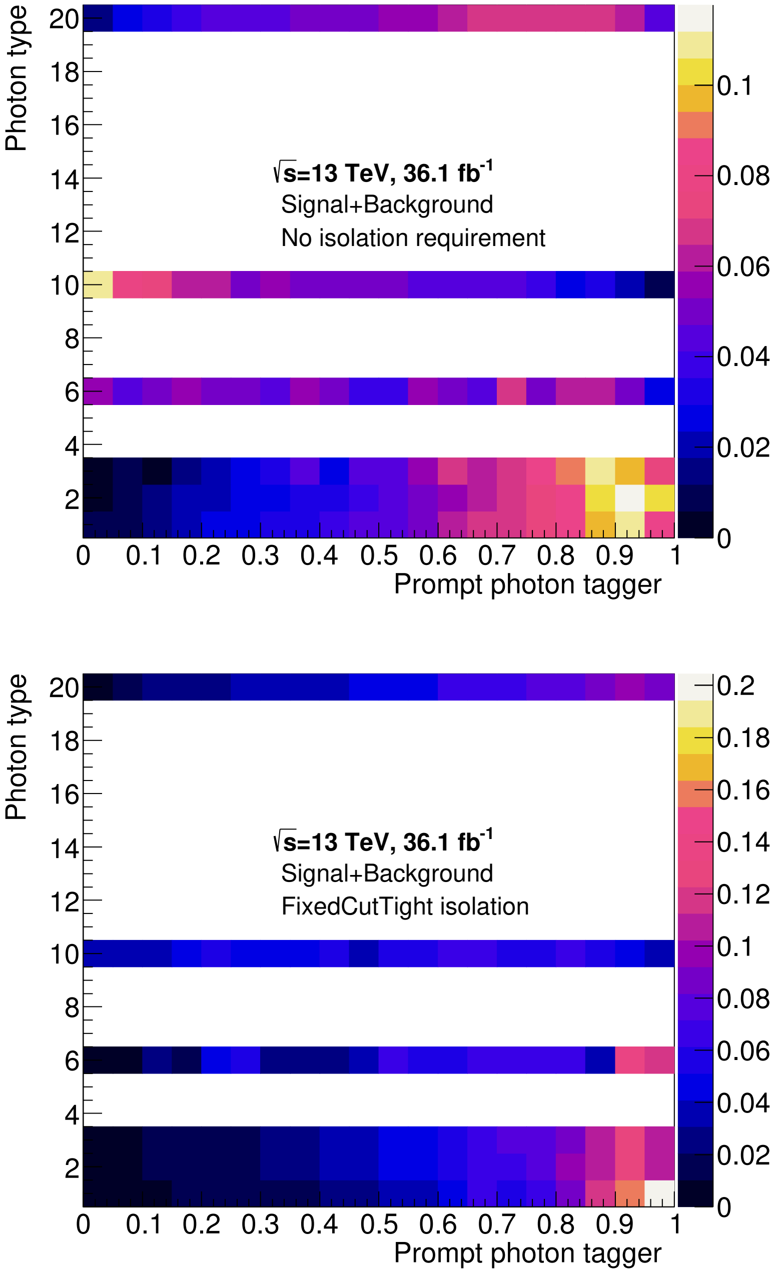


Figure 6.2.: PPT output distributions for different photon truth particle types. The distributions are obtained by applying the PPT to the $t\bar{t}\gamma$ analysis MC samples before event-level cuts, but includes object definitions such as the *Tight* photon requirement. Photon type = 20 corresponds to $e \to \gamma$ fake photons. Photon type = 10 means the photon is classified as a hadronic fake. Photon type < 10 is a signal-like photon. All rows are normalised to unity individually.

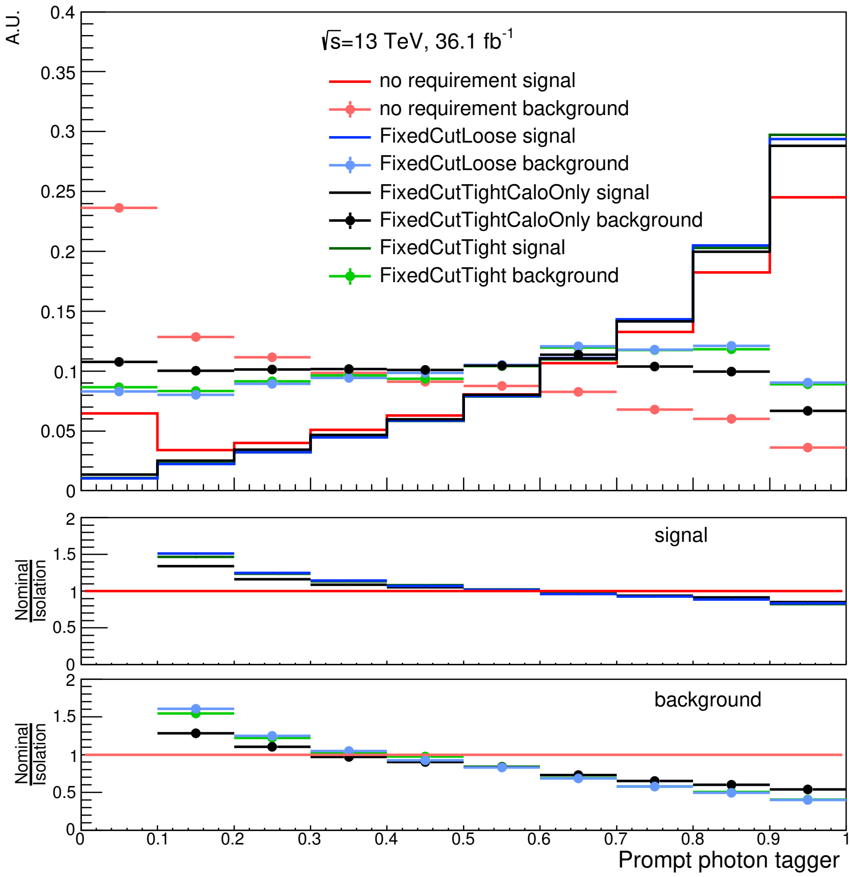


Figure 6.3.: PPT output distributions for increasing isolation cuts. The distributions are obtained by applying the PPT to the $t\bar{t}\gamma$ analysis MC samples before event-level cuts, but includes object definitions such as the *Tight* ID photon requirement.

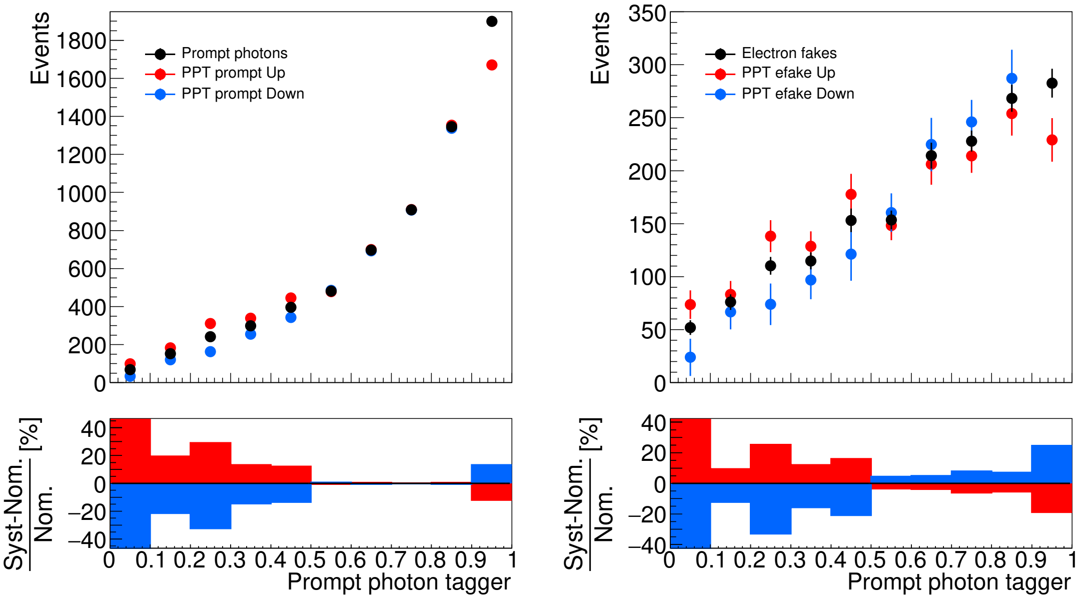
6.1.3. Scale factors and systematic uncertainties

Scale factors to correct for the mis-modelling of the PPT between MC and data are calculated in a dedicated prompt photon control region. This region corrects for prompt photons in a $Z \rightarrow \ell^+\ell^-\gamma$ enriched phase space. Exactly one photon and two opposite sign leptons are required, with an invariant mass cut between the lepton pair to be within [60, 100] GeV. The 3D scale factors for prompt photons are extracted in bins of p_{T}, η and PPT output by taking the difference between the MC and the data.

Closely linked to the scale factors are the systematic uncertainties derived for the PPT. The scale factors are switched on and off and are symmetrised to extract the up and down variations of the PPT modelling. Only a shape effect is considered and the overall number of events must remain constant. The prompt photon uncertainties are also applied to the $e \rightarrow \gamma$ fake background as the shapes are similar. For the hadronic fake background, two further sets of scale factors are derived. The first is from the hadronic fake control region C (Chapter 5.4). The signal contamination is varied by ±50% and the maximum variations between data and MC are used to estimate the uncertainty. This accounts for *Tight* yet non-isolated photons. The second set of scale factors follows a similar approach in the hadronic fake control region A. This accounts for isolated yet non-*Tight* photons.

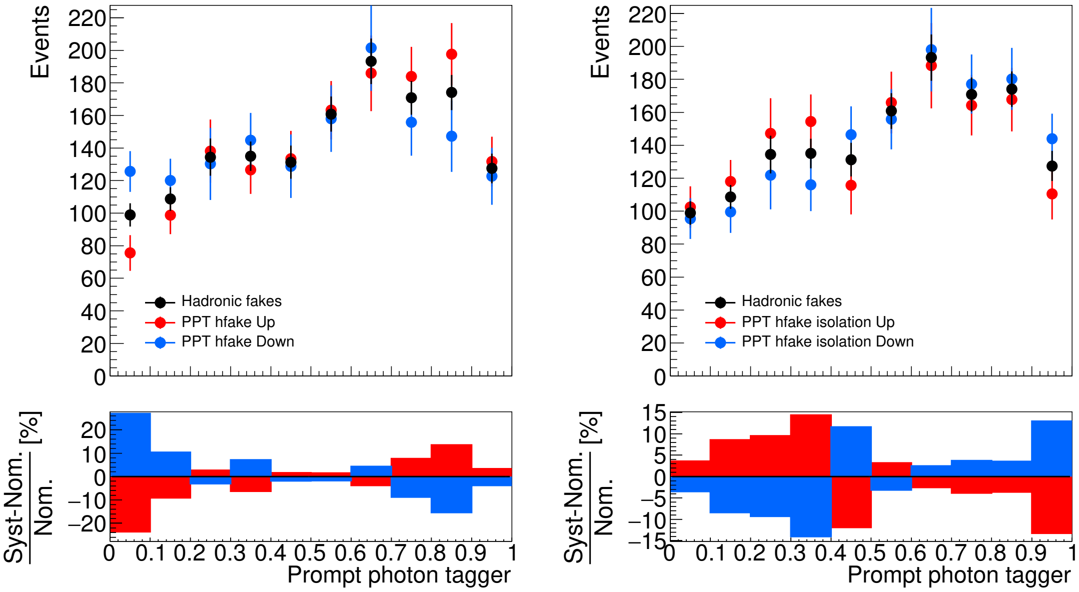
Overall, the derivation of systematic contributions follows a very conservative approach due to the novelty of the tool and the shower-shape variable mis-modelling. The effect of turning on and off the scale factors for each process can be seen in Figure 6.4. The ratio panel shows the relative contribution the systematic uncertainty has for each process, where the "up" and "down" variation labels are arbitrary. For prompt contributions, the extrema of the classifier exhibit large variations. An uncertainty/scale factor of around 15% is seen in the very prompt rich area of the PPT, which for signal rich analyses will inflate the final uncertainty.

The final PPT distribution is shown for the $t\bar{t}\gamma$ analysis in the single-lepton channel in Figure 6.5. The full signal region cuts and scale factors from previous sections are applied and total systematic uncertainties are included in the error band. This includes the PPT scale factors and uncertainties. Also shown is the normalised separation between the signal and the hadronic fake background contribution, which shows good separation for *Tight* and *FixedCutTight* selections. The shape discrepancy seen in Figure 6.5a arises from the input variables used to train the PPT, and thus the scale factors. However, this is covered by systematic uncertainties.



(a) Prompt photons

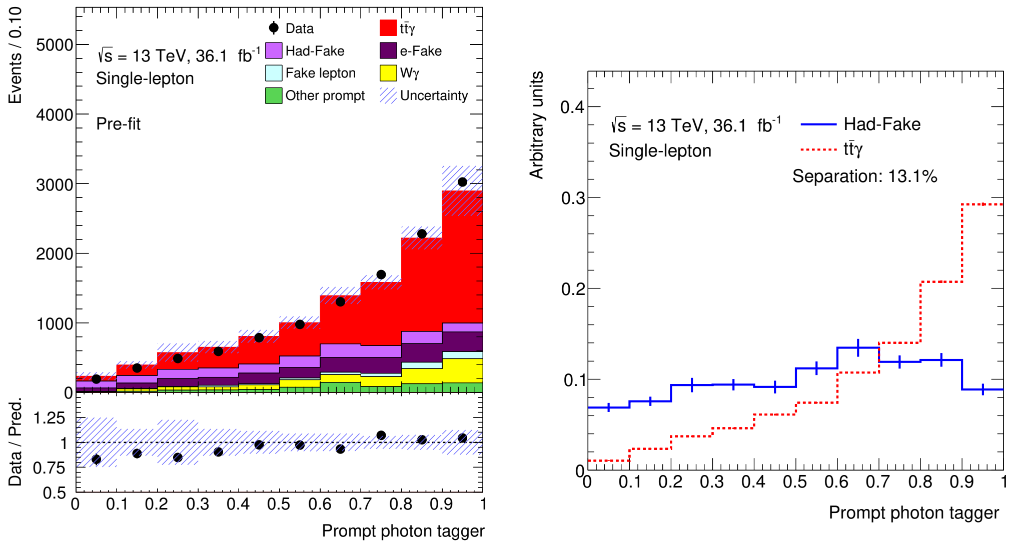
(b) $e \to \gamma$ fake photons



(c) hadronic fake photons from CR C

(d) hadronic fake photons from CR A

Figure 6.4.: The relative sizes of the PPT scale factors/systematic uncertainties for prompt, $e \to \gamma$ fake and hadronic fake photons (for the two sources of uncertainties). “Up” and “down” labels are arbitrary.



(a) (b)

Figure 6.5.: a) The PPT in the $t\bar{t}\gamma$ analysis after all event selections, scale factors, statistical and systematic uncertainties have been applied. b) The normalised separation between signal and the hadronic fake background, in which only statistical uncertainties are shown.

6.2. The Event-level Discriminator

An Event-level discriminator (ELD) has been developed separately for the single-lepton and dilepton channels. Both take the form of NNs. The single-lepton ELD is trained on events from the e+jets and μ+jets channels while the dilepton ELD is trained on events from the ee, $\mu\mu$ and $e\mu$ channels. The driving factor for this decision is due to events in the respective single-lepton and dilepton channels having very similar topologies and compositions. While there are some differences in background producing mechanisms[2], the negative impact this is meant to have on a NN is small (if any) and is outweighed by the positive impacts: simplicity and an increase in statistics. The ELD exploits topological event information (such as the number of b-tagged jets or the $E_{\mathrm{T}}^{\mathrm{miss}}$) and builds a binary classifier. This provides a new variable that helps constrain the signal and background in the maximum likelihood fit, explained in Chapter 5.1. Section 6.2.1 presents the variable selection, Section 6.2.2 explains the training methods used. Finally, Section 6.2.3 evaluates the performance of the ELD.

6.2.1. Variable selection

Two ELDs were simultaneously trained. For the single-lepton channel ELD, the full set of selection cuts defined in Section 5.2 was applied. Looser cuts were used to train the dilepton ELD. This includes dropping the two mass window cuts between the two leptons and the two leptons and the photon, neglecting the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut and not requiring a cut on the number of jets. This allows for more background contamination, enhances statistics and thus makes training more stable. Training only in the signal region (or close to the signal region) means the weights reach optimal values in a shorter amount of time.

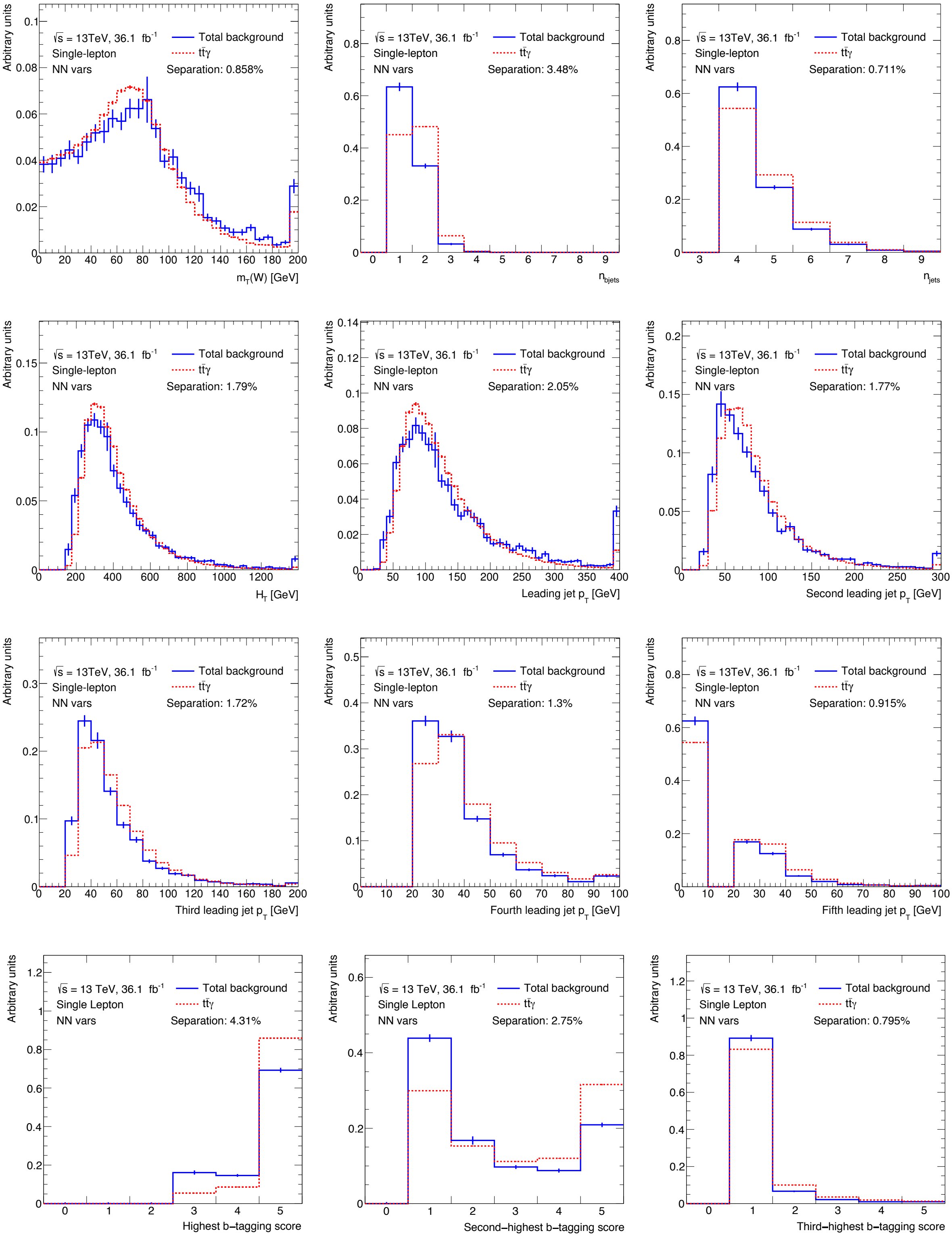
---

[2] The $e\mu$ channel has negligible $Z\gamma$ events due to the requirement of one electron and one muon, for which the Z-boson can not decay into. Electron reconstruction efficiencies are generally lower than muon reconstruction efficiencies in the detector, thus, selections relying on muon final state particles will have more events than those with electron final state particles. Fake electrons contribute to backgrounds in which final state electrons are required. Thus, the $e \to \gamma$ fake backgrounds will be larger for these channels. The fake lepton background is not only made up of fake leptons, but also leptons which are non-prompt. Thus, due to jets creating and even faking electrons, this background is significantly larger in the e+jets channel compared to the μ+jets channel.

Variables are used which highlight the different kinematics of the processes considered as signal and background. Table 6.2 shows a summary of the variables considered for the respective channels. Included is their separation power (Equation 6.1) of signal versus the sum of all backgrounds. A check-mark indicates the variable is used as an input to the final ELD[3]. As will be shown in Section 6.2.2, in general for the single-lepton ELD, more variables lead to a better sensitivity. However, the strategy of how differential cross section measurements would be performed was unclear at this stage. To remove any potential bias, the ΔR variables were excluded in case they were to be used as variables for the differential cross-section measurements. For the same reason, the p_{T} and η of the photon (not shown in the table) were not considered as inputs to the ELD. Thus, a loss of sensitivity in the ELDs was chosen so that the strategy for differential cross-section measurements would not be impaired. For the dilepton ELD, apart from the differential cross-section argument above, it was found that seven variables was the ideal scenario. This will be shown in Section 6.2.2. Initially the PPT was considered for the dilepton ELD as there is a small hadronic fake background contribution. However, after deliberation over the conservative systematic contributions that would need to be applied, it was decided that any gain would be washed out. Thus, it was dropped for the dilepton channel. The most powerful variables in both channels are due to the number of b-tagged jets as well as the pseudo continuous b-tagging weights which are described in Chapter 4.3.1. For the single-lepton channel this is then followed by the PPT. Figure 5.8 and 6.5a (the PPT) show the variables that have been used for training the single-lepton ELD. Figure 5.9 shows a slightly tighter selection (the full signal region) for the variables used in the dilepton ELD. Figure 6.6 and 6.7 show the normalised separation plots for the input variables of the respective ELD (those with check-marks in Table 6.2). Only statistical uncertainties are included. While some variables show little difference in the normalised signal and background shapes ($m(\gamma, \mathrm{lep})$ or $m_{\mathrm{T}}(W)$ in the single-lepton channel for example), two aspects need to be considered. First, these separation plots are shown for the sum of all backgrounds and might be masking the separation for a given process (such as fake lepton). Second, non-linear transformations through a neural network might result in the network learning correlations that are otherwise hard to anticipate, even from seemingly similar shapes of signal and background. Correlations between input variables and between the final ELD will be discussed in Section 6.2.3.

---

[3] For the dilepton channel, the separation is calculated when including all SR cuts as this is the phase space we are actually interested in.


Arbitrary units
$\sqrt{s}$ = 13TeV, 36.1 fb^{-1}
Single-lepton
NN vars
Total background
$t\bar{t}\gamma$
Separation: 0.858%
$m_T(W)$ [GeV]
Separation: 3.48%
n_{bjets}
Separation: 0.711%
n_{jets}
Separation: 1.79%
H_T [GeV]
Separation: 2.05%
Leading jet p_T [GeV]
Separation: 1.77%
Second leading jet p_T [GeV]
Separation: 1.72%
Third leading jet p_T [GeV]
Separation: 1.3%
Fourth leading jet p_T [GeV]
Separation: 0.915%
Fifth leading jet p_T [GeV]
$\sqrt{s}$ = 13 TeV, 36.1 fb^{-1}
Single Lepton
Separation: 4.31%
Highest b−tagging score
Separation: 2.75%
Second−highest b−tagging score
Separation: 0.795%
Third−highest b−tagging score

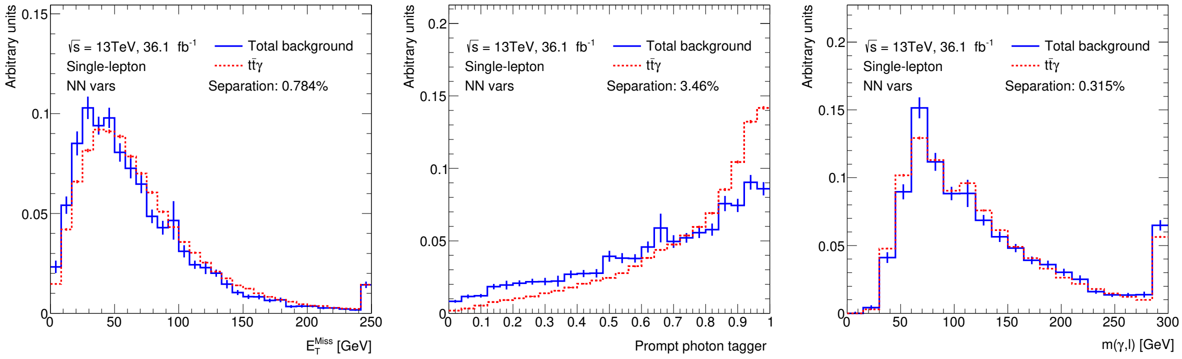


Figure 6.6.: Variables used in the single-lepton ELD showing the separation between the signal and sum of all backgrounds.

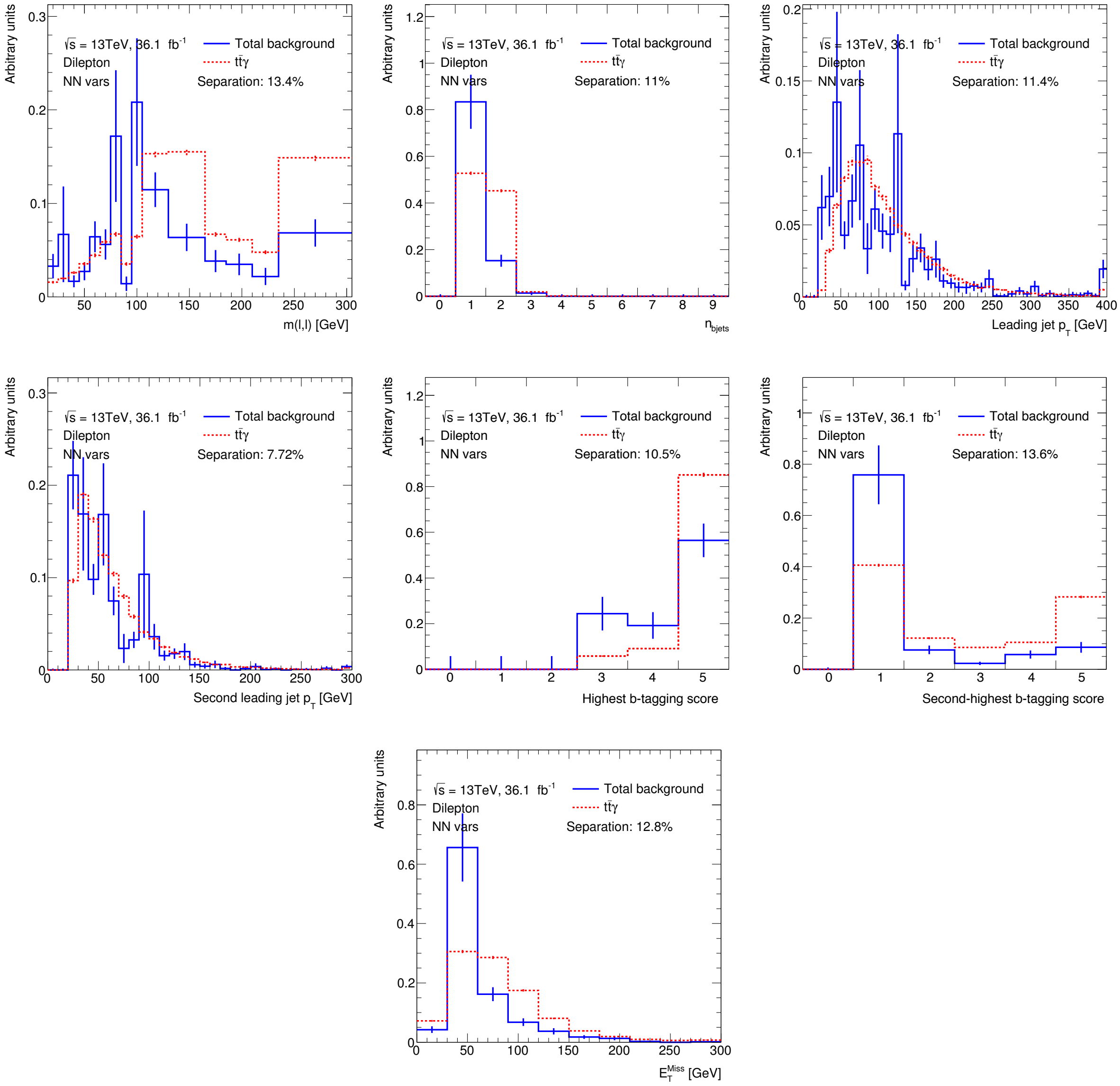


Figure 6.7.: Variables used in the dilepton ELD showing the separation between the signal and sum of all backgrounds.

Variable	Separation [%] single-lepton	Separation [%] dilepton
Prompt Photon Tagger (PPT)	3.5 ✓	6.8
H_{T}	1.8 ✓	1.3
number of jets	0.7 ✓	1.4
number of b-jets	3.5 ✓	11.0 ✓
$E_{\mathrm{T}}^{\mathrm{miss}}$	0.8 ✓	12.8 ✓
$m_{\mathrm{T}}(W)$	0.9 ✓	-
p_{T} of first jet	2.1 ✓	11.4 ✓
p_{T} of second jet	1.8 ✓	7.7 ✓
p_{T} of third jet	1.7 ✓	3.3
p_{T} of fourth jet	1.3 ✓	1.1
p_{T} of fifth jet	0.9 ✓	0.3
highest b-tagging weight	4.3 ✓	10.5 ✓
2nd highest b-tagging weight	2.8 ✓	13.6 ✓
3rd highest b-tagging weight	0.8 ✓	-
$m(\ell, \ell)$	-	13.4 ✓
$m(\gamma, \ell)$	0.3 ✓	5.8
$m(\gamma, \ell, \ell)$	-	2.3
$\Delta R(\gamma, \ell)$	0.5	13.4
$\Delta R(\gamma, jet)_{leading}$	1.8	3.1
$\Delta R(\gamma, jet)_{subleading}$	1.7	4.3

Table 6.2.: Separation of each variable used in the training of the single-lepton and dilepton ELDs. A check mark indicates that the variable was used in that NN. A dash indicates the variable does not apply to that channel.

6.2.2. Training

Training was carried out on the full set of MC samples described in Section 3.2 and the data-driven fake lepton samples described in Section 5.5. Each feature (or input variable) is normalised to give a standard deviation of 1 and a mean of 0. Keras is then used to build and train the NNs. The final output consists of separate architectural and variable input files. The variable input file not only contains a map for the variables used in the training, but also the normalisation parameters for each variable. These are essential inputs to LWTNN when reconstructing a NN inside of ATLAS code. Each of the NNs had hyper-parameter searches performed to achieve the best architecture and thus maximum separation. Presented below is a detailed description of the chosen NN architectures.

Final architecture and hyper-parameters

The NN final topology is the same for single-lepton and dilepton channels: one input dense layer consisting of 30 nodes is followed by the regularisation techniques of a dropout layer then a batch normalisation layer. A dense layer of 20 nodes follows with a final dense output layer consisting of one node with a sigmoid activation function to reduce outputs to values between 0 and 1. All layers' weights are initialised with uniform distributions and input and hidden layers receive a rectified linear unit (ReLU) activation function. The dropout layer has a 30% chance that any node and its weights will randomly be removed from the epoch.

The single-lepton NN uses 75% of the raw input events for training with 25% used for testing. This equates to 144491 events used for training and 48164 events used for testing.

Cumulatively, there are 160679 signal and 31976 background events. The dilepton NN uses 67% of the raw input events for training with 33% used for testing. This equates to 20258 events used for training and 10160 events used for testing. Cumulatively, there are 21591 signal and 8827 background events. The ratio of the split between training and testing is done to ensure that when separate k-fold validation is performed the training and test samples can be split consistently into four (three) different tests for the single-lepton (dilepton) channels. The single-lepton NN is trained using a batch size of 150 events while the dilepton NN uses 300 events. This describes the number of events used for each gradient update. Optimisation for batch size was performed using a grid search. Event weights are accounted for by scaling the binary cross-entropy loss function during training, with negative event weights set to 0.

The previous section introduced the variables considered as inputs for the NNs. Based on their separation a few training scenarios were carried out to find a stable configuration. An NN was trained with three (five) variables ranked by their separation for the single-lepton (dilepton) channel. The NN was then retrained by adding the next highly ranked variables. Tables 6.3 and 6.4 show the area under the curve (AUC) for this process, performed on an increasing number of variables. In both cases, the NN performs the best with the largest number of variables considered. However, for the dilepton NN with more than seven variables we see larger variations between multiple trainings and testings and so with the limited statistics it is deemed more unstable with overtraining likely to occur. As a result the network using seven variables was selected. A schematic of the final NNs is seen in Figure 6.8, with a summary of the hyper-parameters in Table 6.5.

Total vars.	Var. names	AUC (train)	AUC (test)	Selected NN
5	highest b-tagging weight, 2^{nd} highest b-tagging weight, 3^{rd} highest b-tagging weight, PPT, number of b-jets	0.6483	0.6474	
10	+ p_{T} of first jet, H_{T}, p_{T} of second jet, p_{T} of third jet, p_{T} of fourth jet	0.6744	0.6736	
14	+ p_{T} of fifth jet ,$m_T(W)$, $E_{\mathrm{T}}^{\text{miss}}$, number of jets,	0.6996	0.6906	
15	+ $m(\gamma, \ell)$	0.7017	0.6929	✓

Table 6.3.: Variable optimisation for the single-lepton NN. The starting point is the first row. Each subsequent row includes more variables on top of the previous row. Variables are chosen according to separation rank and physics motivation.

The performance of the chosen NNs is shown in Figure 6.9. The left plots describe the NN's response to training (filled histograms) and test (points) data. As shown in the ratio plots, good agreement can be seen. The right plots show the ROC curve with the respective AUC values for the training and test data. A line going from (0,1) to (1,0) would indicate the classifier is no better than random guesswork. If the classifier were to be 100% efficient it would be represented by a box with the outermost corner at (1,1). The AUC values for the test and training set are close enough to indicate no overtraining has occurred. The single-lepton ROC curve represents the actual performance of the ELD in the signal region. Care must be taken when reading the dilepton ROC curve since the training is performed on a larger phase space than the final signal region.

Another important parameter to consider when training a NN is the loss function, shown for both NNs in Figure 6.10. This represents the evolution in the differences between the predicted

Total vars.	Var. names	AUC (train)	AUC (test)	Selected NN
3	2^{nd} highest b-tagging weight, $m(\ell,\ell)$,$E_{\mathrm{T}}^{\mathrm{miss}}$	0.9289	0.9233	
7	+ p_{T} of first jet, number of b-jets, highest b-tagging weight, p_{T} of second jet	0.9313	0.9258	✓
8	+ PPT	0.9329	0.9261	
10	+ p_{T} of third jet, $m(\gamma,\ell,\ell)$	0.9365	0.9313	
13	+ $m_{\mathrm{T}}(W)$,number of jets, H_{T}	0.9428	0.9393	

Table 6.4.: Variable optimisation for the dilepton NN. The starting point is the first row. Each subsequent row includes more variables described on top of the previous row. Variables are chosen according to separation rank and physics motivation.

Parameter	single-lepton	dilepton
Variables	15	7
Raw training events	144491 (75%)	20258 (67%)
Raw test events	48164 (25%)	10160 (33%)
Raw signal events	160679 (83%)	21591 (71%)
Raw background events	31976 (17%)	8827 (29%)
Batch size	150	300
Optimiser	Adam	
Loss function	Binary cross entropy	
Architecture	Dense(30) → Dropout(30%) → BN → Dense(20) → Dense(1)	
Activations	ReLu and Sigmoid	
Total trainable parameters	1181	941

Table 6.5.: Summary of the parameters used in the training of the single-lepton and dilepton ELDs. The value between parenthesis for Dense and Dropout indicate the number of nodes and the dropout percentage, respectively. BN is batch normalisation.

labels and the actual labels (Equation 1.5). Care must be taken when reading the y-axis range on these plots. The single-lepton is trained on a fairly tight signal region and so from the beginning it must work hard to classify events effectively. Thus, a small improvement in the loss is seen up until the first ≈ 40 epochs, at which point the validation loss starts to increase; a clear sign of overtraining. A "patience" parameter is used such that if no improvement is seen for 50 epochs, the training is exited and the best model (at around 40 epochs) is saved. The dilepton loss has a sharply falling curve for the first few epochs due to the looser cuts and these events being easily classified as signal or background.

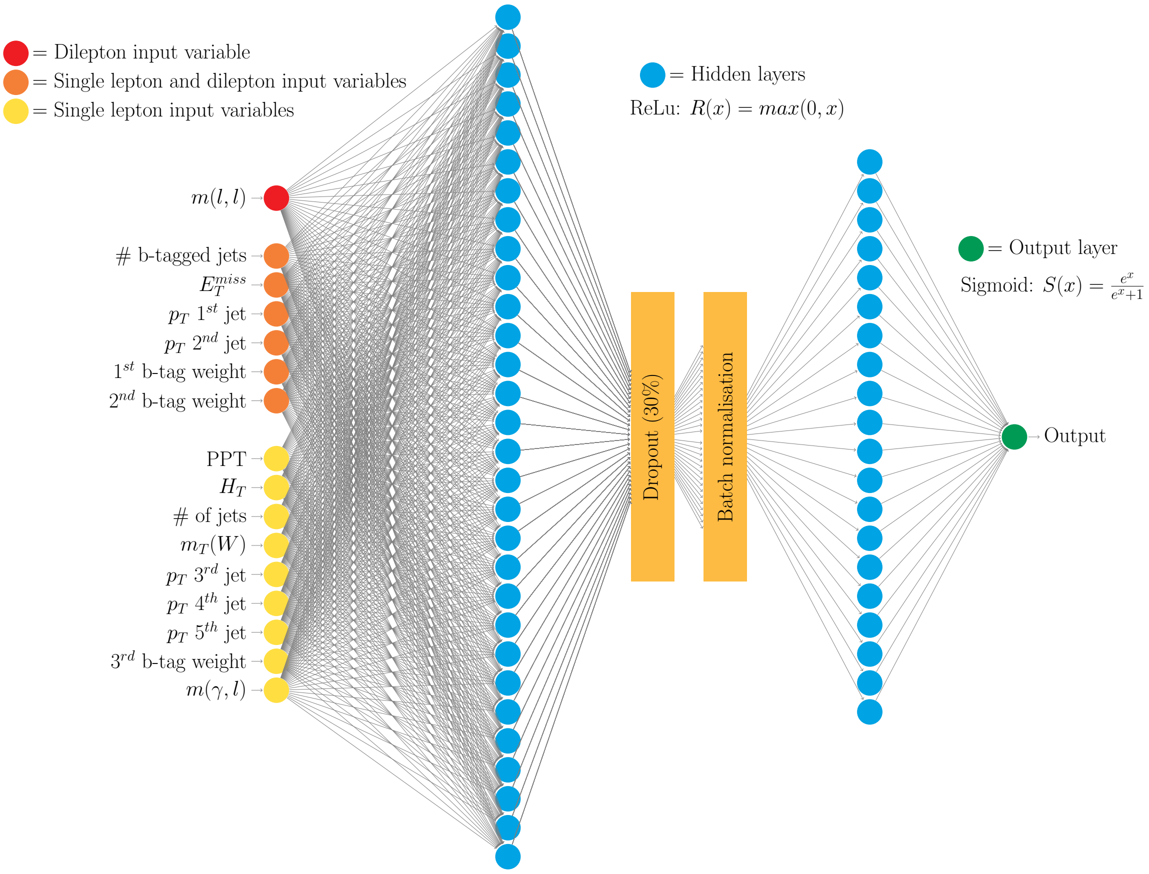
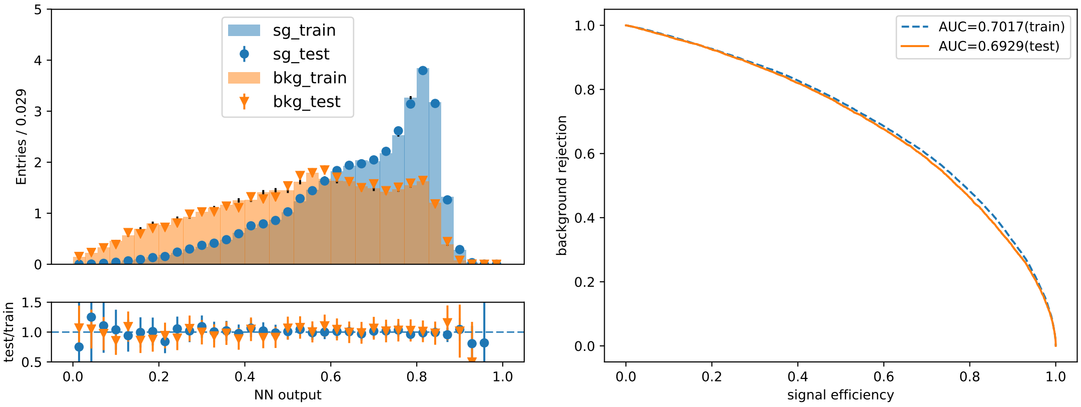
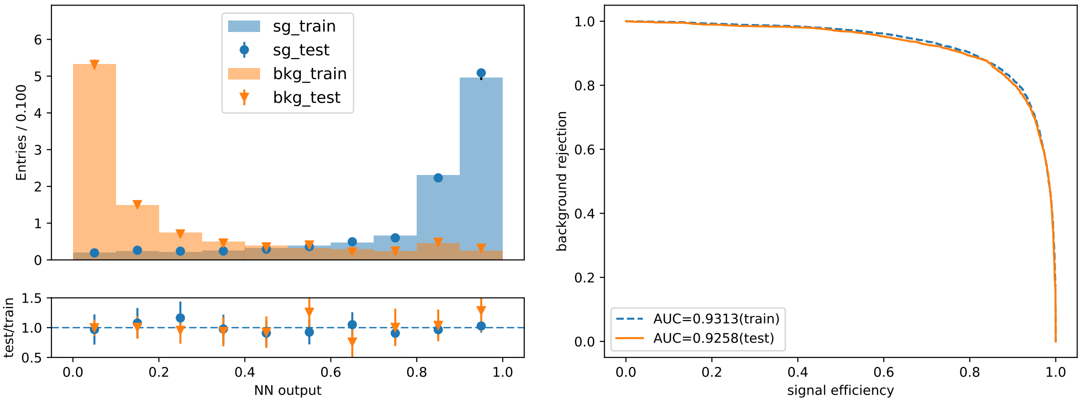

= Dilepton input variable
= Single lepton and dilepton input variables
= Single lepton input variables
= Hidden layers
ReLu: $R(x) = max(0, x)$
= Output layer
Sigmoid: $S(x) = \frac{e^x}{e^x+1}$
$m(l,l)$
# b-tagged jets
E_T^{miss}
p_T 1st jet
p_T 2nd jet
1st b-tag weight
2nd b-tag weight
PPT
H_T
# of jets
$m_T(W)$
p_T 3rd jet
p_T 4th jet
p_T 5th jet
3rd b-tag weight
$m(\gamma, l)$
Dropout (30%)
Batch normalisation
Output

Figure 6.8.: Schematic of the single-lepton and dilepton ELD showing the different input variables and layers.

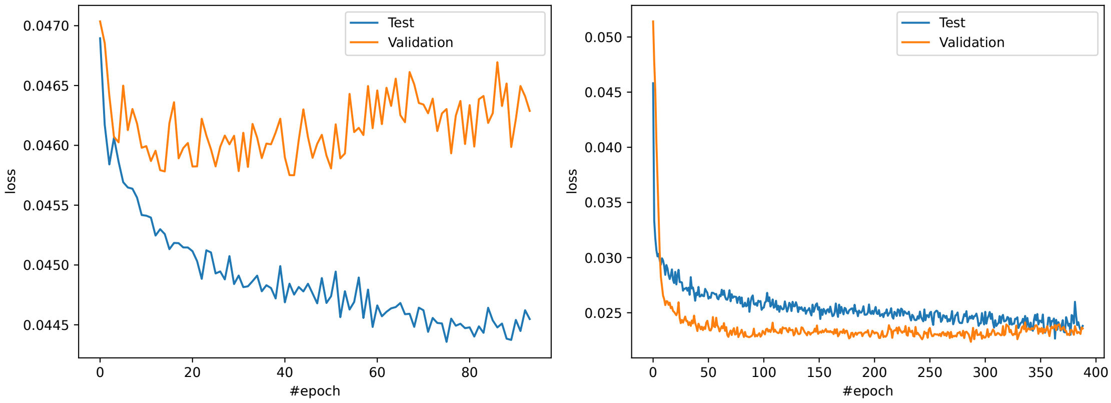


(a) single-lepton (b) single-lepton



(c) dilepton (d) dilepton

Figure 6.9.: Event-level Discriminator training output showing the test and training data agreement for the NN response and the associated ROC curve.

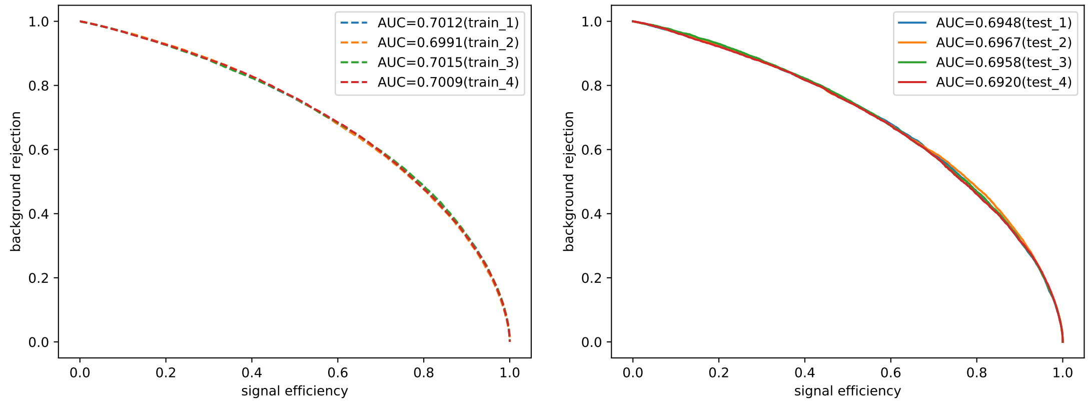


(a) single-lepton (b) dilepton

Figure 6.10.: Test and validation dataset loss curves for the single-lepton and dilepton ELD. Training is carried out so that the best model is saved according to the validation loss curve. The model stops training if no improvement is seen in the validation curve for more than 50 epochs. Note the ranges on the y-axes.

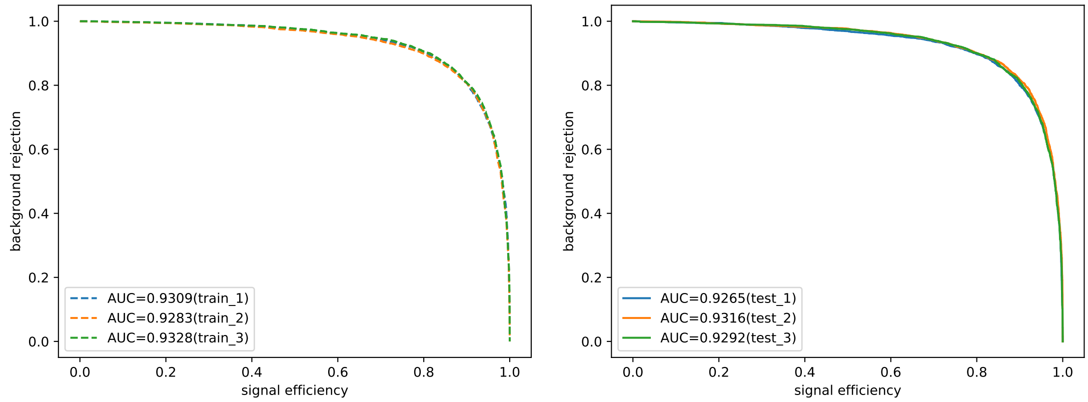
K-fold cross validation

A final check to ensure overtraining has not occurred is to perform k-fold cross validation on training data. In this test, the data has been split into four separate batches for the single-lepton case and three separate batches for the dilepton case. One batch is used as a validation or test set, the other batches are used for training the NN. Each batch has a turn to be the test set. All ROC curves should be comparable with each other and only deviate slightly from the initial ROC curve (due to statistical fluctuations). The k-fold cross validation plots are shown in Figure 6.11 for the single-lepton and dilepton NNs. Very little differences in AUC values and ROC curve shapes indicate that no overtraining occurs.



(a) single-lepton training data

(b) single-lepton test data



(c) dilepton training data

(d) dilepton test data

Figure 6.11.: The k-fold cross validation plots for each NN. The left column shows the ROC curves for the training dataset, the right column for the test dataset.

6.2.3. Evaluation

This section looks into the performance and evaluation of the ELD for each of the backgrounds. All signal and background are considered with all cuts from Chapter 5.2 applied. The output from the neural networks for each channel is shown in normalised Figures 6.12 and 6.13 for the single-lepton and dilepton channels, respectively. For the single-lepton channels the backgrounds are individually compared to the signal in Appendix C.1. Only statistical uncertainties are included. The larger error bars on some of the backgrounds are due to the MC generator weights. In general, the ELD performs well in all channels.

For the single-lepton channels the hadronic fake and $e \to \gamma$ fake backgrounds have the smallest separation at 12% and 14% for the e+jets channel, and 11% and 8% for the μ+jets channel, respectively. For the ee and $\mu\mu$ channels the main background is from $Z\gamma$, which is the background peak on the left of the ELD. A very small hadronic fake contribution makes up the background peak on the right of the ELD that sits under the signal. In the case of the $e\mu$ channel, the background peak consists of mainly hadronic fake events (due to the negligible contribution of $Z\gamma$ events in this channel). As stacked plots will show later, this is a very small contribution. These distributions are used to perform a binned maximum likelihood fit, as will be described in Chapter 5.1.

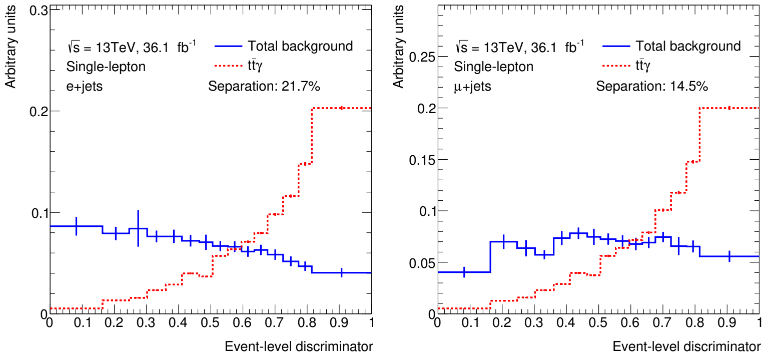


Figure 6.12.: Event-level Discriminator variable for the single-lepton channels with the corresponding separation plots for all summed backgrounds.

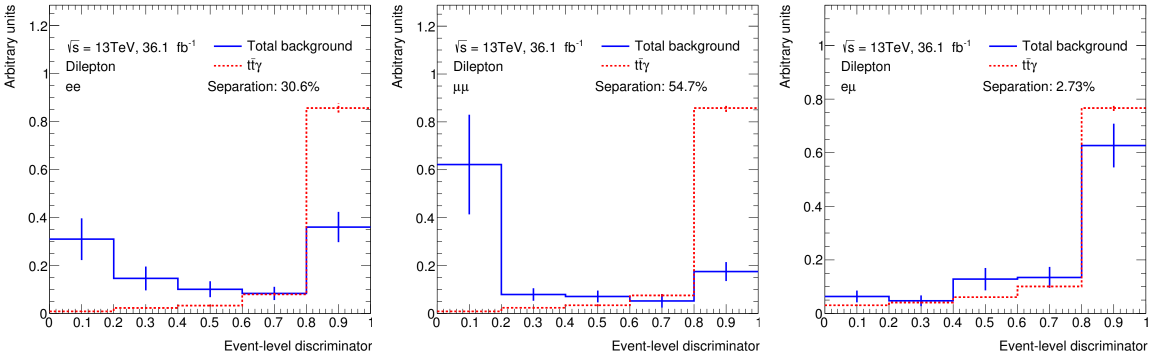
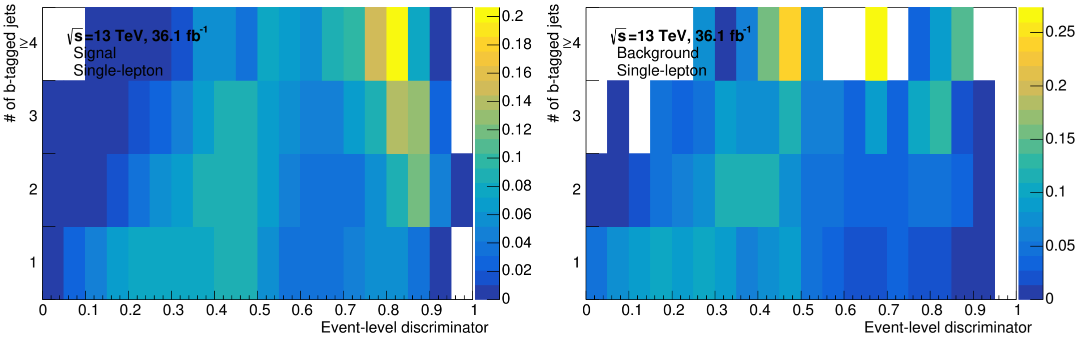


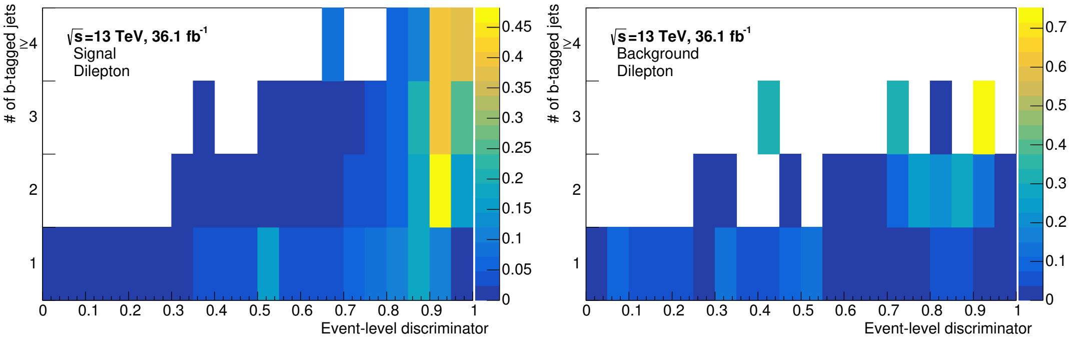
Figure 6.13.: Event-level Discriminator variable for the dilepton channels with the corresponding separation plots for all summed backgrounds.

Since the b-tagging variables play such an import role in the ELD, we can examine the relationship between the b-jet dependence and the ELD output. Figure 6.14 and 6.15 shows this correlation for the single-lepton and dilepton channel NNs, respectively, where each row of each plot has been normalised. In general, as more b-tagged jets enter into the event, the more signal-like it is. Background processes with a higher number of b-tagged jets in the event have an increased chance of being tagged as signal by the ELD.



(a) Signal (b) Background

Figure 6.14.: The number of b-tagged jets versus the single-lepton ELD output for the signal and the background. Each row is normalised separately.



(a) Signal (b) Background

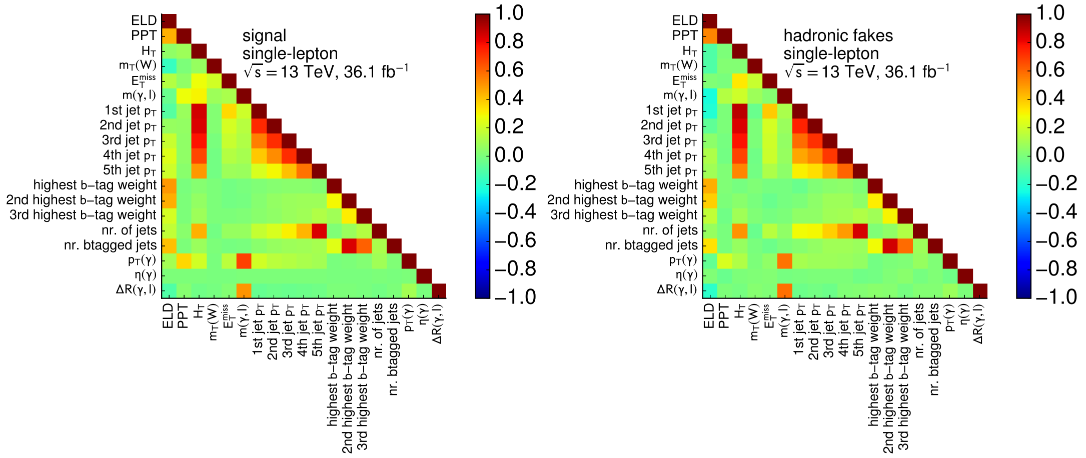
Figure 6.15.: The number of b-tagged jets versus the dilepton ELD output for the signal and the background. Each row is normalised separately.

Correlation plots for ELD output, the training input variables, and the final variables used in the differential measurements (for interest sake) are shown in Figure 6.16 and 6.17 for the single-lepton and dilepton channels, respectively. Full event selections and weights described in the previous chapters have been applied. The single-lepton correlation figures are split into the signal and the different background components. The dilepton figures are simply split into signal and the sum of all backgrounds.

An interesting observation is the correlation between the last three (five) variables used for differential measurements in the single-lepton (dilepton) ELD. These variables are not used in the training and so any correlation to the ELD that is seen has been learned by the NN. It shows that the two NNs learned very little about photon or lepton related quantities, and thus are potentially missing out on crucial event information. This is not too surprising given that very little photon or lepton kinematic information is provided as input. Future analyses can study the impact of introducing more low-level kinematic variables of the photon and leptons.

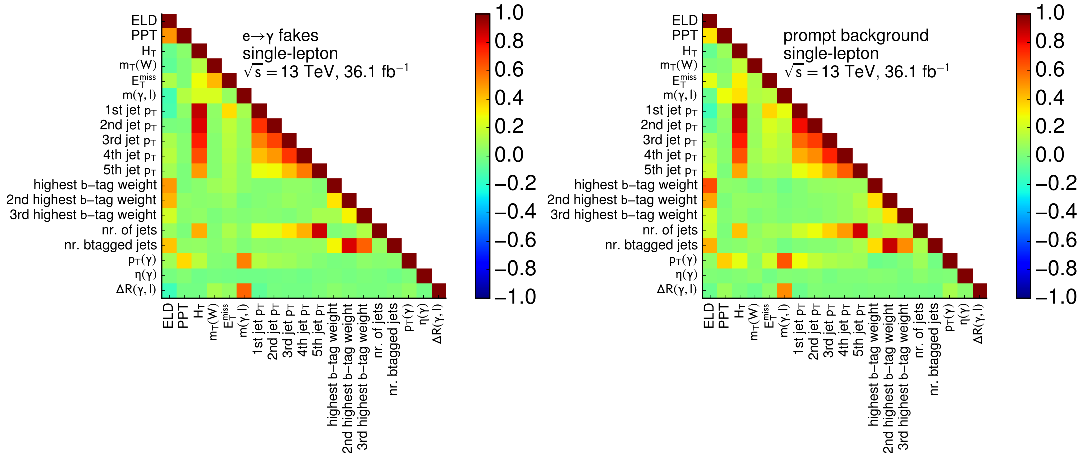
The first column essentially shows how much the ELD learns from each variable. Intuitively, the variables that have the highest separation (as shown in Table 6.2) should be amongst the most highly correlated to the ELD, however, this does not always apply. Also worthwhile to consider is that the separation values in Table 6.2 show signal against the sum of all backgrounds. For the single-lepton channel, by separating out the different backgrounds we can see that some variables are more important than what Table 6.2 leads us to believe. For instance, the invariant mass of the photon and the lepton have a very high anti-correlation for the fake lepton background in the single-lepton channel. Similarly, H_T has a high anti-correlation. In separating out the Other prompt background we see that the $E_\mathrm{T}^\mathrm{miss}$ of an event plays a fairly important role. This is because the $E_\mathrm{T}^\mathrm{miss}$ spectrum is softer for events coming from the $V\gamma$, single top, diboson, and $t\bar{t}V$ processes. As expected some of the highest correlations to the ELDs are consistent with respect to the b-tagging variables in both signal and backgrounds.

The second column for the single-lepton figures shows the correlations between all the variables and the PPT. The general trend is that the PPT is minimally correlated with the rest of the variables. This is expected since it was trained on shower shape variables, which are for the most part analysis independent. An exception occurs for the transverse momentum of the photon, for which shower shape variables are known to have some dependence. Even then, the correlation is never greater than about 30%.



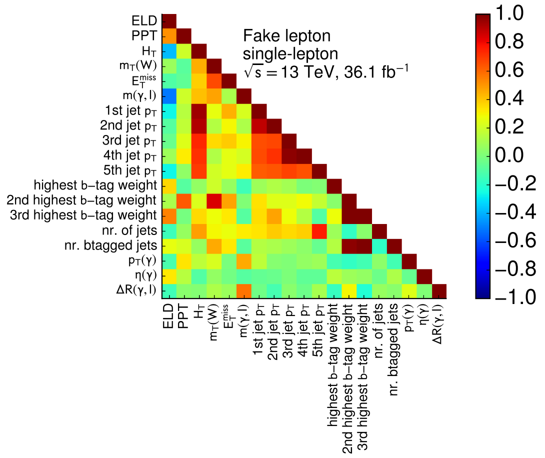
(a) Signal

(b) Hadronic fake background



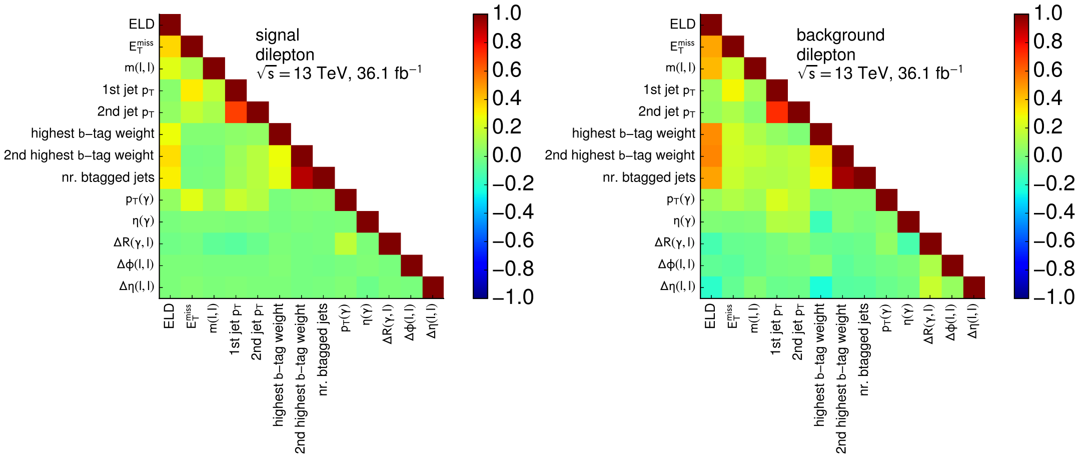
(c) $e \rightarrow \gamma$ fake background

(d) Prompt background



(e) Fake lepton background

Figure 6.16.: Correlation plots for the signal and separate backgrounds for the single-lepton channel comparing the ELD and training variables. Full event selection and event weights have been applied. The last three variables ($p_T(\gamma)$, $\eta(\gamma)$ and $\Delta R(\gamma, l)$) do not enter the training and are used for differential measurements.



(a) Signal

(b) Total background

Figure 6.17.: Correlation plots for the signal and sum of backgrounds for the dilepton channel comparing the ELD and training variables. Full event selection and event weights have been applied. The last five variables ($p_T(\gamma)$, $\eta(\gamma)$, $\Delta R(\gamma, l)$, $\Delta\phi(l, l)$ and $\Delta\eta(l, l)$) do not enter the training and are used for differential measurements.

CHAPTER 7

Systematic uncertainties

Many sources of uncertainties need to be considered for any analysis in HEP. They fall under two classes: statistical and systematic. Statistical uncertainties are related to the amount of data collected or MC available. Thus more data or MC events means lower statistical errors, which is needed for precision measurements. Systematic errors are essentially due to limitations when building a detector, or our lack of full understanding for a given physics object/algorithm or detector component. Lack of theoretical knowledge also contributes to systematic uncertainties. Within ATLAS some systematic uncertainties are provided as "recommendations" due to studies carried out by other teams, for example luminosity or Jet Energy Resolution (JER). There are also many systematics that arise from analysis specific techniques, such as those when training a NN or deriving a data-driven background.

Systematic uncertainties in template fits have two components; a shape and an overall normalisation. A shape component can change the shape of the distribution by assigning more or less events in various bins, but the overall integral must remain the same. A normalisation component shifts the integral up or down by some constant amount. A systematic can have either shape, normalisation, or both. A concrete example will be shown in Section 7.3.

This section introduces all the systematics considered for this analysis. The modelling uncertainties of the signal and backgrounds are shown in Section 7.1. The experimental uncertainties, which are common to both signal and backgrounds are explained in Section 7.2. Finally, a discussion on how systematics can be smoothed, symmetrised, and pruned for the final fit is presented in Section 7.3.

7.1. Signal and background modelling uncertainties

In the following systematic sources, references to hadronic fake and $e \to \gamma$ fake modelling refer to the $t\bar{t}$ components that make up these backgrounds, since this is the dominant process.

7.1.1. QCD renormalisation and factorisation scales

The QCD renormalisation and factorisation scale uncertainties are applied with a normalisation and shape component to the signal and hadronic fake processes. For $e \to \gamma$ fake, $W\gamma$ and $Z\gamma$ background processes (for the respective channel) only the shape is considered. The effect of the choice of the renormalisation (μ_r) and factorisation (μ_f) scales is estimated by varying μ_r and μ_f simultaneously or independently up and down by a factor of two with respect to the nominal sample value, thus resulting in three sets of up-and-down variations. This implies changing the dependence of the finite order perturbation calculations in QCD. The final uncertainty on the normalisation is the quadratic sum of the three sets of variations, while for the shape the maximum variation is chosen.

7.1.2. Parton Shower

The parton shower uncertainty is derived and applied to the signal process as well as the hadronic fake, $e \to \gamma$ fake and where applicable, the $Z\gamma$ backgrounds.

For $t\bar{t}\gamma$, the uncertainty is estimated by comparing the nominal sample (`MG5_aMC` + `Pythia 8`) to the same `MG5_aMC` events showered with `Herwig 7`. For hadronic fake and $e \to \gamma$ fake backgrounds the uncertainty is estimated by comparing the nominal sample (`PowHeg-Box` +`Pythia 8`) to a $t\bar{t}$ sample produced by `Sherpa`. The $Z\gamma$ shape uncertainty in the dilepton channels is estimated by comparing the nominal sample (`Sherpa`) to events generated with `MG5_aMC` + `Pythia 8`.

7.1.3. Initial and final state radiation

The initial and final state QCD radiation systematics (ISR/FSR) are applied to the signal process as well as the hadronic fake and $e \to \gamma$ fake background processes. The uncertainty is estimated by comparing the nominal sample to samples produced with variations of the `A14` tune.

7.1.4. PDF uncertainty

The PDF systematics uncertainties are only applied to the signal process. The shape and normalisation aspect is evaluated by the envelope error of the 100 PDF error sets in the `NNPDF2.3LO` PDF set, which are stored as weights in the nominal sample. For the shape component this is inconvenient as it impacts the speed of the fit. Thus, separate studies were done to see the impact on the signal strength and its sensitivity. If this shape affect is negligible, it can be left out of the final fit. The impact is calculated by performing 100 fits (a fit for each variation) and computing the standard deviation with the respect to the central value of μ. The errors relative to μ for each channel are shown in Table 7.1. Tests for an ideal pruning value were performed and indicate all of these values are well below the chosen value of 0.7% (Section 7.3).

Channel	Relative error on μ [%]
e+jets	0.066
μ+jets	0.066
ee	0.137
$\mu\mu$	0.109
$e\mu$	0.027
single-lepton	0.066
dilepton	0.200

Table 7.1.: The relative uncertainty on μ for the 100 PDF shape uncertainty fits.

7.1.5. Other background modelling uncertainties

Further sources of systematic uncertainties for the hadronic fake and $e \to \gamma$ fake backgrounds have already been presented in Chapter 5.3 and 5.4. These result from the tag-and-probe and ABCD methods.

For non-floating MC backgrounds without a modelling systematic error (single top, diboson and $t\bar{t}V$ backgrounds) a flat uncertainty of 50% is placed on the cross section. As a cross-check, this uncertainty is doubled and leads to negligible changes. This is shown in Appendix D.1. In the single-lepton channels the $Z\gamma$ background is also assigned a 50% normalisation uncertainty, similarly the same is done for the $W\gamma$ background in the dilepton channels.

The estimation for the fake lepton background systematics is presented in Section 5.5. An envelope of the different parameterisations used in the matrix method is calculated, with the outlying parameterisations taken as the up/down variations. This source of uncertainty has both shape and normalisation components. This uncertainty is large and so no other contributions are considered.

7.2. Experimental uncertainties

A large source of experimental systematic uncertainties arise from algorithms that simulate and reconstruct leptons, photons, jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ in the ATLAS detector. Different pileup profiles also need to be simulated and therefore have an associated error. Another source of error associated with the LHC and detectors is the integrated luminosity (used in the normalisation of the MC). The above mentioned sources can change the overall shape and normalisation of the MC, and thus this needs to be taken into account. For data-driven corrected backgrounds the normalisations are already corrected but the shapes of the distributions could still be biased. Thus, only shape components are considered. Similarly for when the $W\gamma$ background is a free parameter in the final fit, the overall normalisation is adjusted. Thus, only shape systematics are considered for the $W\gamma$ background in the single-lepton channels.

Leptons and photons

Leptons (electrons and muons) and photons are corrected with $E_{\mathrm{T}}/p_{\mathrm{T}}$ and η data-driven scale factors from identification and isolation efficiency measurements discussed in Chapter 4.1 and 4.2. Corresponding efficiencies are derived, for which a few examples include: varying the amount of material in front of the calorimeters, studying the effects of different generators, varying the background contamination or background subtraction in the fits, varying the fraction of converted photons in data and simulation (for photon related systematics) and varying the isolation cone sizes [140, 142, 143].

Prompt photon tagger

The systematic uncertainties associated with the PPT are discussed in Chapter 6.1.3. They are derived separately for prompt and hadronic fake photons in the form of scale factors, which are then turned on and off. One source is assigned to the $e \rightarrow \gamma$ fake background (due to similarity in shape to the prompt photons), a source is assigned to each of the prompt contributions (both signal and background and correlated in the fit), and two sources enter for the hadronic fake background. All PPT systematic variations have only a shape component.

Jets

Systematic contributions assigned to jets include: pileup, flavour composition (quarks and gluons), η-calibrations, calorimeter responses to different jet flavours, punch-through corrections and single-particles (for high p_{T} uncertainties). These are derived from simulation, test beam data and *in-situ* measurements [147, 162–164]. An important contribution to the JES uncertainty scheme is due to pileup corrections. Specifically, the uncertainty of the per-event p_{T}

density modelling in the η-ϕ plane for MC simulations. This is termed *Pileup RhoTopology*. The JVT systematic contribution is derived by varying the cut on the multivariate output.

b-tagging

The numerous uncertainties on b-tagging come from separate data-driven methods applied to different jet flavours: b-jets, c-jets and *light*-jets, which have 30, 15 and 80 NPs, respectively [165].

Missing Transverse Energy

The $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated from all objects in an event according to Equation 4.3. Thus, the associated uncertainties of each object are propagated through. Additionally, the scale and resolution of the soft terms are considered.

Pileup

Events are re-weighted in MC to match the number of interactions per bunch crossing in data (Chapter 3.1). Systematic uncertainties for pileup are evaluated by scaling these distributions up and down.

Luminosity

The integrated luminosity of the 2015/2016 data at 13 TeV is measured to a precision of 2.1%. This uncertainty is used to normalise the MC signal and background samples, and as such the uncertainty is applied to these MC samples. The derivation of this uncertainty has been derived using x-y beam separation scans using a similar methodology described in [166].

Table 7.2 details all the systematics and the number of components considered for this analysis.

7.3. Smoothing, symmetrisation and pruning

Smoothing and *symmetrisation* are methods in which statistical fluctuations in various systematic sources are minimised. In the case of symmetrisation the systematic is centred around a mean value. *Pruning* occurs when a systematic contribution is below some threshold in the fit and removed. This serves to remove a potential cause of instabilities and speeds up the fit.

Smoothing and symmetrisation

Two-sided symmetrisation is performed when an up and a down variation is provided for any given systematic. The difference between the two variations is calculated then divided by the mean of the variations. This value is then taken as positive (for up) and negative (for down). This is described by Equation 7.1.

$$\text{Variation up/down} = \pm \left| \frac{\text{up} - \text{down}}{(\text{up} + \text{down})/2} \right| \tag{7.1}$$

Thus, the $\pm 1\sigma$ variation is centred around the nominal value. It is important to notice that if the up/down variations are symmetric then applying this procedure has no effect. *One-sided* symmetrisation is when only an up or down variation is provided. An example is the PPT systematics. In this case the variation is simply mirrored to reflect the supplementary variation.

Systematic name	Type	Components	Notes
Signal modelling			
ISR/FSR	SN	1	
Parton Shower	SN	1	
QCD Scales	SN	1	
PDF	SN	1	
Background modelling			
ISR/FSR	SN*, S**	2	*(hadronic fake), **($e \to \gamma$ fake)
Parton Shower	SN*, S**	3	*(hadronic fake), **($e \to \gamma$ fake, $Z\gamma$)
QCD Scales	S	4	hadronic fake, $e \to \gamma$ fake, $W\gamma$, $Z\gamma$
cross section normalisation	N	4	50% for VV, ST, $t\bar{t}V$, $Z\gamma$, $W\gamma$ in dilepton
fake lepton	SN	1	
$e \to \gamma$ fake	SN	11	
hadronic fake	SN	25	
Prompt Photon Tagger			
Prompt photons	S	1	Split into $t\bar{t}\gamma$, $W\gamma$ and Other prompt contributions and correlated in fit
$e \to \gamma$ fake	S	1	
hadronic fake	S	1	
hadronic fake isolation	S	1	
Object reconstruction			All, except data-corrected backgrounds receive only S
Electrons (trigger, reco, ID, isolation)	SN	5	
Egamma (resolution, scale)	SN	2	
Muons (trigger, reco, ID, isolation)	SN	15	
$E_{\mathrm{T}}^{\mathrm{miss}}$ (resolution, scale)	SN	3	
Photons (efficiency, isolation)	SN	3	
Jet energy scale	SN	21	
Jet energy resolution	SN	1	
Jet vertex tagger	SN	1	
b-tagging efficiency	SN	30	
c-tagging efficiency	SN	15	
Light-jet tagging efficiency	SN	80	
b-tagging extrapolation	SN	1	
Miscellaneous			
Luminosity	N	1	All except data-corrected backgrounds
Pileup re-weighting	SN	1	All, except data-corrected backgrounds receive only S

Table 7.2.: Complete list of systematics considered. A "N" indicates that only normalisation has been considered as a systematic. A "S" indicates that only the shape is considered as a systematic. "SN" indicates that both shape and normalisation have been considered as a systematic.

In general, experimental systematic sources are symmetrised, while signal and background modelling contributions are not.

Smoothing algorithms average statistics across bins. This prevents large statistical spikes in many of the systematic uncertainties that are expected to give small contributions. In certain cases such as for signal, $t\bar{t}$ and $Z\gamma$ modelling, this option is turned off to allow for maximum shape variations[1]. The smoothing algorithms make use of bin and neighbouring bin information such as integrals, statistical uncertainties and derivative sign changes. The smoothing does not change overall normalisations.

An arbitrary nuisance parameter is shown for three toy examples in Figure 7.1 as an example of smoothing and symmetrisation. Dotted lines show the original nominal and up/down variations before any smoothing or symmetrisation has been performed. The solid red/blue lines represent the systematic effect that will enter the fit after the symmetrisation and smoothing has been performed. If a systematic uncertainty is not symmetrised and is not smoothed, or if these operations have very little effect, the respective dotted and solid lines will be overlaid. The hatched black line is the statistical error for the nominal signal or background sample. Figure 7.1a shows an arbitrary, non-negligible systematic uncertainty applied to some arbitrary signal or background process. The overall normalisation contribution is shown in the parenthesis on the top right (for both the up and down variations). The ratio plot shows the shape and the normalisation contribution. It is important to note that the shape contribution is the relative difference of one bin to the bordering bins. One can not have a shape-only systematic contribution if there is only a single bin. Figure 7.1b shows this same systematic uncertainty applied to the same MC, but with the normalisation effect removed. Thus, this systematic uncertainty only has a shape aspect. This is shown in the ratio plot by looking at bin height and derivative changes, with respect to the previous bin. Figure 7.1c shows this same systematic uncertainty where only an overall normalisation effect of 3.7% has been applied. An example of such a systematic uncertainty would be the luminosity or the cross-section normalisation for the Other prompt backgrounds.

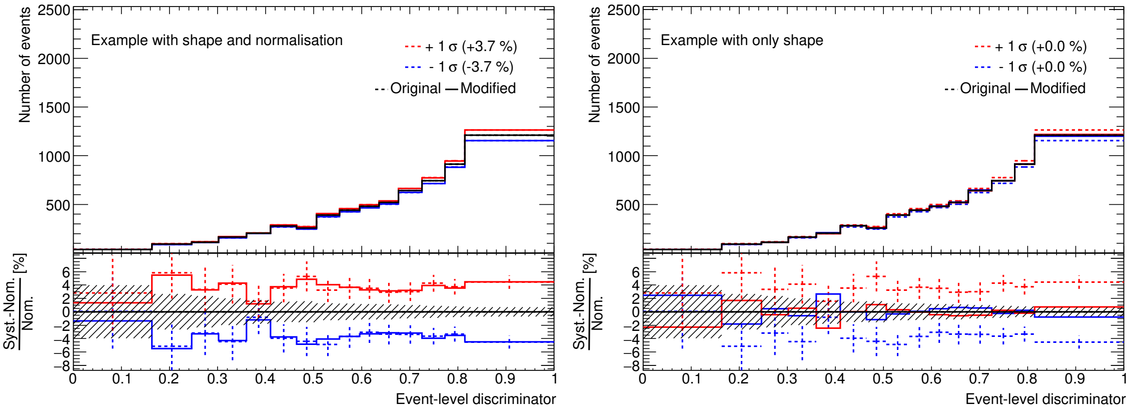
These plots are crucial in diagnosing problematic NPs in the final fit and so will be shown frequently in Chapter 8.

Pruning

The pruning of systematic uncertainties is done for the normalisation and shape separately. The $\pm 1\sigma$ variation is calculated for each NP and an initial fit performed. For either shape or normalisation, if the effect on the uncertainty is less than the given threshold, the component is removed from further fits. To determine the threshold a range of scenarios were tested. The chosen value of 0.7% sees little difference between all systematic uncertainties used in the fit and only those that are greater than this threshold. The NPs that pass this threshold are shown in Figure 7.2 for the single-lepton channels and Figure 7.3 for the dilepton channels. Differences between the single-lepton, e+jets and μ+jets (dilepton, ee, $\mu\mu$ and $e\mu$) channels mainly arise from different efficiencies for electrons and muons. Small differences in background compositions will also result in very slight pruning disparities. “Not Present” indicates that either the NP does not apply for that process (for example “$t\bar{t}\gamma$ Parton shower” for the hadronic fake background) or that the NP has been pruned.

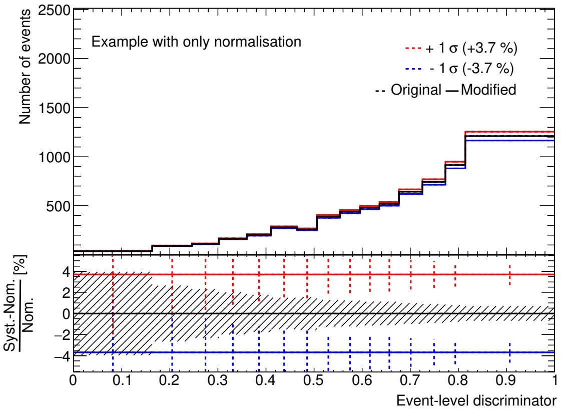
---

[1] For the dilepton channels, the $t\bar{t}$ modelling is smoothed due to very small background contributions and large statistical fluctuations.



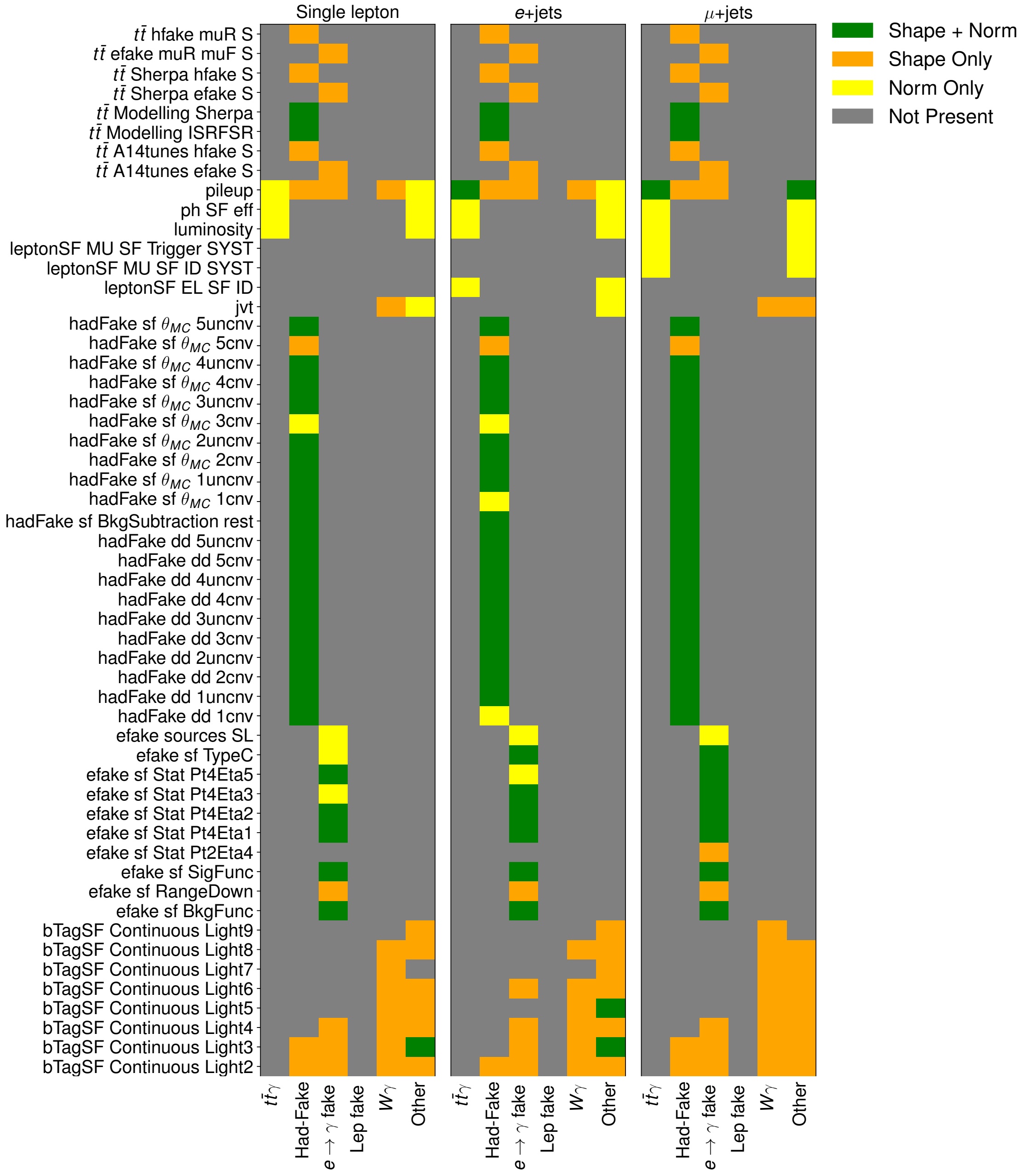
(a) Shape and normalisation

(b) Shape only



(c) Normalisation only

Figure 7.1.: An arbitrary example showing the effects of smoothing and symmetrisation for the same systematic source applied to the same physics process for systematic uncertainties that have contributions from shape and normalisation, shape only, and normalisation only. The dotted line represents the uncertainty before smoothing and symmetrisation has been applied. The solid line represents the after effect and the nuisance parameter that will enter the fit.


Single lepton
e+jets
μ+jets
Shape + Norm
Shape Only
Norm Only
Not Present
$t\bar{t}$ hfake muR S
$t\bar{t}$ efake muR muF S
$t\bar{t}$ Sherpa hfake S
$t\bar{t}$ Sherpa efake S
$t\bar{t}$ Modelling Sherpa
$t\bar{t}$ Modelling ISRFSR
$t\bar{t}$ A14tunes hfake S
$t\bar{t}$ A14tunes efake S
pileup
ph SF eff
luminosity
leptonSF MU SF Trigger SYST
leptonSF MU SF ID SYST
leptonSF EL SF ID
jvt
hadFake sf θ_{MC} 5uncnv
hadFake sf θ_{MC} 5cnv
hadFake sf θ_{MC} 4uncnv
hadFake sf θ_{MC} 4cnv
hadFake sf θ_{MC} 3uncnv
hadFake sf θ_{MC} 3cnv
hadFake sf θ_{MC} 2uncnv
hadFake sf θ_{MC} 2cnv
hadFake sf θ_{MC} 1uncnv
hadFake sf θ_{MC} 1cnv
hadFake sf BkgSubtraction rest
hadFake dd 5uncnv
hadFake dd 5cnv
hadFake dd 4uncnv
hadFake dd 4cnv
hadFake dd 3uncnv
hadFake dd 3cnv
hadFake dd 2uncnv
hadFake dd 2cnv
hadFake dd 1uncnv
hadFake dd 1cnv
efake sources SL
efake sf TypeC
efake sf Stat Pt4Eta5
efake sf Stat Pt4Eta3
efake sf Stat Pt4Eta2
efake sf Stat Pt4Eta1
efake sf Stat Pt2Eta4
efake sf SigFunc
efake sf RangeDown
efake sf BkgFunc
bTagSF Continuous Light9
bTagSF Continuous Light8
bTagSF Continuous Light7
bTagSF Continuous Light6
bTagSF Continuous Light5
bTagSF Continuous Light4
bTagSF Continuous Light3
bTagSF Continuous Light2
$t\bar{t}\gamma$
Had-Fake
$e\rightarrow\gamma$ fake
Lep fake
$W\gamma$
Other

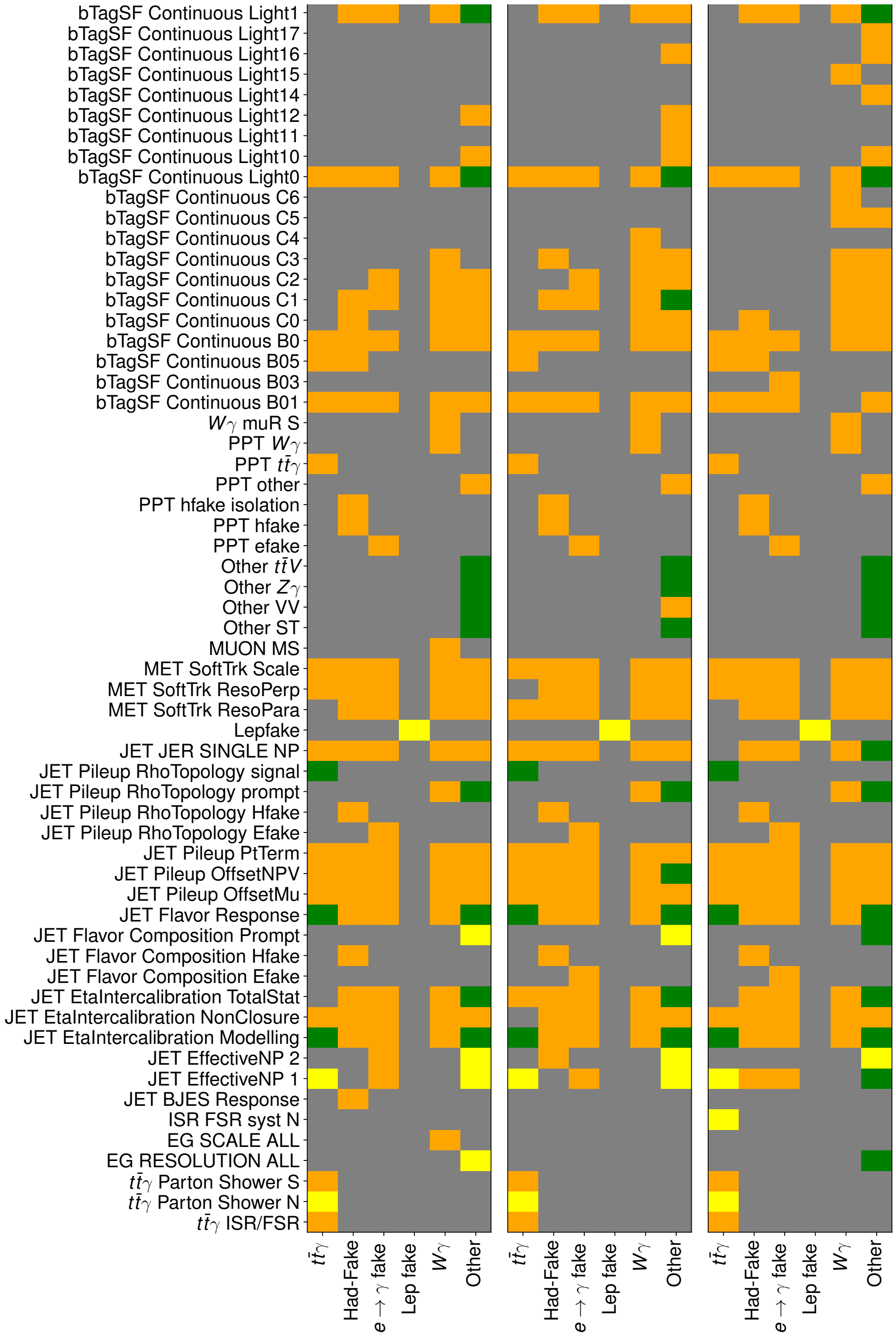
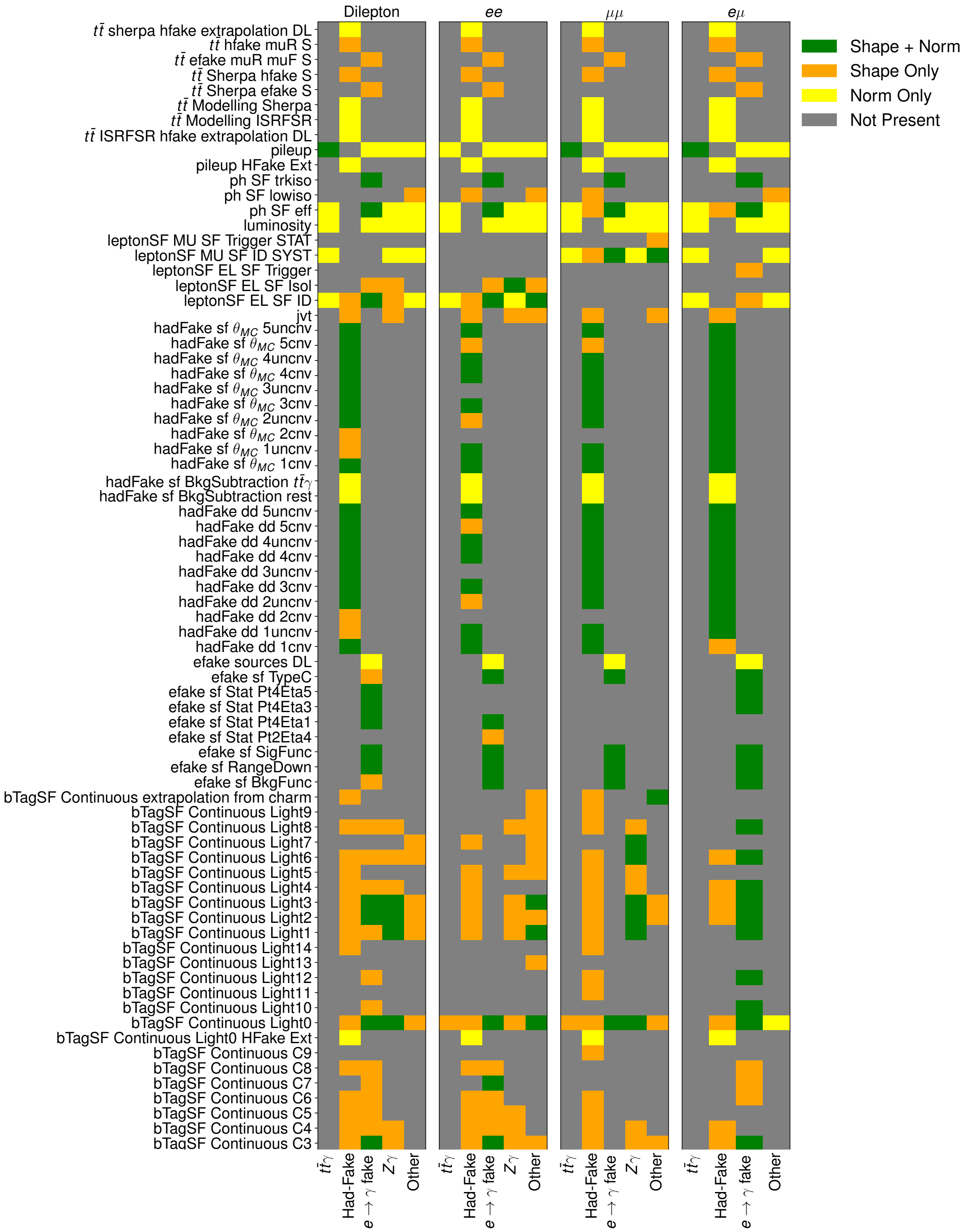


Figure 7.2.: Schematic showing the systematic sources that survive the pruning procedure for each process. From left to right the single-lepton, e+jets and μ+jets channels are shown. Only systematic uncertainties that pass the pruning for at least one process are included. Grey either indicates the systematic was pruned, or it does not apply for the process.


Dilepton
ee
$\mu\mu$
$e\mu$
Shape + Norm
Shape Only
Norm Only
Not Present
$t\bar{t}$ sherpa hfake extrapolation DL
$t\bar{t}$ hfake muR S
$t\bar{t}$ efake muR muF S
$t\bar{t}$ Sherpa hfake S
$t\bar{t}$ Sherpa efake S
$t\bar{t}$ Modelling Sherpa
$t\bar{t}$ Modelling ISRFSR
$t\bar{t}$ ISRFSR hfake extrapolation DL
pileup
pileup HFake Ext
ph SF trkiso
ph SF lowiso
ph SF eff
luminosity
leptonSF MU SF Trigger STAT
leptonSF MU SF ID SYST
leptonSF EL SF Trigger
leptonSF EL SF Isol
leptonSF EL SF ID
jvt
hadFake sf θ_{MC} 5uncnv
hadFake sf θ_{MC} 5cnv
hadFake sf θ_{MC} 4uncnv
hadFake sf θ_{MC} 4cnv
hadFake sf θ_{MC} 3uncnv
hadFake sf θ_{MC} 3cnv
hadFake sf θ_{MC} 2uncnv
hadFake sf θ_{MC} 2cnv
hadFake sf θ_{MC} 1uncnv
hadFake sf θ_{MC} 1cnv
hadFake sf BkgSubtraction $t\bar{t}\gamma$
hadFake sf BkgSubtraction rest
hadFake dd 5uncnv
hadFake dd 5cnv
hadFake dd 4uncnv
hadFake dd 4cnv
hadFake dd 3uncnv
hadFake dd 3cnv
hadFake dd 2uncnv
hadFake dd 2cnv
hadFake dd 1uncnv
hadFake dd 1cnv
efake sources DL
efake sf TypeC
efake sf Stat Pt4Eta5
efake sf Stat Pt4Eta3
efake sf Stat Pt4Eta1
efake sf Stat Pt2Eta4
efake sf SigFunc
efake sf RangeDown
efake sf BkgFunc
bTagSF Continuous extrapolation from charm
bTagSF Continuous Light9
bTagSF Continuous Light8
bTagSF Continuous Light7
bTagSF Continuous Light6
bTagSF Continuous Light5
bTagSF Continuous Light4
bTagSF Continuous Light3
bTagSF Continuous Light2
bTagSF Continuous Light1
bTagSF Continuous Light14
bTagSF Continuous Light13
bTagSF Continuous Light12
bTagSF Continuous Light11
bTagSF Continuous Light10
bTagSF Continuous Light0
bTagSF Continuous Light0 HFake Ext
bTagSF Continuous C9
bTagSF Continuous C8
bTagSF Continuous C7
bTagSF Continuous C6
bTagSF Continuous C5
bTagSF Continuous C4
bTagSF Continuous C3
$t\bar{t}\gamma$
Had-Fake
$e \to \gamma$ fake
$Z\gamma$
Other

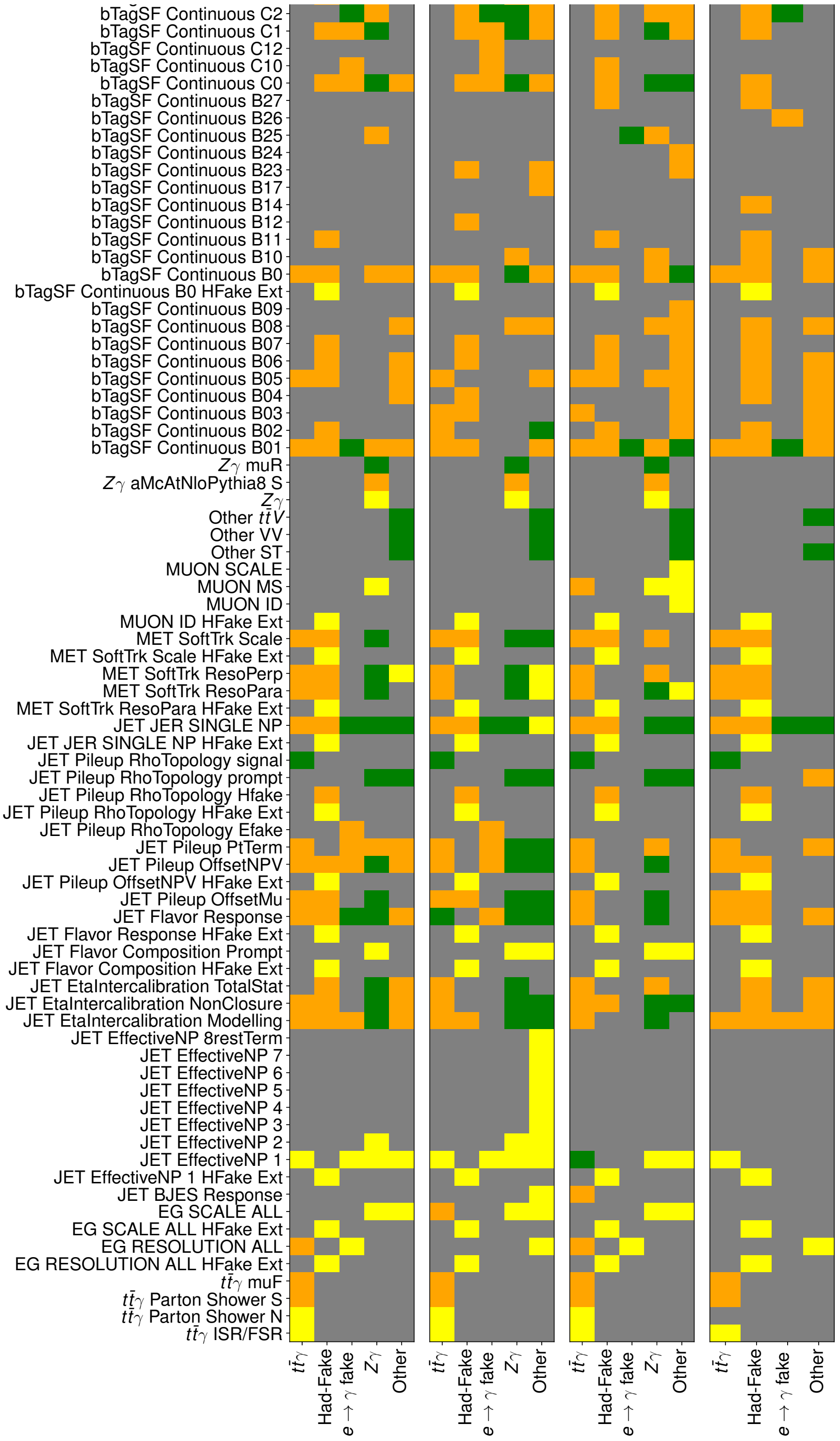


Figure 7.3.: Schematic showing the systematic sources that survive the pruning procedure for each process. From left to right the dilepton, ee, $\mu\mu$ and $e\mu$ channels are shown. Only systematic uncertainties that pass the pruning for at least one process are included. Grey either indicates the systematic was pruned, or it does not apply for the process.

CHAPTER 8

Results

The section presents the culmination of the material introduced and studies performed in the previous chapters. Results directly relating to the final measurements, as well as the final measurements themselves are presented here. Section 8.1 documents studies and tests done using pseudo-datasets. This includes binning optimisations for the ELD as well as preliminary nuisance parameter checks. Section 8.2 presents the final fit-to-data results along with further nuisance parameter discussions. Lastly, Section 8.3 presents the final fiducial cross-section results.

8.1. Asimov fits

Asimov[1] datasets are pseudo datasets that essentially imply our signal and background shapes and normalisations are completely understood, thus pseudo data can be built as signal+background for each bin. A fit using this dataset (an *Asimov fit*) will by design yield a signal strength of one with some uncertainty. Asimov fits can be used to check the expected sensitivity[2] of a model as well as how the NPs behave.

8.1.1. ELD binning optimisation

The choice of binning for the final discriminating variable used in the maximum likelihood fit plays an important role in constraining signal and backgrounds. A bin that contains only a single background process will help constrain the NPs that apply to that background for the rest of the distribution. Three different Asimov fit scenarios are tested on the single-lepton and dilepton channels to determine the best bin sizes for the ELD. All systematic uncertainties are included in these scenarios. The first scenario uses a uniform binning distribution with ten and six bins for the single-lepton and dilepton channels, respectively. Another scenario uses an automatic binning algorithm. The algorithm scans the distribution and merges bins from the right until a certain fraction of signal and background remains in the bin. For each bin, a threshold for this merging is defined as

$$Z = z_b \frac{n_b}{N_b} + z_s \frac{n_s}{N_s}, \tag{8.1}$$

---

[1] The name “Asimov” originates from Isaac Asimov’s short story, *Franchise* [167]. In it, the single most representative member of the population is elected to vote on behalf of everyone.

[2] The sensitivity of a model is a measure of how large the (relative or absolute) uncertainties are.

where n_s and n_b are the signal and background content, respectively. N_s and N_b are the total number of signal and background events in the full ELD distribution. This function takes two free parameters, z_s and z_b, which define the maximum allowed fraction of signal and background events in each bin for $z_s + z_b =$ number of bins. The algorithm stops iterating when $Z > 1$.

The results for each of the scenarios are shown in Table 8.1. For the single-lepton channel small differences are seen between each scenario. However, the AutoBin 2 binning algorithm consisting of 15 bins yields the highest sensitivity and so will be used. For simplicity, this binning is also applied to the individual e+jets and μ+jets channels. The final binning is $[0, 0.16, 0.25, 0.30, 0.36, 0.41, 0.46, 0.51, 0.55, 0.60, 0.64, 0.68, 0.73, 0.77, 0.81, 1]$. For the dilepton channel, similarly, small differences are seen. The AutoBin 2 algorithm performs slightly better, however with fewer events as is seen in the ee, $\mu\mu$ and $e\mu$ channels it seems prudent to limit the number of bins. For simplicity sake, the uniform binning with six bins is chosen for all dilepton (combined and separate) channels.

Channel	Binning	"+" error	"-" error
single-lepton	Uniform	0.0934	0.0915
	AutoBin 1 ($z_s = 5, z_b = 5$)	0.0923	0.0897
	AutoBin 2 ($z_s = 12, z_b = 3$)	**0.0911**	**0.0888**
dilepton	**Uniform**	**0.0678**	**0.0629**
	AutoBin 1 ($z_s = 5, z_b = 5$)	0.0663	0.0626
	AutoBin 2 ($z_s = 12, z_b = 3$)	0.0652	0.0606

Table 8.1.: The different Asimov fit scenarios for the single-lepton and dilepton channels. All systematics are included in the fit. The error represents the absolute up/down error of μ. The chosen binning is highlighted in bold text.

8.1.2. Asimov fit cross checks

NPs enter the fit with a Gaussian prior constraint with a mean of 0 and a width (root mean square) associated with the normalisation uncertainty of 1σ. If an NP should be *constrained* in the fit, the posterior width will be narrower than the prior. If an NP should be *pulled* in a fit, the central value of the Gaussian distribution will no longer be 0 (thus having a normalisation different to the prediction). In an Asimov fit all NPs are kept constant at 0 and the constraining power of each is checked. Significant constraints need to be explored as it could mean we do not understand our data (or our NP) as well as we should.

The posteriors for all NPs and channels are shown in Figure 8.1 and 8.2 in *pull-plots* for the individual and combined single-lepton and dilepton channels, respectively[3]. Also included is the signal strength point-of-interest (POI) and in the case of the single-lepton channels, the floating $W\gamma$ normalisation (which in an Asimov fit is held constant at one). From Section 7.3, a pruning threshold of 0.7% is applied to all systematics in each channel. One NP might be over the threshold for one channel but fail it for another. Thus, the pull-plots could have a "missing" contribution for a given channel. For those systematics that are constrained, further discussions on the priors are presented to ensure that no problems exist.

For the single-lepton channels there are four constrained NPs that warrant closer checks, of which three are related to $t\bar{t}$ modelling. This is not too surprising since dedicated scale factors are derived for the nominal sample and not the systematic variation samples.

---

[3] The Asimov pull-plot for the 5-channel inclusive fit is not shown as any observed constraints would need to be traced back to the individual channels.

- $t\bar{t}\gamma$ parton shower: In all single-lepton channels this systematic uncertainty is constrained. The shape component of the NP that enters each fit is shown in Figure 8.3 for the single-lepton channel (these figures are similar for the e+jets and μ+jets channels). Large up/down variations from the `Herwig` sample can be seen. Given that this NP is correlated to the signal (and thus the signal strength), it is not surprising there is a potential for it to be constrained in the fit.

- $t\bar{t}$ parton shower: This NP is constrained in all single-lepton channels. The shape-only contributions (due to being corrected with data-driven scale factors) for the hadronic fake and $e \to \gamma$ fake backgrounds are shown in Figure 8.4. In the low end of the ELD distribution large shape discrepancies are seen.

- $t\bar{t}$ ISR/FSR: This NP is constrained in all single-lepton channels, with the un-symmetrised, shape-only contributions for the hadronic fake and $e \to \gamma$ fake backgrounds shown in Figure 8.5. Large shape discrepancies are seen.

- Fake lepton background: The up/down variations for the fake lepton background are from the extremes of the parameterisation "envelope" used in the data driven method (Chapter 5.5). Thus, one expects large relative differences to the nominal distribution as shown in Figure 8.6. Given that the overall contribution of the fake lepton background is small and spread out due to the ELD, the effect that this NP has on the final result is small.

For the dilepton channels the main constrained NPs arise because of the modelling of the $Z\gamma$ background. This includes the shape due to the parton shower and the conservative normalisation of 50% placed on the cross section. The $Z\gamma$ parton shower shape can be seen in Figure 8.7, which shows clear differences to the nominal sample. The outcome of this mis-modelling is that uncertainties in the ee and $\mu\mu$ channels will be larger and thus more conservative.

In most channels the JER is also seen to be slightly constrained. This is simply due to the size of the NP. This and the other few constraints observed in the Asimov fits will be further discussed in the upcoming section with fits to real data.

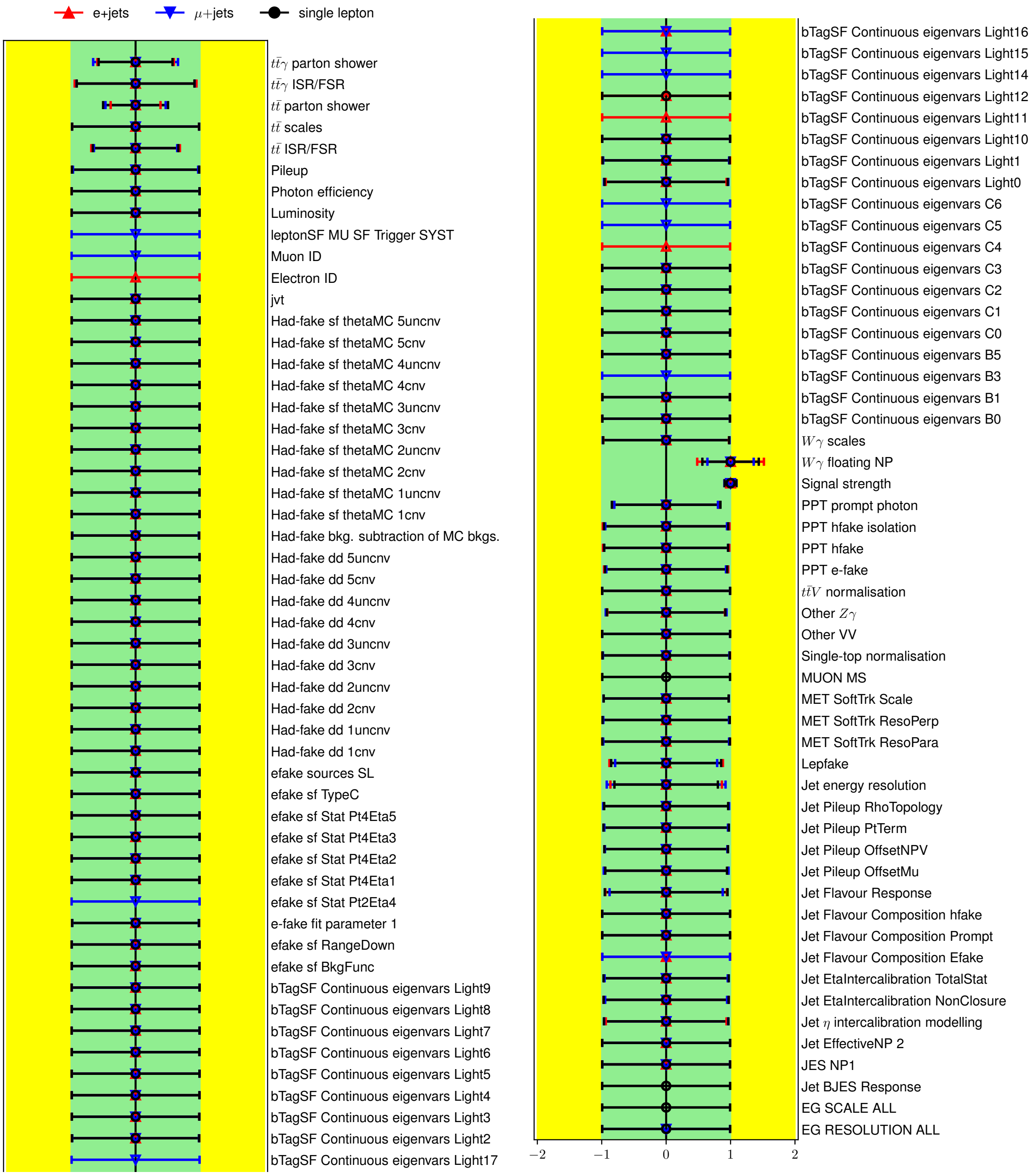


Figure 8.1.: Asimov post-fit pull plots for all nuisance parameters in the single-lepton channels. The “Signal strength” and “$W\gamma$ floating NP” both have expectation values held fixed at one. “Had-fake” and “efake” NPs relate to the data-driven methods and the uncertainties that arise from estimating these backgrounds. The “bTagSF” NPs are associated with pseudo-continuous b-tagging for different flavour jets.

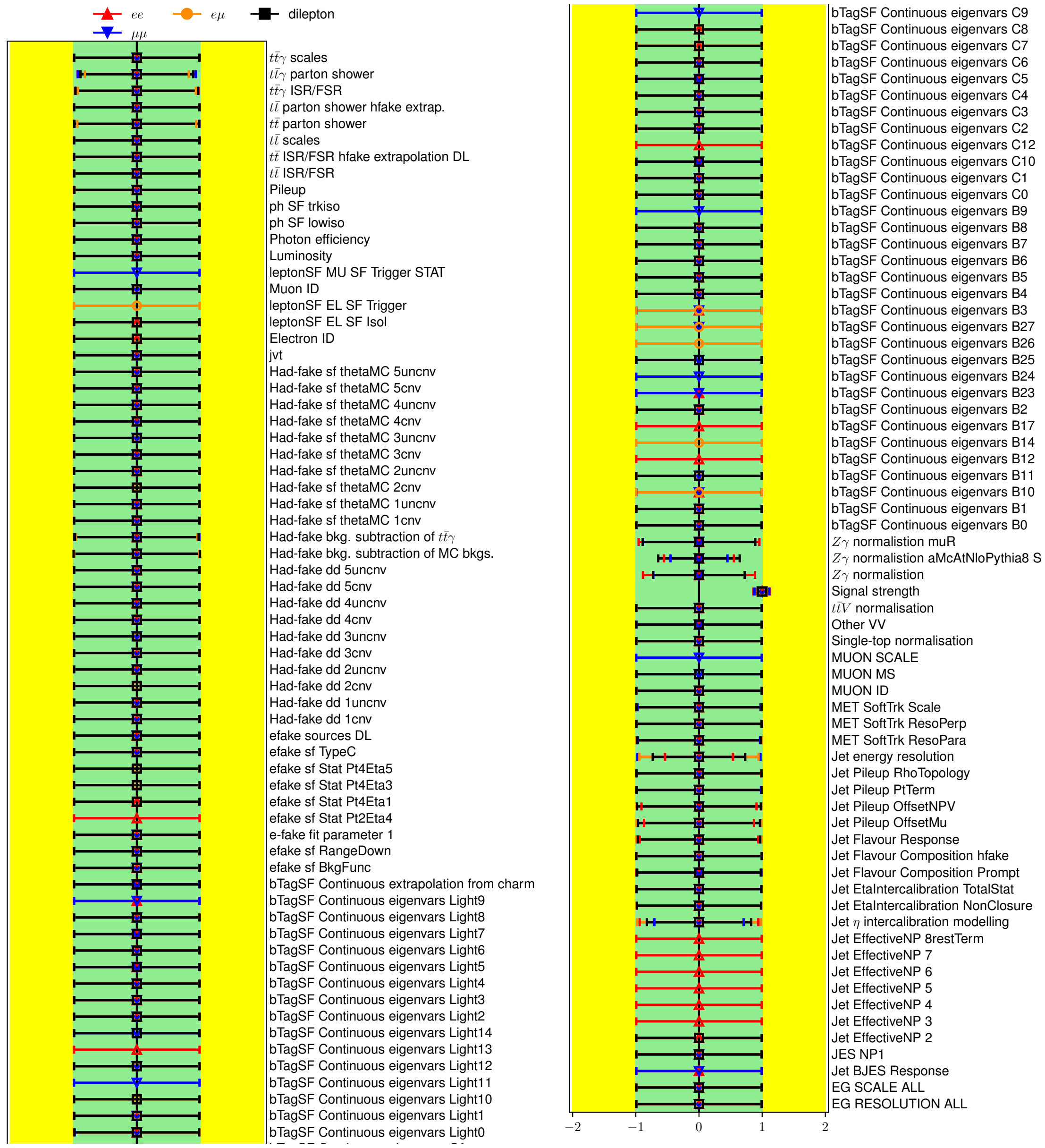
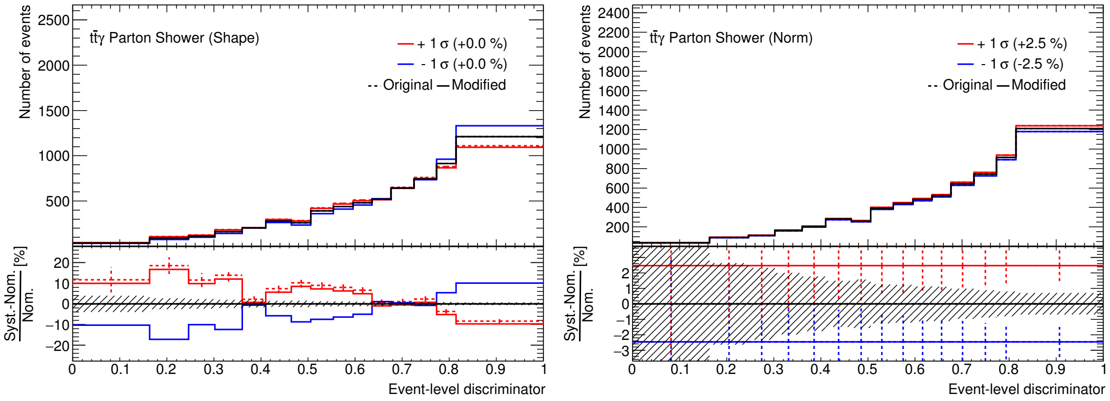
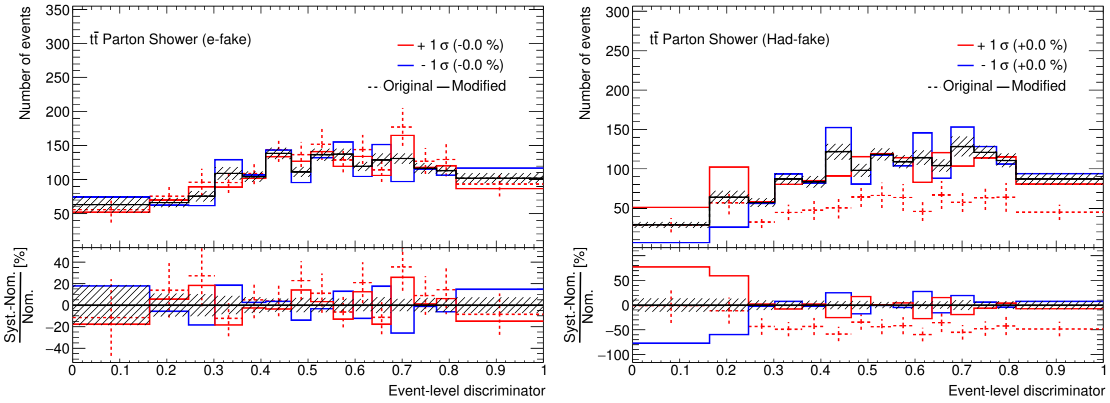


Figure 8.2.: Asimov post-fit pull plots for all nuisance parameters in the dilepton channels. The “Signal strength” has an expectation value held fixed at one. “Had-fake” and “efake” NPs relate to the data-driven methods and the uncertainties that arise from estimating these backgrounds. The “bTagSF” NPs are associated with pseudo-continuous b-tagging for different flavour jets.



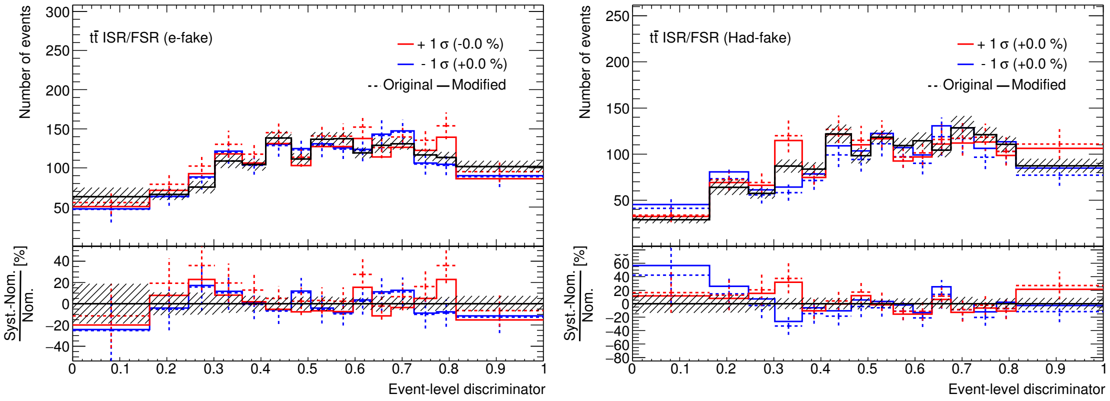
(a) Shape (b) Normalisation

Figure 8.3.: The signal parton shower systematic for the single-lepton channel.



(a) $e \rightarrow \gamma$ fake (b) hadronic fake

Figure 8.4.: The shape only component of the $t\bar{t}$ parton shower NP for the hadronic fake and $e \rightarrow \gamma$ fake background processes in the single-lepton channel.



(a) $e \rightarrow \gamma$ fake (b) hadronic fake

Figure 8.5.: The shape only component of the $t\bar{t}$ ISR/FSR NP for the hadronic fake and $e \rightarrow \gamma$ fake background processes in the single-lepton channel.

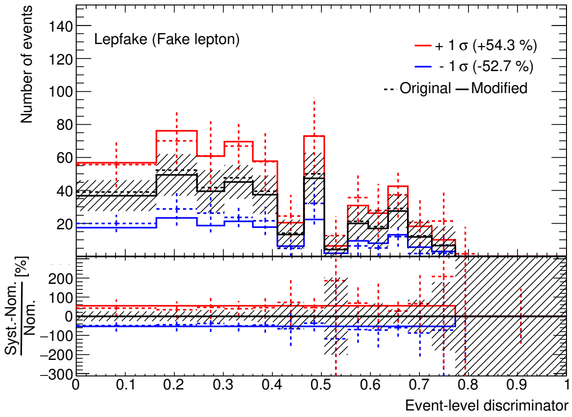


Figure 8.6.: The fake lepton background NP for the single-lepton channel.

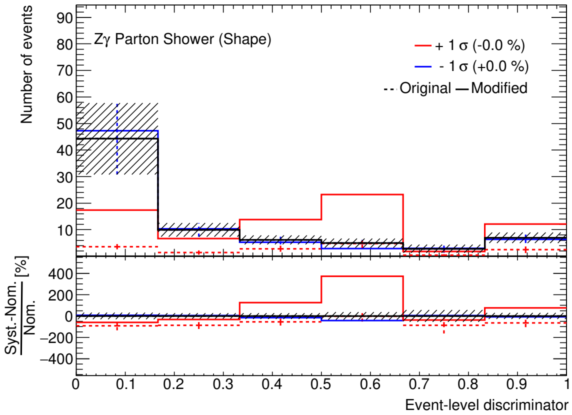


Figure 8.7.: The $Z\gamma$ parton shower NP for the dilepton channel.

8.2. Full fit results

Final fits to data are presented in this section for all channels. The signal strength measurements are presented in Section 8.2.1, followed by post-fit plots, yields, and dominant uncertainties in Section 8.2.2. This is followed by a discussion on the nuisance parameters in Section 8.2.3.

8.2.1. Signal strength

The signal strength (μ) for each fit is shown in Figure 8.8 along with the total uncertainty split into statistical and systematic components. The vertical dotted line at one represents the SM prediction of the signal strength. The shaded region represents the theoretical uncertainty derived from Equation 5.5, which includes the uncertainty on the k-factor calculation.

All single-lepton measurements indicate that we are dominated by systematic uncertainties. For dilepton channels, while the systematic uncertainty is dominant, much can still be gained from more data. There also seems to be a clear trend of signal strengths greater than one. Given that all measurements agree within the theory prediction uncertainties, nothing conclusive can be said. Future measurements need to put high priority on reducing the theoretical uncertainty. For completeness, this same figure but with the addition of all fit scenarios detailed in Chapter 5.1.3 is shown in Appendix E, where negligible differences are seen.

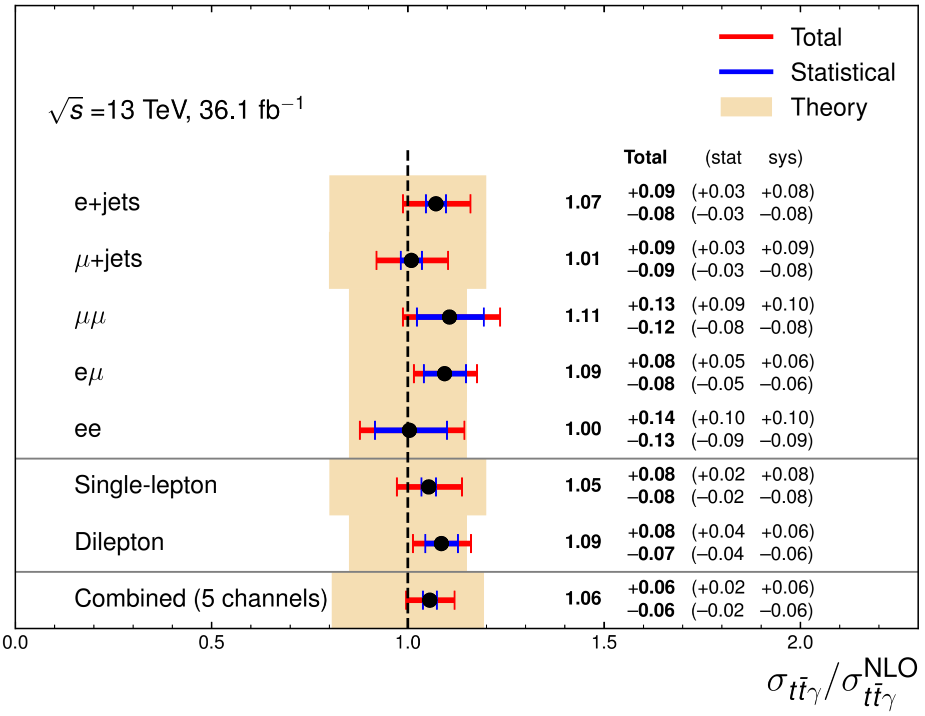


Figure 8.8.: The observed signal strength for individual, single-lepton, dilepton and combined channels. The NLO SM prediction is represented by the dashed vertical line. The theoretical uncertainty for each fit is represented by the shaded region.

For the single-lepton channels the $W\gamma$ background is a free parameter in the fit and so has a corresponding normalisation. These are:

$$\begin{aligned} W\gamma\mathrm{SF}_{e\text{+jets}} &= 0.70 \pm 0.40 \text{ (tot.)}, \\ W\gamma\mathrm{SF}_{\mu\text{+jets}} &= 0.94 \pm 0.32 \text{ (tot.)}, \\ W\gamma\mathrm{SF}_{\mathrm{SL}} &= 0.80 \pm 0.34 \text{ (tot.)}, \\ W\gamma\mathrm{SF}_{5\text{-chan}} &= 0.86 \pm 0.26 \text{ (tot.)}. \end{aligned}$$

These normalisations agree within uncertainties with the expectation value of one, as well as the validation region fit results in Chapter 5.6.

8.2.2. Distributions, yields and uncertainties

The pre-fit and post-fit ELD distributions for the single-lepton channel are shown in Figure 8.9, while the distributions for the dilepton channel are shown in Figure 8.10. Individual channel distributions are shown in Appendix E. The post-fit yields for all channels are shown in Table 8.2, which includes all uncertainties. In all bins the ratio of predicted MC and measured data agree within uncertainties. By looking at the uncertainty bands for pre- and post-fit distributions, the reduction of the size is a clear indication that the fitting mechanism is able to constrain NPs.

The extracted signal strengths and NP results from the ELDs can be applied to other distributions in the respective channels[4]. For the single-lepton and dilepton channels, a selection of distributions are shown in Figure 8.11 and Figure 8.12. For the single-lepton distributions the prediction and data agree. Similarly, this is the case for all but one of the dilepton variables. The $\Delta\phi(l,l)$ distribution is sensitive to the $t\bar{t}$ spin correlation and represents the azimuthal opening between the two leptons. This distribution translates to observing a higher spin correlation than that predicted by the SM. This is consistent with findings found in [57] for the $e\mu$ channel in $t\bar{t}$ events[5]. For further differential $t\bar{t}\gamma$ measurements of this variable one can refer to [1].

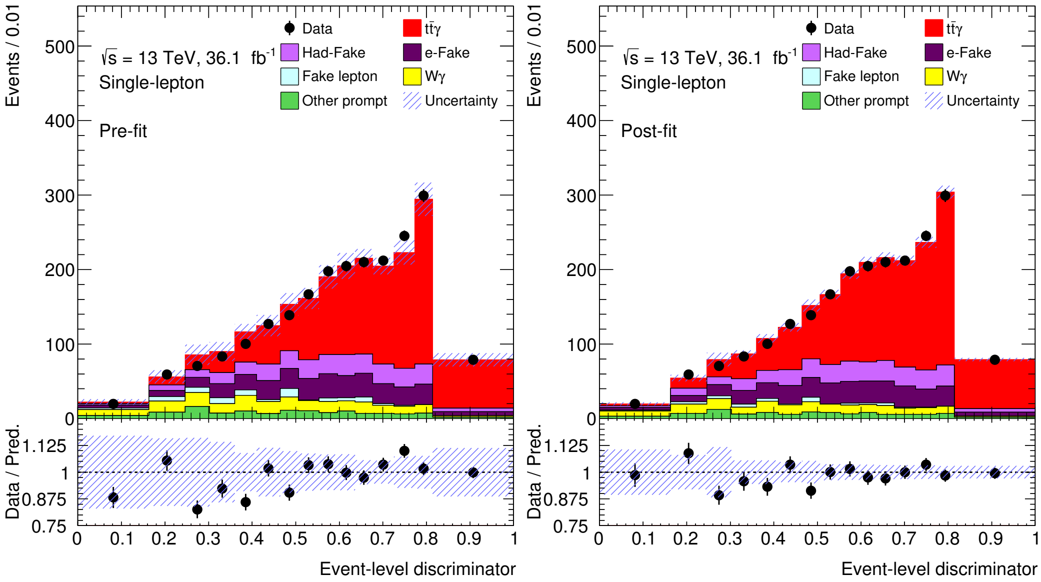
	single-lepton	dilepton	e+jets	μ+jets	ee	$\mu\mu$	$e\mu$
$t\bar{t}\gamma$	7044 ± 350	782 ± 44	3548 ± 190	3326 ± 220	147 ± 17	192 ± 19	440 ± 27
hadronic fake	1473 ± 180	49 ± 26	734.0 ± 83	736 ± 110	13 ± 7	8 ± 7	25 ± 13
$e \to \gamma$ fake	1621 ± 160	2 ± 1	939 ± 95	723 ± 74	1 ± 1	0 ± 1	1 ± 1
fake lepton	186 ± 68	-	171 ± 58	36 ± 30	-	-	-
$W\gamma$	896 ± 370	-	397 ± 210	545 ± 180	-	-	-
$Z\gamma$	-	55 ± 29	-	-	29 ± 17	31 ± 20	-
Other prompt	569 ± 180	18 ± 7	312 ± 110	304 ± 99	5 ± 2	4 ± 2	8 ± 3
Total	11787 ± 180	906 ± 38	6102 ± 140	5669 ± 110	195 ± 19	235 ± 24	474 ± 24
Data	11662	902	6002	5660	196	233	473

Table 8.2.: Post-fit yields for all single-lepton and dilepton channels. All uncertainties are included.

---

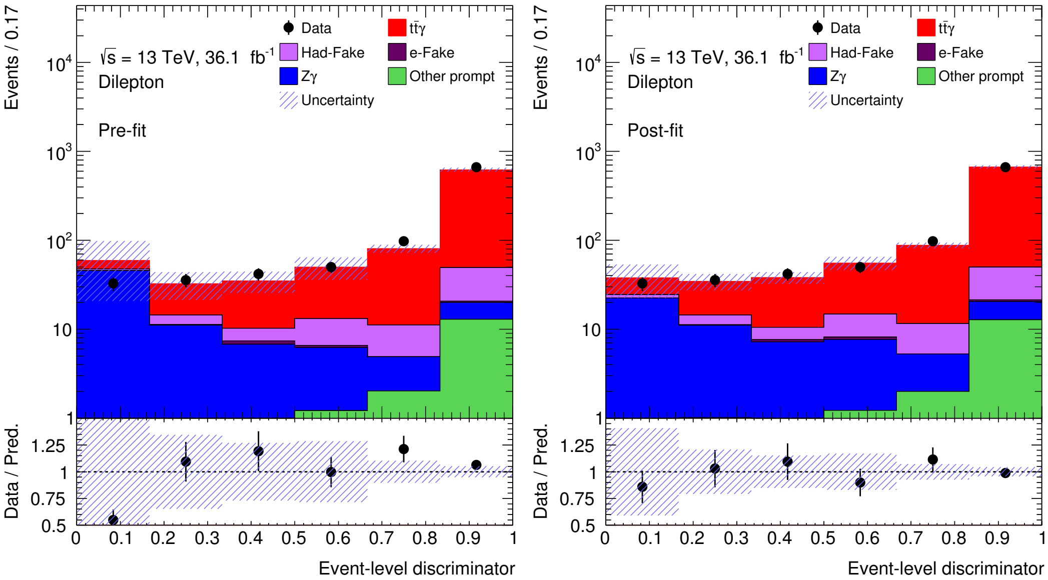
[4] Additionally, the post-fit results from the 5-channel combined fit can be applied to distributions of each channel, the advantage being that the total uncertainties will be smaller. This has been done and uncertainties still cover most of the deviations seen (the exception being the $\Delta\phi(l,l)$ distribution). This is not presented in this thesis due to the quantity of figures.

[5] Since the $t\bar{t}\gamma$ events are a subset of the $t\bar{t}$ events, this is not a new finding, but rather corroborates what the other analysis measured.



(a) Pre-fit single-lepton (b) Post-fit single-lepton

Figure 8.9.: Pre- and post-fit plots for the single-lepton channel where the ELD distribution is used as the discriminating variable in the fit.



(a) Pre-fit dilepton (b) Post-fit dilepton

Figure 8.10.: Pre- and post-fit plots for the dilepton channel where the ELD distribution is used as the discriminating variable in the fit.

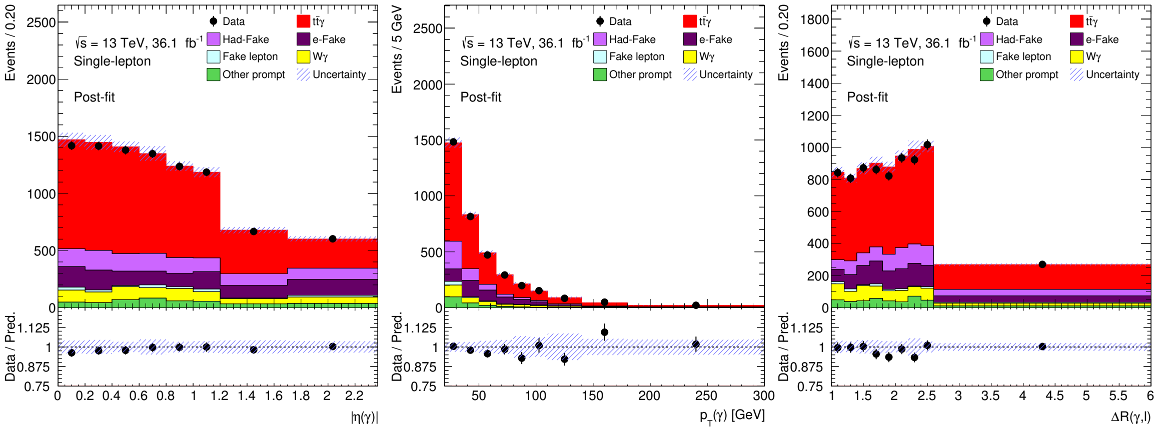


Figure 8.11.: Post-fit distributions for the single-lepton channel where the single-lepton ELD distribution is used to extract uncertainties and signal strength parameters.

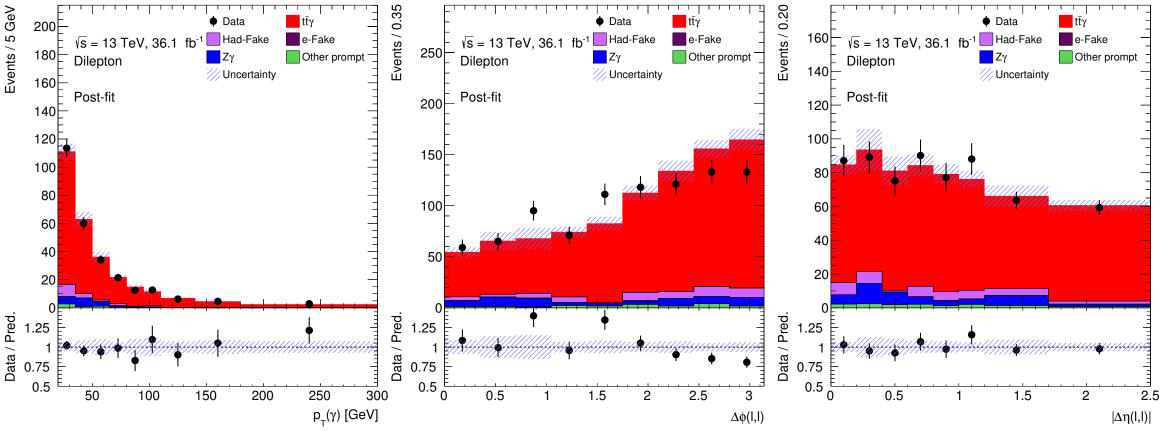


Figure 8.12.: Post-fit distributions for the dilepton channel where the dilepton ELD distribution is used to extract uncertainties and signal strength parameters.

All uncertainties that enter the fit can be grouped into appropriate categories and their impact on the fit explored. Table 8.3 shows grouped systematics and their relative contributions to the uncertainty on the signal strength, $\Delta\mu/\mu$. To calculate $\Delta\mu$, all NPs in a category are held fixed while the rest are free to float. The resulting uncertainty is then subtracted in quadrature from the initial total uncertainty. Thus, what is shown in the table is a relative uncertainty given in percentage for an "up" and "down" variation. The same table for the individual channels is shown in Appendix E.

Some systematics of interest on Table 8.3 are shaded in grey. We can see that the largest contributions come from jets, background modelling and in the case of single-lepton channels, the PPT. For the dilepton channels the signal modelling also plays a large role which is no surprise given how much signal there is. These groups can be further dissected.

- Jets: Table 8.4 further breaks down the jet contributions. The Jet Pileup RhoTopology component [148] is the largest contribution in all cases. This has already been studied within ATLAS and future analyses will make use of a reduced Jet Pileup RhoTopology systematic uncertainty.

- Background modelling: Table 8.5 categorises the background modelling according to the three major backgrounds ($e \to \gamma$ fake, hadronic fake and Other prompt) where contributions largely come from analysis specific techniques used in Chapter 5. Included separately is the $t\bar{t}$ modelling, which is essentially theoretical uncertainties that are more difficult to reduce. For the single-lepton channels a significant contribution of uncertainty in background modelling comes from the Other prompt category. This is because it includes all cross-section normalisation uncertainties for the Other prompt backgrounds, and also the uncertainty on the floating $W\gamma$ background, which is the largest contribution. The $e \to \gamma$ fake modelling is one of the dominant systematics from the single-lepton channels, which is not surprising since it is the largest background. For the dilepton channel the dominant systematic component comes from the hadronic fake background. This analysis took a generally conservative approach when deriving the hadronic fake systematics, even more so when extrapolating to the dilepton channels. For future analyses this component can certainly be reduced using different techniques.

- PPT: Table 8.6 breaks down the systematic contributions of the PPT for the single-lepton channels. As expected, the dominant component comes from the signal and background sources that have the most statistics. The PPT's systematic uncertainties need to be revised as the current approach is too conservative. An additional study was done to assess the impact the PPT has on the ELD and the potential gain it can bring for future analyses. The ELD was retrained in the identical fashion with identical parameters. The only difference being that the PPT variable was removed. As expected, the ROC curve values are lower, and the discrimination for hadronic fake events disappears. The s/b and $s/\sqrt{b}$ figures of merit (s=signal, b=sum of all backgrounds) are shown in Figure 8.13. From left to right, at each bin on the respective ELD distribution a cut is made and the two figures of merit calculated using the remaining events. Thus, the point at zero reflects how this analysis was conducted. It can be seen that our total uncertainty is slightly larger due to the inclusion of the PPT. However, the figure also shows the increase in sensitivity the PPT can bring should we cut on the ELD distribution in future analyses.

A summary of the above discussion is shown in Figure 8.14 for all fits. The relative uncertainty ($\Delta\mu/\mu$) of a few selected groups is shown along with the total, systematic and statistical contributions. This plot gives a better feeling for the size of the uncertainties discussed above. The next section discusses the stability (and thus the reliability of the results) of the fit by examining the individual NPs in more detail.

	Inclusive		Single-lepton		Dilepton	
	+ [%]	- [%]	+ [%]	- [%]	+ [%]	- [%]
Signal Modelling	2.50	2.36	1.59	1.47	2.87	2.75
Jets	3.14	2.83	5.40	4.80	2.02	1.89
Luminosity	2.28	2.06	2.32	2.01	2.29	2.04
Pileup	1.99	1.86	2.04	1.80	2.32	2.12
Photon Efficiencies	1.06	0.97	1.08	0.95	1.07	0.96
b-Tagging	0.41	0.40	0.76	0.86	0.36	0.40
Background modelling	2.81	2.74	4.80	4.80	2.91	2.86
Leptons	0.96	0.88	0.26	0.27	1.30	1.20
Prompt photon tagger (shape)	1.45	1.45	3.80	4.00	-	-
Eγ (Resolution and scale)	0.06	0.08	0.02	0.03	0.16	0.20
Template Statistics	1.50	1.38	1.89	1.78	1.65	1.56
Total systematic	5.8	5.5	7.9	7.6	5.8	5.4
Total statistical	1.4	1.4	1.5	1.5	3.8	3.8
Total	6.0	5.7	8.1	7.7	7.0	6.6

Table 8.3.: Relative effects on μ due to the grouped systematic sources. Fits are performed with nuisance parameters in each group held constant with the rest floating. This new uncertainty is subtracted in quadrature from the total uncertainty to obtain $\Delta\mu$.

Jets	Inclusive		Single-lepton		Dilepton	
	+ [%]	- [%]	+ [%]	- [%]	+ [%]	- [%]
Jet Pileup (RhoTopology)	2.31	2.10	3.80	3.38	1.66	1.52
Jet Flavour Composition	0.05	0.05	0.13	0.14	0.10	0.10
Jet EtaInterCalibration	0.13	0.12	1.10	1.01	0.21	0.18
Jet Effective NP	1.79	1.61	2.91	2.53	0.97	0.89
Jet Flavour Resp.	0.41	0.39	1.94	1.79	0.02	0.00

Table 8.4.: Breakdown of the effects that the larger jet systematic components have on μ.

Background modelling	Inclusive		Single-lepton		Dilepton	
	+ [%]	- [%]	+ [%]	- [%]	+ [%]	- [%]
$e \to \gamma$ fake modelling	1.23	1.19	2.51	2.54	0.05	0.05
$t\bar{t}$ modelling	1.10	1.06	2.02	2.05	0.98	1.13
hadronic fake modelling	1.93	1.87	1.41	1.42	2.13	2.17
Other prompt bkg modelling	1.57	1.55	3.55	3.70	1.26	1.25

Table 8.5.: Breakdown of the effects that various background modelling components have on μ. The $W\gamma$ floating normalisation uncertainty is included in "Other prompt bkg modelling".

PPT	Single-lepton	
	+ [%]	- [%]
Prompt sources	3.17	3.30
$e \to \gamma$ fake sources	1.74	1.92
hadronic fake sources	0.60	0.60
hadronic fake isolation	0.25	0.23

Table 8.6.: Breakdown of the effects that the PPT components have on μ.

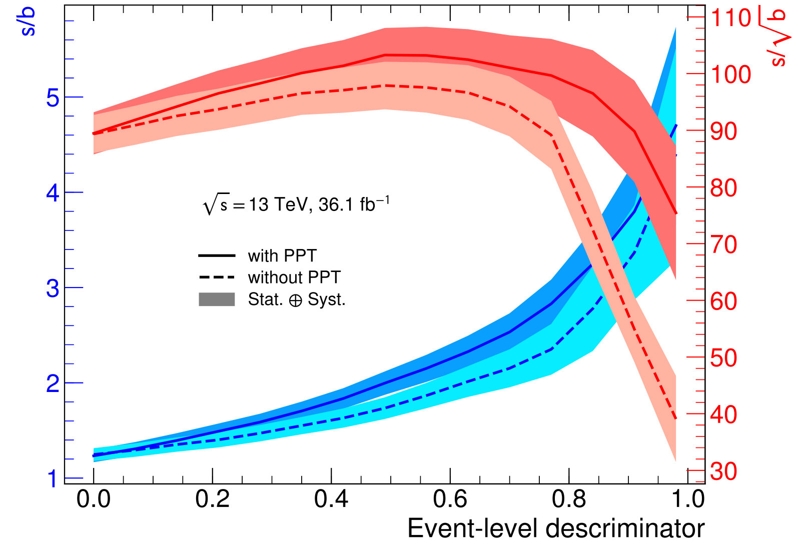


Figure 8.13.: Curves showing s/b and $s/\sqrt{b}$ for different cuts on the ELD distributions trained with and without the PPT. The point at zero on the x-axis represents how this current analysis was conducted. Total uncertainties are included in the shaded bands.

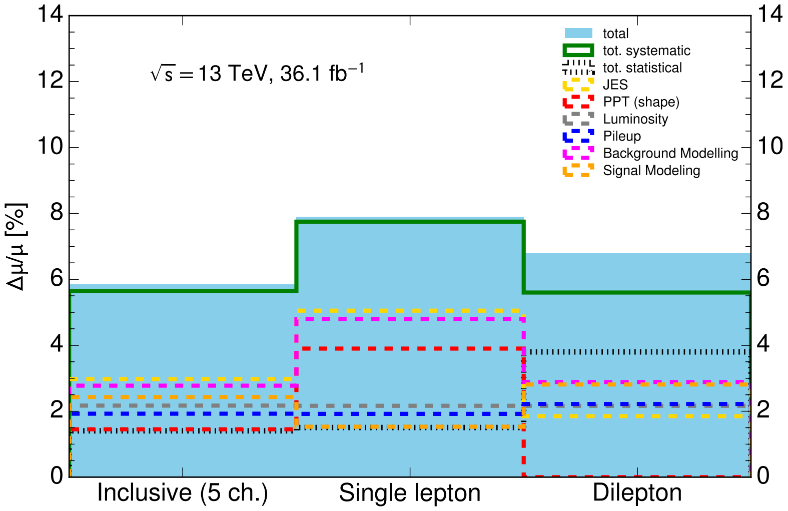


Figure 8.14.: The relative contribution of uncertainties shown in groups. A fit is performed where all NPs in a group are kept fixed while the rest are left floating. The resulting uncertainty is subtracted in quadrature from the original uncertainty to obtain $\Delta\mu$.

8.2.3. Fit cross checks

This section presents cross checks similar to Section 8.1.2 in order to ensure the fits are reliable as well as gain a deeper insight into the NPs. The major difference compared to the Asimov studies is that in a full fit the NPs and POIs are allowed to float. Figure 8.15 shows the posterior NPs in separate single-lepton channels (overlaid into one figure). Similarly, Figure 8.16 shows the same for all the dilepton channels. It is important to note that for the two above mentioned plots each of the overlaid set of posterior NPs makes up the first seven fits in Figure 8.8. The posterior NPs for the final fit on Figure 8.8 (5-channel combined) are shown in Figure 8.17.

The fits in general are stable with only a few pulls and constraints that need to be further checked. For the single-lepton channels the main pulls/constraints are for: $t\bar{t}\gamma$ parton shower, $t\bar{t}$ parton shower, $t\bar{t}$ ISR/FSR, $Z\gamma$ normalisation, fake lepton and JER. For the dilepton channels the main pulls/constraints are for: $Z\gamma$ factorisation and renormalisation scales, $Z\gamma$ parton shower, $Z\gamma$ normalisation and JER. For the most part, pulls are seen for the same NPs that are constrained in the Asimov fit. First, it is useful to see how much of a role these NPs actually play in the final fit. Figure 8.18 shows ranking plots for the top ten systematic uncertainties of the single-lepton, dilepton and 5-channel inclusive fit. The individual channels (e+jets, μ+jets, ee, $\mu\mu$ and $e\mu$) are not shown due to being very similar. The top axis shows absolute difference and applies to the pre-fit and post-fit impact on μ. Here, similar to tables presented earlier the NP in question is held fixed while the fit is performed with the remaining NPs floating. This is then subtracted in quadrature from the original uncertainty. The bottom axis applies to the black points and shows the constraints in the same way as the previous pull plots.

For the single-lepton rankings the Jet Pileup RhoTopology NP contributes the largest uncertainty. It has already been mentioned that dedicated groups within ATLAS have reduced this for future analyses. The PPT NPs have both prompt and $e \to \gamma$ fake sources within the top 5, evidence that the approach to deriving the systematics for PPT can be revisited. The dilepton channel ranking shows that the overall systematic contributions are already quite small. The highly ranked systematics have little dependence on the techniques used to analyse this channel (with the exception of one hadronic fake derived NP in the top ten). Thus, it is a more straightforward channel to work with, which offers a very clean signal (largely due to the $e\mu$ channel). With 36 fb^{-1} of data collected, the dilepton ($e\mu$) channel can already be considered the “golden channel”. For the 5-channel inclusive measurement common systematics between the five channels (or subsets of the five channels) are treated as fully correlated. This allows for larger possible constraining power. Table 8.7 summarises the pulled or constrained NPs from the single-lepton, dilepton and 5-channel combined fits and their ranked impact on the signal strength. A dash indicates the NP lies outside the largest 40 contributions. The main NPs that are constrained are not ranked particularly high, and so do not impact the fit as much.

A selection of NPs is presented based on whether or not they are pulled in the fit as well as how high they are ranked, and is shown in Figures 8.19 and 8.20 for the single-lepton and dilepton channels, respectively. These figures are different to the previous NP figures in that they show the systematic contribution compared to the full signal + background prediction, as well as data. This enables us to see why certain NPs are pulled, and in which bin(s) the mis-modelling originates from. For the single-lepton channel there are a few aspects to notice:

- Figure 8.19a and 8.19b show the two main contributions of parton shower modelling to the $e \to \gamma$ fake and hadronic fake background. More sources are included (and correlated) with these NPs, which arise from the data-driven methods and the effect varying the parton shower has when deriving the scale factors. However, since the pulls are in opposite directions for the e+jets and μ+jets channels, and well within 1σ, this is no cause for concern.

- For the $t\bar{t}\gamma$ parton shower the largest shape differences can be seen in the most signal-rich bin (Figure 8.19c). This can account for the constraint. While in the third bin from the right a fluctuation in data, also in a signal-rich bin, will cause a small pull towards $+1\sigma$. This is compounded by the normalisation part of the parton shower (Figure 8.19d), where the third bin from the right also prefers the $+1\sigma$ variation.

- Figure 8.19e shows the highest ranked NP for the single-lepton channel, which does not show any major pulls or constraints.

- For JER (Figure 8.19f), the third bin from the right includes a data fluctuation. In this bin the "up" and "down" variations (which are arbitrary) flip, and so the -1σ variation is preferred. This is indicated by the pull towards -1σ on the pull plots.

For the dilepton channel two NPs are shown.

- The $Z\gamma$ parton shower NP (Figure 8.20a) is heavily pulled and constrained in the dilepton fits. This is due to the mis-modelling between the nominal SHERPA and systematic variation `MG5_aMC` samples. The data prefers the shape provided by the systematic variation. The overall impact this and other $Z\gamma$ related NPs have on the fit are small, as can be seen from Table 8.7.

- The JER (Figure 8.20b) has a mild preference for the $+1\sigma$ variation as can be seen in the signal-rich (and generally event-rich) right side of the ELD.

Finally, Figure 8.21 shows the correlations from the maximum likelihood fit for the single-lepton and dilepton channels. Only NPs that have at least one correlation above 15% are included. For the single-lepton channel, large correlations between PPT prompt, Jet Pileup RhoTopology and the signal strength are unsurprising given their very conservative nature and thus their impact on the measurement. These are large NPs that predominantly affect the signal. For the dilepton channel there are no alarming or surprisingly large correlations. Thus, taking into account all studies presented above, we can conclude that the fit behaves as expected and is stable.

	Nuisance Parameter	Rank
single-lepton	$t\bar{t}$ Parton shower	6
	$t\bar{t}\gamma$ Parton shower	9
	JER	17
	$t\bar{t}$ ISR/FSR	18
	$Z\gamma$ Normalisation	23
	Lepfake	-
dilepton	$Z\gamma$ Normalisation	12
	JER	14
	$Z\gamma$ Parton shower	28
	$Z\gamma$ F+R scales	-
5-channel combined	Pileup	3
	$t\bar{t}\gamma$ ISR/FSR	5
	PPT prompt photons	7
	$t\bar{t}$ Parton shower	17
	$t\bar{t}\gamma$ Parton shower	18
	$t\bar{t}$ ISR/FSR	19
	Lepfake	-
	$Z\gamma$ Normalisation	-
	$Z\gamma$ Parton shower	-

Table 8.7.: The most pulled/constrained NPs in the single-lepton and dilepton fits and their ranking. A dash indicates the NP is not within the top 40 entries.

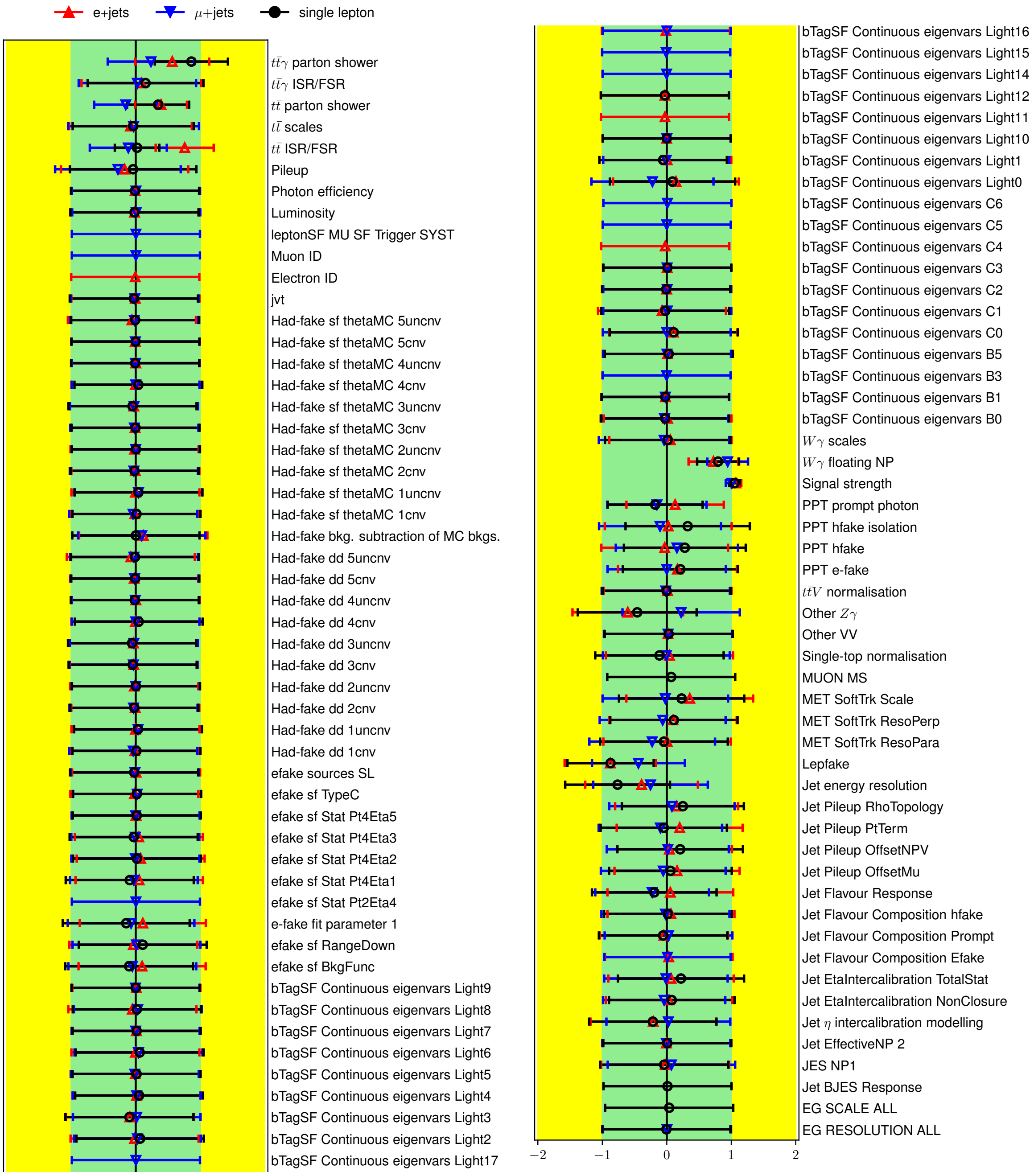


Figure 8.15.: Post-fit pull plots for all nuisance parameters in the single-lepton channels. Each set of points constitutes posteriors for a different fit. The "Signal strength" and "$W\gamma$ floating NP" both have expectation values of one. "Had-fake" and "efake" NPs relate to the data-driven methods and the uncertainties that arise from estimating these backgrounds. The "bTagSF" NPs are associated with pseudo-continuous b-tagging for different flavour jets.

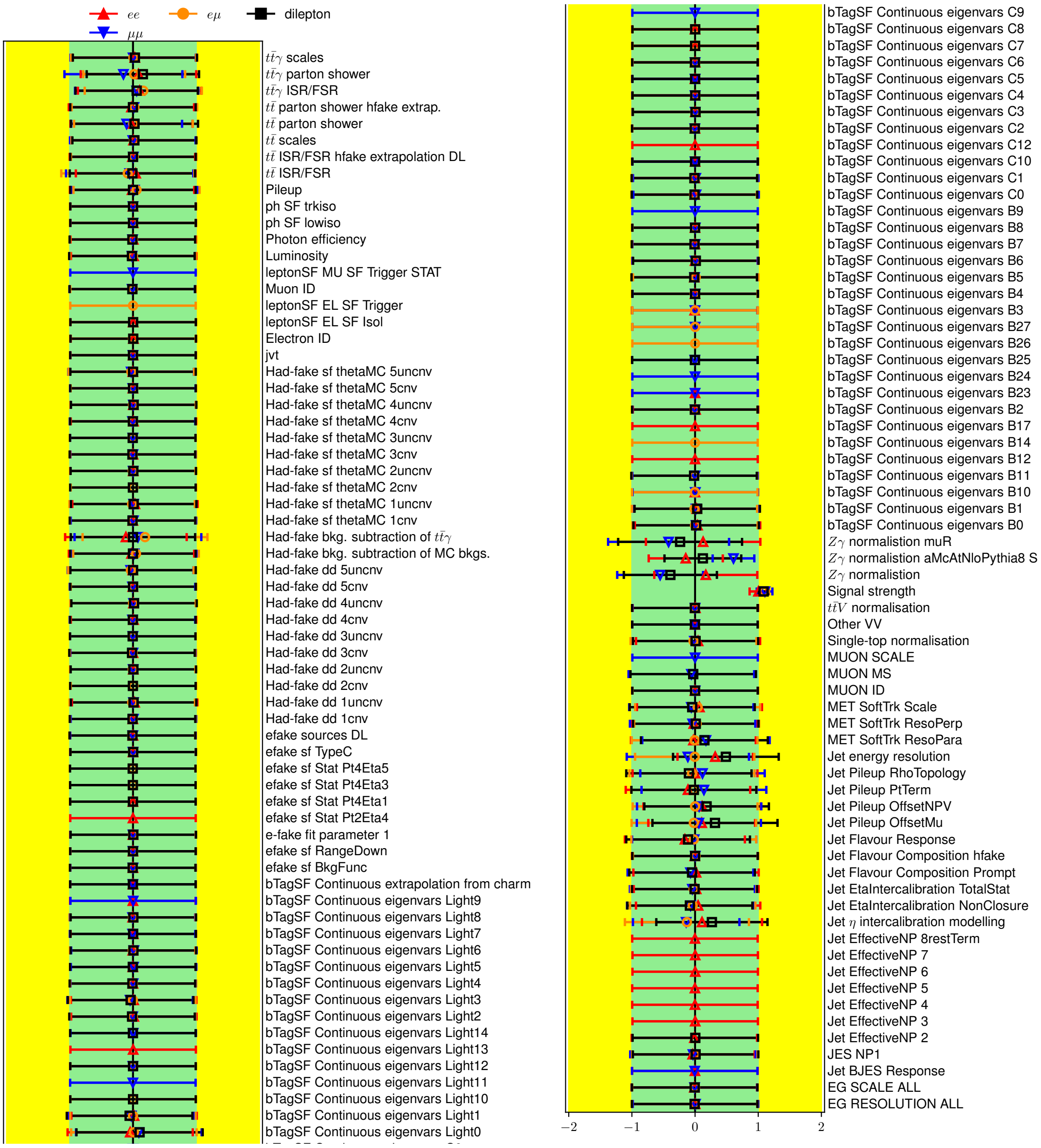


Figure 8.16.: Post-fit pull plots for all nuisance parameters in the dilepton channels. Each set of points constitutes posteriors for a different fit. The "Signal strength" has an expectation value of one. "Had-fake" and "efake" NPs relate to the data-driven methods and the uncertainties that arise from estimating these backgrounds. The "bTagSF" NPs are associated with pseudo-continuous b-tagging for different flavour jets.

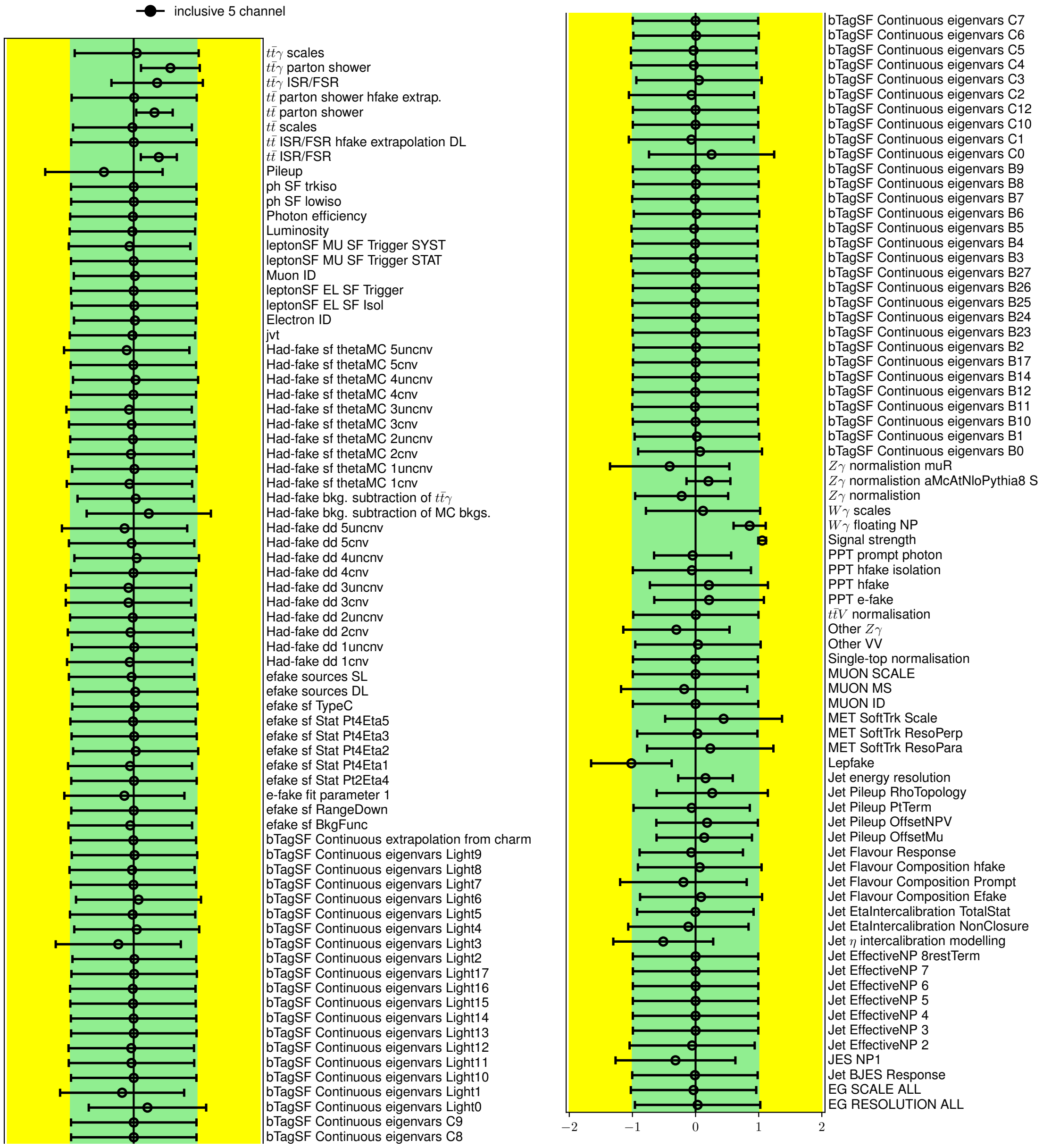
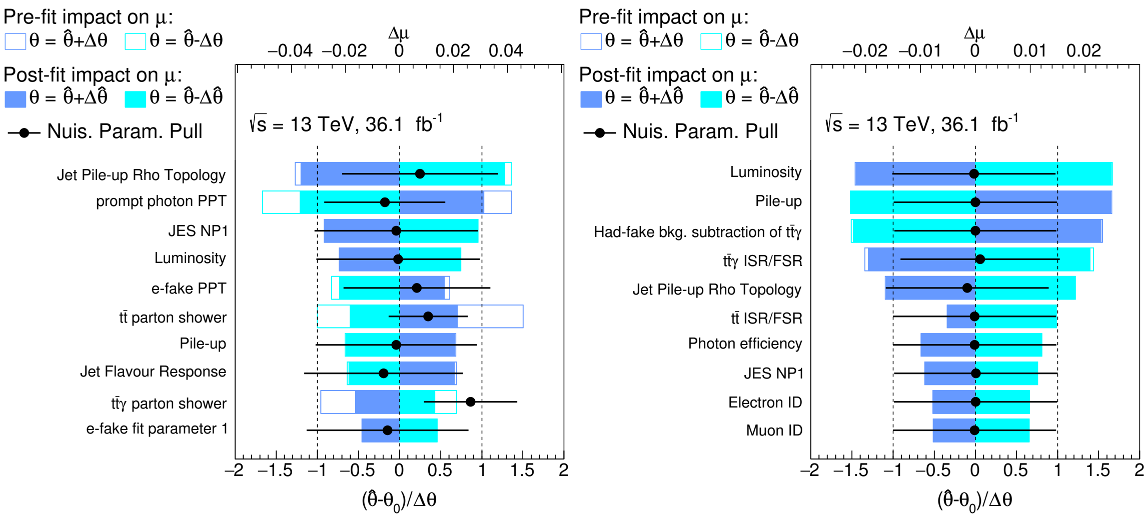
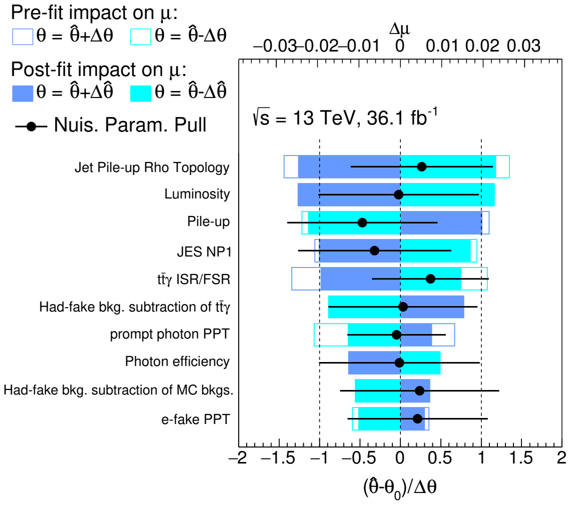


Figure 8.17.: Post-fit pull plots for all nuisance parameters in the 5-channel inclusive fit. The "Signal strength" and "$W\gamma$ floating NP" both have expectation values of one. "Had-fake" and "efake" NPs relate to the data-driven methods and the uncertainties that arise from estimating these backgrounds. The "bTagSF" NPs are associated with pseudo-continuous b-tagging for different flavour jets.



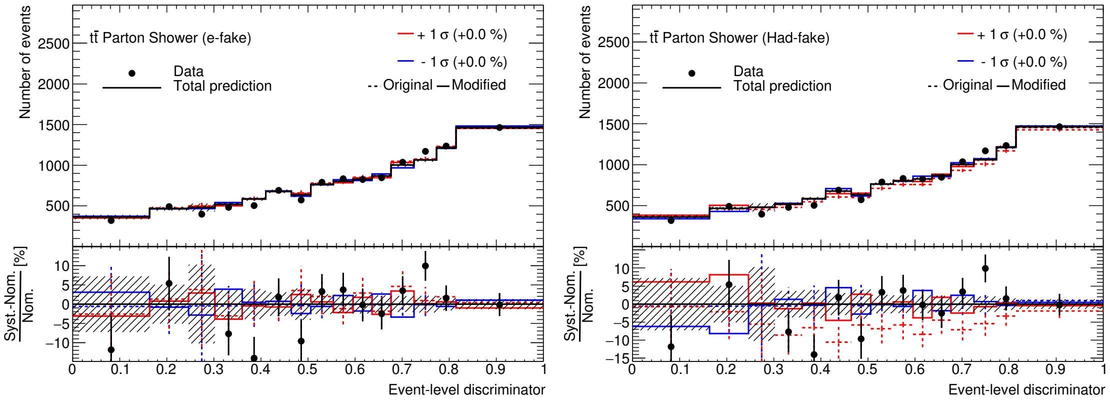
(a) single-lepton

(b) dilepton

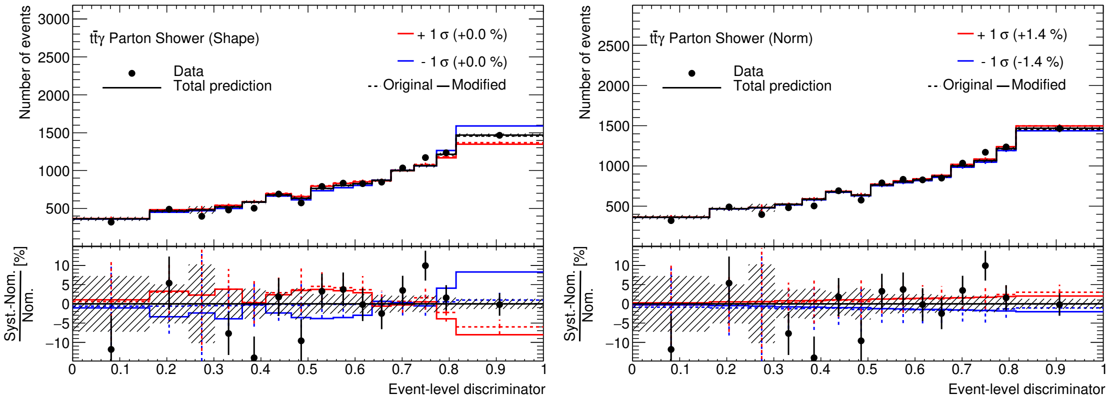


(c) 5-channel inclusive

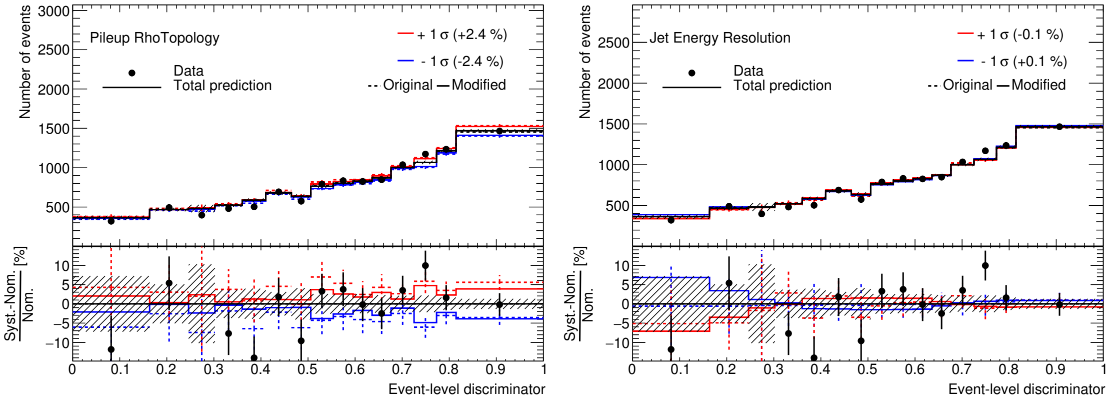
Figure 8.18.: Largest ten NPs ranked according to their impact on the single-lepton, dilepton and 5-channel inclusive fits.



(a) $t\bar{t}$ parton shower ($e \to \gamma$ fake)
(b) $t\bar{t}$ parton shower (hadronic fake)

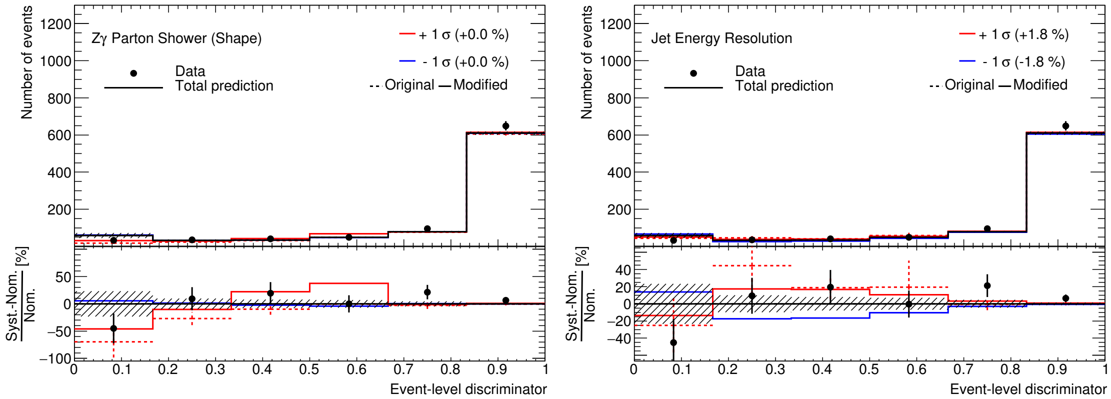


(c) $t\bar{t}\gamma$ parton shower (shape)
(d) $t\bar{t}\gamma$ parton shower (norm)



(e) Pileup RhoTopology
(f) Jet energy resolution

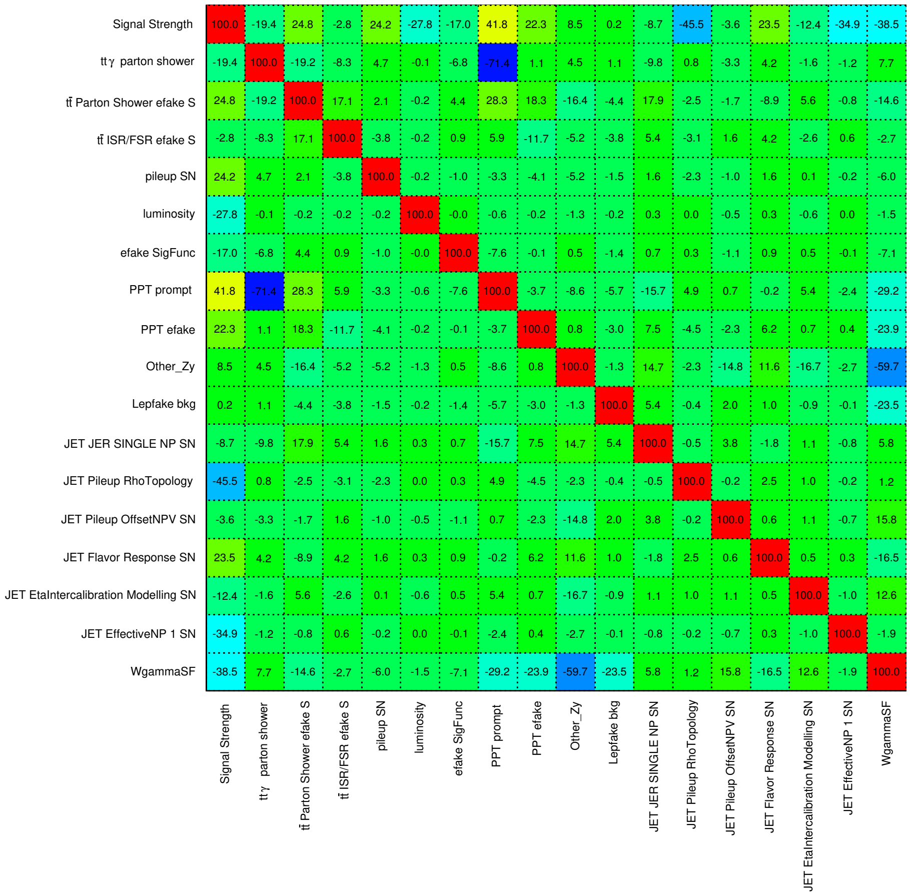
Figure 8.19.: A selection of NPs that are either pulled or constrained, or have large contributions in the final single-lepton fit.



(a) $Z\gamma$ parton shower

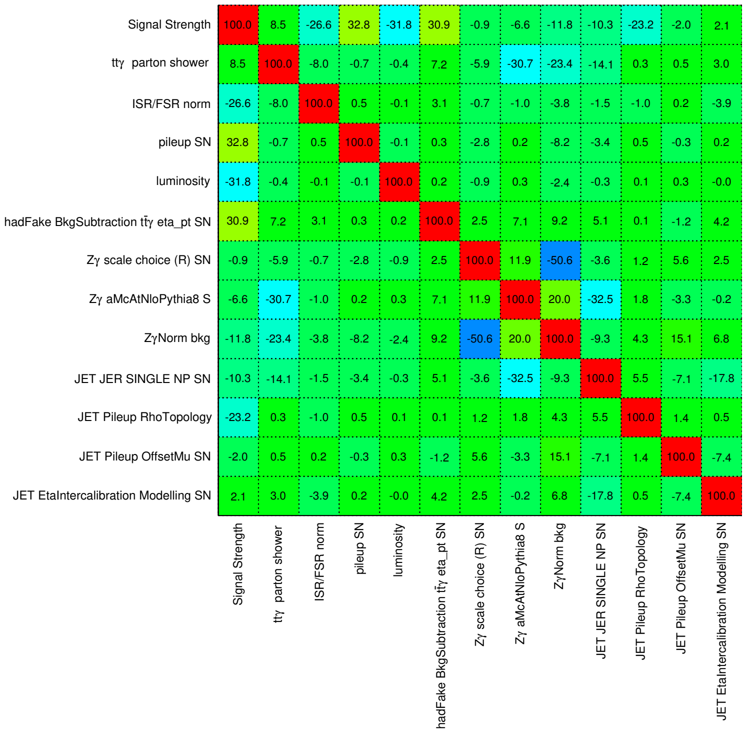
(b) Jet energy resolution

Figure 8.20.: A selection of NPs that are either pulled or constrained, or have large contributions in the final dilepton fit.



	Signal Strength	ttγ parton shower	tt̄ Parton Shower efake S	tt̄ ISR/FSR efake S	pileup SN	luminosity	efake SigFunc	PPT prompt	PPT efake	Other_Zy	Lepfake bkg	JET JER SINGLE NP SN	JET Pileup RhoTopology	JET Pileup OffsetNPV SN	JET Flavor Response SN	JET EtaIntercalibration Modelling SN	JET EffectiveNP 1 SN	WgammaSF
Signal Strength	100.0	-19.4	24.8	-2.8	24.2	-27.8	-17.0	41.8	22.3	8.5	0.2	-8.7	-45.5	-3.6	23.5	-12.4	-34.9	-38.5
ttγ parton shower	-19.4	100.0	-19.2	-8.3	4.7	-0.1	-6.8	-71.4	1.1	4.5	1.1	-9.8	0.8	-3.3	4.2	-1.6	-1.2	7.7
tt̄ Parton Shower efake S	24.8	-19.2	100.0	17.1	2.1	-0.2	4.4	28.3	18.3	-16.4	-4.4	17.9	-2.5	-1.7	-8.9	5.6	-0.8	-14.6
tt̄ ISR/FSR efake S	-2.8	-8.3	17.1	100.0	-3.8	-0.2	0.9	5.9	-11.7	-5.2	-3.8	5.4	-3.1	1.6	4.2	-2.6	0.6	-2.7
pileup SN	24.2	4.7	2.1	-3.8	100.0	-0.2	-1.0	-3.3	-4.1	-5.2	-1.5	1.6	-2.3	-1.0	1.6	0.1	-0.2	-6.0
luminosity	-27.8	-0.1	-0.2	-0.2	-0.2	100.0	-0.0	-0.6	-0.2	-1.3	-0.2	0.3	0.0	-0.5	0.3	-0.6	0.0	-1.5
efake SigFunc	-17.0	-6.8	4.4	0.9	-1.0	-0.0	100.0	-7.6	-0.1	0.5	-1.4	0.7	0.3	-1.1	0.9	0.5	-0.1	-7.1
PPT prompt	41.8	-71.4	28.3	5.9	-3.3	-0.6	-7.6	100.0	-3.7	-8.6	-5.7	-15.7	4.9	0.7	-0.2	5.4	-2.4	-29.2
PPT efake	22.3	1.1	18.3	-11.7	-4.1	-0.2	-0.1	-3.7	100.0	0.8	-3.0	7.5	-4.5	-2.3	6.2	0.7	0.4	-23.9
Other_Zy	8.5	4.5	-16.4	-5.2	-5.2	-1.3	0.5	-8.6	0.8	100.0	-1.3	14.7	-2.3	-14.8	11.6	-16.7	-2.7	-59.7
Lepfake bkg	0.2	1.1	-4.4	-3.8	-1.5	-0.2	-1.4	-5.7	-3.0	-1.3	100.0	5.4	-0.4	2.0	1.0	-0.9	-0.1	-23.5
JET JER SINGLE NP SN	-8.7	-9.8	17.9	5.4	1.6	0.3	0.7	-15.7	7.5	14.7	5.4	100.0	-0.5	3.8	-1.8	1.1	-0.8	5.8
JET Pileup RhoTopology	-45.5	0.8	-2.5	-3.1	-2.3	0.0	0.3	4.9	-4.5	-2.3	-0.4	-0.5	100.0	-0.2	2.5	1.0	-0.2	1.2
JET Pileup OffsetNPV SN	-3.6	-3.3	-1.7	1.6	-1.0	-0.5	-1.1	0.7	-2.3	-14.8	2.0	3.8	-0.2	100.0	0.6	1.1	-0.7	15.8
JET Flavor Response SN	23.5	4.2	-8.9	4.2	1.6	0.3	0.9	-0.2	6.2	11.6	1.0	-1.8	2.5	0.6	100.0	0.5	0.3	-16.5
JET EtaIntercalibration Modelling SN	-12.4	-1.6	5.6	-2.6	0.1	-0.6	0.5	5.4	0.7	-16.7	-0.9	1.1	1.0	1.1	0.5	100.0	-1.0	12.6
JET EffectiveNP 1 SN	-34.9	-1.2	-0.8	0.6	-0.2	0.0	-0.1	-2.4	0.4	-2.7	-0.1	-0.8	-0.2	-0.7	0.3	-1.0	100.0	-1.9
WgammaSF	-38.5	7.7	-14.6	-2.7	-6.0	-1.5	-7.1	-29.2	-23.9	-59.7	-23.5	5.8	1.2	15.8	-16.5	12.6	-1.9	100.0

(a) single-lepton



	Signal Strength	ttγ parton shower	ISR/FSR norm	pileup SN	luminosity	hadFake BkgSubtraction tt̄γ eta_pt SN	Zγ scale choice (R) SN	Zγ aMcAtNloPythia8 S	ZγNorm bkg	JET JER SINGLE NP SN	JET Pileup RhoTopology	JET Pileup OffsetMu SN	JET EtaIntercalibration Modelling SN
Signal Strength	100.0	8.5	-26.6	32.8	-31.8	30.9	-0.9	-6.6	-11.8	-10.3	-23.2	-2.0	2.1
ttγ parton shower	8.5	100.0	-8.0	-0.7	-0.4	7.2	-5.9	-30.7	-23.4	-14.1	0.3	0.5	3.0
ISR/FSR norm	-26.6	-8.0	100.0	0.5	-0.1	3.1	-0.7	-1.0	-3.8	-1.5	-1.0	0.2	-3.9
pileup SN	32.8	-0.7	0.5	100.0	-0.1	0.3	-2.8	0.2	-8.2	-3.4	0.5	-0.3	0.2
luminosity	-31.8	-0.4	-0.1	-0.1	100.0	0.2	-0.9	0.3	-2.4	-0.3	0.1	0.3	-0.0
hadFake BkgSubtraction tt̄γ eta_pt SN	30.9	7.2	3.1	0.3	0.2	100.0	2.5	7.1	9.2	5.1	0.1	-1.2	4.2
Zγ scale choice (R) SN	-0.9	-5.9	-0.7	-2.8	-0.9	2.5	100.0	11.9	-50.6	-3.6	1.2	5.6	2.5
Zγ aMcAtNloPythia8 S	-6.6	-30.7	-1.0	0.2	0.3	7.1	11.9	100.0	20.0	-32.5	1.8	-3.3	-0.2
ZγNorm bkg	-11.8	-23.4	-3.8	-8.2	-2.4	9.2	-50.6	20.0	100.0	-9.3	4.3	15.1	6.8
JET JER SINGLE NP SN	-10.3	-14.1	-1.5	-3.4	-0.3	5.1	-3.6	-32.5	-9.3	100.0	5.5	-7.1	-17.8
JET Pileup RhoTopology	-23.2	0.3	-1.0	0.5	0.1	0.1	1.2	1.8	4.3	5.5	100.0	1.4	0.5
JET Pileup OffsetMu SN	-2.0	0.5	0.2	-0.3	0.3	-1.2	5.6	-3.3	15.1	-7.1	1.4	100.0	-7.4
JET EtaIntercalibration Modelling SN	2.1	3.0	-3.9	0.2	-0.0	4.2	2.5	-0.2	6.8	-17.8	0.5	-7.4	100.0

(b) dilepton

Figure 8.21.: Correlations for the single-lepton and dilepton channel for each nuisance parameter and the signal strength. Only NPs with at least one correlation above 15% are shown.

8.3. Fiducial cross sections

The fiducial cross sections can be calculated using the theoretical calculations from Chapter 5.1.1 and the results in Section 8.2. These are summarised in Table 8.8, which also show theoretical predictions. All measurements agree with the predictions within uncertainty.

	σ_{fid} [fb]	Statistical		Systematic		Theory [fb]
		+ [fb]	- [fb]	+ [fb]	- [fb]	
e+jets	265	6	6	21	20	247 ± 49
μ+jets	250	7	7	22	21	248 ± 50
ee	16	2	1	2	1	16 ± 2
$\mu\mu$	18	1	1	2	1	16 ± 2
$e\mu$	34	2	2	2	2	31 ± 5
single-lepton	521	9	9	41	39	495 ± 99
dilepton	69	3	3	4	4	63 ± 9
Inclusive (5 channels)	589	10	10	34	32	558 ± 110

Table 8.8.: Fiducial cross-section summary for all channels with uncertainty split into statistical and systematic components. The theoretical prediction is included in the last column.

CHAPTER 9

Conclusions

This thesis presents fiducial cross-section measurements of the $pp \to t\bar{t}\gamma$ process using 36.1 fb^{-1} of data collected with the ATLAS detector at $\sqrt{s} = 13$ TeV. The data were collected during the LHC running periods of 2015 and 2016. In total, eight cross sections are measured including the individual single-lepton and dilepton channels (e+jets, μ+jets, ee, $\mu\mu$ and $e\mu$), the merged single-lepton and dilepton channels, and a 5-channel inclusive cross section, where no distinction between leptons and jets from tau decays are made. This thesis presents the first measurements of dilepton, and single- and dilepton inclusive cross sections.

The strategy for this analysis deviated significantly from the previous results at 7 and 8 TeV, both in ATLAS and CMS. Previously, the discriminating variable was the track isolation of the photon. This analysis places a cut on this variable, which removes a significant number of background events. Then, two separate neural network algorithms trained using Keras and applied using LWTNN are introduced to further separate signal from background. These tools present a paradigm shift for how machine learning is typically carried out at the LHC. The first neural network, called Prompt Photon Tagger, makes use of energy deposits in the calorimeter to help discriminate prompt photons from those coming from hadron decays or from hadrons misidentified as photons. The second network (the Event-level Discriminator) makes use of a range of variables related to the kinematics and topology of an event. This includes information about b-tagging, $E_{\mathrm{T}}^{\mathrm{miss}}$, number of jets etc. In the case of the single-lepton channels the Prompt Photon Tagger serves as a powerful input to the Event-level Discriminator.

The various backgrounds fall into four main classes; $e \to \gamma$ fakes, hadronic fakes, prompt photons and fake leptons. The $e \to \gamma$ fake background is dominant in the single-lepton channels and consists of events where electrons have been misidentified as photons. The main contribution comes from the $t\bar{t}$ dileptonic decays in the ee and $e\mu$ channels. This background is negligible in the dilepton channels. The estimation of this background follows a data-driven approach in which a fake photon enriched control region is created and compared to another control region of similar phase space but with the photon required to be an electron. Thus, the fake photon can be probed, efficiencies calculated and correction factors derived.

Hadrons, or photons from hadron decays that are misidentified as prompt photons, form the hadronic fake background. This background is mainly seen in the single-lepton channels, but also a small contribution exists in the dilepton channels. This background is estimated following a data-driven approach where three orthogonal control regions to the signal region are defined. Isolation and identification algorithms in these control regions are reversed. Correction factors to the fake photon contribution can then be derived.

The fake lepton background is predominantly found in the single-lepton channels with a negligible contribution in the dilepton channels. It is estimated using a purely data-driven approach in which looser criteria on the leptons are required, thus the sample is assumed to contain mainly fake leptons. Weights can be derived and are applied to data.

Prompt photons that are not from top quarks contribute towards a significant portion of the total background for the single-lepton and dilepton channels. In the single-lepton channel the dominant process is from $W\gamma$, while in the dilepton channels this contribution is from $Z\gamma$. Small contributions arise from single top, diboson and $t\bar{t}V$ processes. The estimation of these backgrounds is based on MC prediction.

A profile maximum likelihood fit is performed on the ELD distribution to extract the signal strength, and thus the final cross sections. In the single-lepton channels the $W\gamma$ background enters as a free-floating parameter. Dedicated k-factors are used in the respective channels to scale the leading-order predictions to next-to-leading order.

A summary of all measured cross sections is shown in Table 9.1. This includes the five individual channels, the two combined channels, and the inclusive measurement. Theoretical cross sections are also included. All measurements agree with theoretical predictions.

	$\sigma_{\text{fid}}^{t\bar{t}\gamma}$ [fb]	
Channel	Measured ±(stat.) ±(syst.)	Theory ±(total)
e+jets	$265 \pm 6 \pm 21$	247 ± 49
μ+jets	$250 \pm 7 \pm 22$	248 ± 50
ee	$16 \pm 2 \pm 2$	16 ± 2
$\mu\mu$	$18 \pm 1 \pm 2$	16 ± 2
$e\mu$	$34 \pm 2 \pm 2$	31 ± 5
Single-lepton	$521 \pm 9 \pm 41$	495 ± 99
Dilepton	$69 \pm 3 \pm 4$	63 ± 9
Inclusive (5 channels)	$589 \pm 10 \pm 34$	558 ± 110

Table 9.1.: Fiducial cross-section summary for all channels, as well as theoretical predictions.

For the single-lepton channels the largest systematic contributions come from the modelling of jets, background estimation techniques and the PPT. For the dilepton channel more statistics will help increase the sensitivity, as well as reducing signal and background modelling systematics. To reduce background modelling one could focus on just the $e\mu$ channel, which is dominated by signal with very little (to almost negligible) background contributions.

Further studies on the $t\bar{t}\gamma$ process should focus on reducing the theoretical uncertainties associated with the k-factor, developing techniques to reduce the contamination from photons being radiated from the top quark decay products, and EFT interpretations. The relationship between the cross section and the $t\gamma$ coupling (Q_t) is expected to be quadratic. Thus, one method to determine the $t\gamma$ coupling could be achieved by performing template fits with various hypotheses of Q_t. This will further help to improve our understanding of the electromagnetic coupling to the top quark.

Bibliography

[1] ATLAS Collaboration, *Measurements of inclusive and differential cross-sections of $t\bar{t}\gamma$ production in leptonic final states in a fiducial volume at $\sqrt{s} = 13$ TeV in ATLAS*, ATLAS-CONF-2018-048, 2018, `http://cds.cern.ch/record/2639675`.

[2] P. A. M. Dirac, *Quantum theory of emission and absorption of radiation*, Proc. Roy. Soc. Lond. **A114** (1927) 243.

[3] E. Fermi, *Quantum Theory of Radiation*, Rev. Mod. Phys. **4** (1932) 87–132.

[4] R. P. Feynman, *Space-time approach to nonrelativistic quantum mechanics*, Rev. Mod. Phys. **20** (1948) 367–387.

[5] J. S. Schwinger, *On Quantum electrodynamics and the magnetic moment of the electron*, Phys. Rev. **73** (1948) 416–417.

[6] J. S. Schwinger, *Quantum electrodynamics. I A covariant formulation*, Phys. Rev. **74** (1948) 1439.

[7] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579–588.

[8] A. Salam and J. C. Ward, *Electromagnetic and weak interactions*, Phys. Lett. **13** (1964) 168–171.

[9] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.

[10] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. **D2** (1970) 1285–1292.

[11] H. D. Politzer, *Asymptotic Freedom: An Approach to Strong Interactions*, Phys. Rept. **14** (1974) 129–180.

[12] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973) 1346–1349.

[13] D. J. Gross and F. Wilczek, *Asymptotically Free Gauge Theories - I*, Phys. Rev. **D8** (1973) 3633–3652.

[14] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323.

[15] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509.

[16] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1.

[17] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30–61.

[18] K. G. Wilson, *Confinement of Quarks*, Phys. Rev. **D10** (1974) 2445–2459.

[19] LHCb Collaboration, *Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays*, Phys. Rev. Lett. **115** (2015) 072001.

[20] D0 Collaboration, *Evidence for a $B_s^0\pi^\pm$ state*, Phys. Rev. Lett. **117** (2016) 022003.

[21] Super-Kamiokande Collaboration, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562–1567.

[22] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531–533.

[23] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652–657.

[24] Particle Data Group Collaboration, M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. **D98** (2018) 030001.

[25] V. C. Rubin and W. K. Ford, Jr., *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, Astrophys. J. **159** (1970) 379–403.

[26] M. Bartelmann and P. Schneider, *Weak gravitational lensing*, Phys. Rept. **340** (2001) 291–472.

[27] D. Clowe, et al., *A direct empirical proof of the existence of dark matter*, Astrophys. J. **648** (2006) L109–L113.

[28] A. Quadt, *Top quark physics at hadron colliders*, Eur. Phys. J. **C48** (2006) 835–1000.

[29] S. W. Herb et al., *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*, Phys. Rev. Lett. **39** (1977) 252–255.

[30] D0 Collaboration, *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632–2637.

[31] CDF Collaboration, F. Abe et al., *Observation of top quark production in $\bar{p}p$ collisions*, Phys. Rev. Lett. **74** (1995) 2626–2631.

[32] S. Mandelstam, *Determination of the pion - nucleon scattering amplitude from dispersion relations and unitarity. General theory*, Phys. Rev. **112** (1958) 1344–1360.

[33] D0 Collaboration, *Observation of Single Top Quark Production*, Phys. Rev. Lett. **103** (2009) 092001.

[34] CDF Collaboration, *First Observation of Electroweak Single Top Quark Production*, Phys. Rev. Lett. **103** (2009) 092002.

[35] J. C. Collins, D. E. Soper, and G. F. Sterman, *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 1–91.

[36] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. **B126** (1977) 298–318.

[37] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438–450.

[38] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and* e^+e^- *Annihilation by Perturbation Theory in Quantum Chromodynamics*, Sov. Phys. JETP **46** (1977) 641–653.

[39] M. Cacciari, et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710** (2012) 612–622.

[40] ATLAS Collaboration, *Summary plots from the ATLAS Top physics group*, 26th September 2018.
`https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots`.

[41] CDF and D0 Collaboration, *Combination of CDF and D0 results on the mass of the top quark using up to* 9.7 *fb*$^{-1}$ *at the Tevatron*, `arXiv:1608.01881 [hep-ex]`.

[42] ATLAS Collaboration, *Measurement of the top quark mass in the* $t\bar{t} \to$ *dilepton channel from* $\sqrt{s} = 8$ *TeV ATLAS data*, Phys. Lett. B **761** (2016) 350.

[43] CMS Collaboration, *Measurement of the top quark mass using proton-proton data at* $\sqrt{s}$ *= 7 and 8 TeV*, Phys. Rev. **D93** (2016) 072004.

[44] CMS Collaboration, *Measurement of the top quark mass using single top quark events in proton-proton collisions at* $\sqrt{s} = 8$ *TeV*, Eur. Phys. J. **C77** (2017) 354.

[45] D0 Collaboration, *Measurement of the Inclusive* $t\bar{t}$ *Production Cross Section in* $p\bar{p}$ *Collisions at* $\sqrt{s} = 1.96$ *TeV and Determination of the Top Quark Pole Mass*, Phys. Rev. **D94** (2016) 092004.

[46] ATLAS Collaboration, *Measurement of the* $t\bar{t}$ *production cross-section using* $e\mu$ *events with b-tagged jets in pp collisions at* $\sqrt{s} = 7$ *and* 8 *TeV with the ATLAS detector*, Eur. Phys. J. C **74** (2014) 3109.

[47] ATLAS Collaboration, *Determination of the top-quark pole mass using* $t\bar{t} + 1$*-jet events collected with the ATLAS experiment in* 7 *TeV pp collisions*, JHEP **10** (2015) 121.

[48] ATLAS Collaboration, *Measurement of lepton differential distributions and the top quark mass in* $t\bar{t}$ *production in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 804.

[49] CMS Collaboration, *Measurement of the* $t\bar{t}$ *production cross section in the* $e\mu$ *channel in proton-proton collisions at* $\sqrt{s}$ *= 7 and 8 TeV*, JHEP **08** (2016) 029.

[50] CMS Collaboration, *Measurement of the* $t\bar{t}$ *production cross section using events with one lepton and at least one jet in pp collisions at* $\sqrt{s}$ *= 13 TeV*, JHEP **09** (2017) 051.

[51] D. Chang, W. Chang, and E. Ma, *Alternative interpretation of the Tevatron top events*, Phys. Rev. **D59** (1999) 091503.

[52] DØ Collaboration, *Experimental discrimination between charge 2e/3 top quark and charge 4e/3 exotic quark production scenarios*, Phys. Rev. Lett. **98** (2007) 041801.

[53] ATLAS Collaboration, *Measurement of the top quark charge in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **11** (2013) 031.

[54] DØ Collaboration, *Evidence for spin correlation in $t\bar{t}$ production*, Phys. Rev. Lett. **108** (2012) 032004.

[55] ATLAS Collaboration, *Observation of spin correlation in $t\bar{t}$ events from pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Phys. Rev. Lett. **108** (2012) 212001.

[56] ATLAS Collaboration, *Measurements of spin correlation in top–antitop quark events from proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Phys. Rev. D **90** (2014) 112016.

[57] ATLAS Collaboration, *Measurements of top-quark pair spin correlations in the $e\mu$ channel at $\sqrt{s} = 13$ TeV using pp collisions in the ATLAS detector*, ATLAS-CONF-2018-027, 2018, `https://cds.cern.ch/record/2628770`.

[58] CDF Collaboration, *Measurement of the top quark forward-backward production asymmetry and its dependence on event kinematic properties*, Phys. Rev. **D87** (2013) 092002.

[59] ATLAS Collaboration, *Measurement of the charge asymmetry in top-quark pair production in the lepton-plus-jets final state in pp collision data at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. C **76** (2016) 87.

[60] ATLAS Collaboration, *Measurements of the charge asymmetry in top-quark pair production in the dilepton final state at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Phys. Rev. D **94** (2016) 032006.

[61] ATLAS Collaboration, *Measurement of the charge asymmetry in highly boosted top-quark pair production in $\sqrt{s} = 8$ TeV pp collision data collected by the ATLAS experiment*, Phys. Lett. B **756** (2016) 52.

[62] ATLAS and CMS Collaborations, *Combination of inclusive and differential $t\bar{t}$ charge asymmetry measurements using ATLAS and CMS data at $\sqrt{s} = 7$ TeV and 8 TeV*, JHEP **04** (2018) 033.

[63] CMS Collaboration, *Inclusive and differential measurements of the $t\bar{t}$ charge asymmetry in pp collisions at $\sqrt{s} = 8$ TeV*, Phys. Lett. **B757** (2016) 154–179.

[64] CMS Collaboration, *Measurement of the charge asymmetry in top quark pair production in pp collisions at $\sqrt{(s)} = 8$ TeV using a template method*, Phys. Rev. **D93** (2016) 034014.

[65] CMS Collaboration, *Measurements of $t\bar{t}$ charge asymmetry using dilepton final states in pp collisions at $\sqrt{s} = 8$ TeV*, Phys. Lett. **B760** (2016) 365–386.

[66] J. A. Aguilar-Saavedra, et al., *Shedding light on the $t\bar{t}$ asymmetry: the photon handle*, JHEP **04** (2014) 188.

[67] U. Baur, et al., *Probing electroweak top quark couplings at hadron colliders*, Phys. Rev. **D71** (2005) 054013.

[68] A. O. Bouzas and F. Larios, *Electromagnetic dipole moments of the Top quark*, Phys. Rev. **D87** (2013) 074015.

[69] M. Schulze and Y. Soreq, *Pinning down electroweak dipole operators of the top quark*, Eur. Phys. J. **C76** (2016) 466.

[70] O. Bessidskaia Bylund, et al., *Probing top quark neutral couplings in the Standard Model Effective Field Theory at NLO in QCD*, JHEP **05** (2016) 052.

[71] CDF Collaboration, *Evidence for $t\bar{t}\gamma$ Production and Measurement of $\sigma_{t\bar{t}\gamma}/\sigma_{t\bar{t}}$*, Phys. Rev. **D84** (2011) 031104.

[72] ATLAS Collaboration, *Observation of top-quark pair production in association with a photon and measurement of the $t\bar{t}\gamma$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Phys. Rev. **D91** (2015) 072007.

[73] CMS Collaboration, *Measurement of the semileptonic* $\mathrm{t\bar{t}}$ + *production cross section in pp collisions at $\sqrt{s} = 8$ TeV*, JHEP **10** (2017) 006.

[74] ATLAS Collaboration, *Measurement of the $t\bar{t}\gamma$ production cross section in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **11** (2017) 086.

[75] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444–450.

[76] K. Albertsson et al., *Machine Learning in High Energy Physics Community White Paper*, `arXiv:1807.02876 [physics.comp-ph]`.

[77] D. Guest, K. Cranmer, and D. Whiteson, *Deep Learning and its Application to LHC Physics*, `arXiv:1806.11484 [hep-ex]`.

[78] M. Paganini, L. de Oliveira, and B. Nachman, *Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters*, Phys. Rev. Lett. **120** (2018) 042003.

[79] I. J. Goodfellow, et al., *Generative Adversarial Networks*, `arXiv:1406.2661 [stat.ML]`.

[80] C. Shimmin, et al., *Decorrelated Jet Substructure Tagging using Adversarial Neural Networks*, Phys. Rev. **D96** (2017) 074034.

[81] G. Louppe, et al., *QCD-Aware Recursive Neural Networks for Jet Physics*, `arXiv:1702.00748 [hep-ph]`.

[82] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, *Deep learning in color: towards automated quark/gluon jet discrimination*, JHEP **01** (2017) 110.

[83] M. Kagan, et al., *Boosted Jet Tagging with Jet-Images and Deep Neural Networks*, EPJ Web Conf. **127** (2016) 00009.

[84] J. Barnard, et al., *Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks*, Phys. Rev. **D95** (2017) 014018.

[85] L. de Oliveira, et al., *Jet-images deep learning edition*, JHEP **07** (2016) 069.

[86] M. Stoye, et al., *DeepJet: Generic physics object based jet multiclass classification for LHC experiments*, in *Proceedings of the Deep Learning for Physical Sciences Workshop at NIPS (2017)*. 2017.
`https://dl4physicalsciences.github.io/files/nips_dlps_2017_10.pdf`.

[87] D. Guest, et al., *Jet flavor classification in high-energy physics with deep neural networks*, Phys. Rev. D **94** (2016) 112002.

[88] ATLAS Collaboration, *Identification of Hadronically-Decaying W Bosons and Top Quarks Using High-Level Features as Input to Boosted Decision Trees and Deep Neural Networks in ATLAS at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2017-004, 2017, `https://cds.cern.ch/record/2259646`.

[89] G. Louppe, M. Kagan, and K. Cranmer, *Learning to Pivot with Adversarial Networks*, `arXiv:1611.01046 [stat.ME]`.

[90] M. Wielgosz, et al., *The model of an anomaly detector for HiLumi LHC magnets based on Recurrent Neural Networks and adaptive quantization*, Eng. Appl. Artif. Intell. **74** (2018) 166–185.

[91] G. E. Hinton, S. Osindero, and Y.-W. Teh, *A Fast Learning Algorithm for Deep Belief Nets*, Neural Computation **18** (2006) 1527–1554, PMID: 16764513.

[92] Y. Bengio, et al., *Greedy Layer-Wise Training of Deep Networks*, `http://www.iro.umontreal.ca/~lisa/pointeurs/BengioNips2006All.pdf`.

[93] E. M. Metodiev, B. Nachman, and J. Thaler, *Classification without labels: Learning from mixed samples in high energy physics*, JHEP **10** (2017) 174.

[94] L. M. Dery, et al., *Weakly Supervised Classification in High Energy Physics*, JHEP **05** (2017) 145.

[95] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning representations by back-propagating errors*, Nature **323** (1986) 533 – 536.

[96] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, ArXiv e-prints (2015).

[97] N. Srivastava, et al., *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, Journal of Machine Learning Research **15** (2014) 1929-1958.

[98] M. Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015. `https://www.tensorflow.org/`.

[99] Theano Development Team, *Theano: A Python framework for fast computation of mathematical expressions*, ArXiv e-prints (2016).

[100] F. Chollet et al., *Keras*, `https://keras.io`, 2015.

[101] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81–86.

[102] A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis*, `arXiv:physics/0703039 [physics.data-an]`.

[103] D. H. Guest et al., *lwtnn: Version 2.3*, July, 2017. `https://doi.org/10.5281/zenodo.836375`.

[104] Gargamelle Neutrino Collaboration, *Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment*, Phys. Lett. **B46** (1973) 138–140.

[105] UA1 Collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at* $\sqrt{s} = 540$ *GeV*, Phys. Lett. **B122** (1983) 103–116.

[106] UA1 Collaboration, *Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/*c^2 *at the CERN SPS Collider*, Phys. Lett. **B126** (1983) 398–410.

[107] J. H. Christenson, et al., *Evidence for the* 2π *Decay of the* K_2^0 *Meson*, Phys. Rev. Lett. **13** (1964) 138–140.

[108] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[109] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[110] CMS Collaboration, *The CMS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08004.

[111] ALICE Collaboration, *The ALICE experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08002.

[112] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[113] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, Tech. Rep. CERN-LHCC-2010-013. ATLAS-TDR-19, Sep, 2010. `https://cds.cern.ch/record/1291633`.

[114] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*. 1996. `http://cds.cern.ch/record/331061`.

[115] ATLAS Collaboration, *Luminosity Public Results*, 2018. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2`.

[116] J. Pumplin, et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012.

[117] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040.

[118] H.-L. Lai, et al., *New parton distributions for collider physics*, Phys.Rev. **D82** (2010) 074024.

[119] J. Alwall, et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079.

[120] T. Gleisberg, et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007.

[121] R. Frederix, E. Re, and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, JHEP **09** (2012) 130.

[122] S. Frixione, P. Nason, and G. Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126.

[123] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040.

[124] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070.

[125] S. Alioli, et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043.

[126] T. Sjstrand, et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159–177.

[127] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. **C58** (2008) 639–707.

[128] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. **C76** (2016) 196.

[129] ATLAS Collaboration, *ATLAS Pythia 8 tunes to* 7 *TeV data*, ATL-PHYS-PUB-2014-021, 2014, `https://cds.cern.ch/record/1966419`.

[130] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D82** (2010) 074018.

[131] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[132] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, 2012, `https://cds.cern.ch/record/1474107`.

[133] A. D. Martin, et al., *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189–285.

[134] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823.

[135] D. Peng-Fei, et al., *QCD corrections to associated production of $t\bar{t}\gamma$ at hadron colliders*, `arXiv:0907.1324 [hep-ph]`.

[136] K. Melnikov, M. Schulze, and A. Scharf, *QCD corrections to top quark pair production in association with a photon at hadron colliders*, Phys. Rev. **D83** (2011) 074013.

[137] ATLAS Collaboration, *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008, `https://cds.cern.ch/record/1099735`.

[138] ATLAS Collaboration, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, Tech. Rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, Geneva, Mar, 2007. `https://cds.cern.ch/record/1020106`.

[139] T. G. Cornelissen, et al., *The global χ^2 track fitter in ATLAS*, J. Phys. Conf. Ser. **119** (2008) 032013.

[140] ATLAS Collaboration, *Photon identification in 2015 ATLAS data*, ATL-PHYS-PUB-2016-014, 2016, `https://cds.cern.ch/record/2203125`.

[141] ATLAS Collaboration, *Public plots from the ATLAS Egamma group*, 5th October 2018. `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/EGAM-2017-004/index.html`.

[142] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016, `https://cds.cern.ch/record/2157687`.

[143] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **76** (2016) 292.

[144] ATLAS Collaboration, *Properties of jets and inputs to jet reconstruction and calibration with the ATLAS detector using proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-036, 2015, `https://cds.cern.ch/record/2044564`.

[145] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 063.

[146] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490.

[147] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002.

[148] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-015, 2015, `https://cds.cern.ch/record/2037613`.

[149] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014, `https://cds.cern.ch/record/1700870`.

[150] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022, 2015, `https://cds.cern.ch/record/2037697`.

[151] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, `arXiv:0901.0512 [hep-ex]`.

[152] ATLAS Collaboration, *Commissioning of the ATLAS high performance b-tagging algorithms in the* 7 *TeV collision data*, ATLAS-CONF-2011-102, 2011, `https://cds.cern.ch/record/1369219`.

[153] ATLAS Collaboration, *Expected performance of missing transverse momentum reconstruction for the ATLAS detector at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-023, 2015, `https://cds.cern.ch/record/2037700`.

[154] M. Cacciari, G. P. Salam, and G. Soyez, *The Catchment Area of Jets*, JHEP **04** (2008) 005.

[155] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, ArXiv Physics e-prints (2003).

[156] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. C **77** (2017) 317.

[157] ATLAS Collaboration, *Trigger Menu in 2016*, ATL-DAQ-PUB-2017-001, 2017, `https://cds.cern.ch/record/2242069`.

[158] ATLAS Collaboration, *The ATLAS Electron and Photon Trigger*, `https://cds.cern.ch/record/2290123`.

[159] 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*, ATLAS-CONF-2014-058, 2014, `https://cds.cern.ch/record/1951336`.

[160] 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*, `http://cds.cern.ch/record/1951336`.

[161] Benedikt Völkel, *Studies of the discrimination between prompt photons and hadron fakes using neural networks*, Master's thesis, II.Physik-UniGö-MSc-2017/07 (2017).

[162] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. C **73** (2013) 2304.

[163] ATLAS Collaboration, *Single hadron response measurement and calorimeter jet energy scale uncertainty with the ATLAS detector at the LHC*, Eur. Phys. J. C **73** (2013) 2305.

[164] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 17.

[165] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, JINST **11** (2016) P04008.

[166] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **76** (2016) 653.

[167] I. Asimov and D. Shannon, *Franchise*. Creative Classic. Creative Education, 1989.

[168] D. Abdurachmanov et al., *Explorations of the viability of ARM and Xeon Phi for physics processing*, J. Phys. Conf. Ser. **513** (2014) 052008.

[169] D. Abdurachmanov, et al., *Initial explorations of ARM processors for scientific computing*, J. Phys. Conf. Ser. **523** (2014) 012009.

[170] D. Abdurachmanov, et al., *Heterogeneous High Throughput Scientific Computing with APM X-Gene and Intel Xeon Phi*, J. Phys. Conf. Ser. **608** (2015) 012033.

[171] M. A. Cox and B. Mellado, *The Massive Affordable Computing Project: Prototyping of a High Data Throughput Processing Unit*, J. Phys. Conf. Ser. **608** (2015) 012002.

[172] J. W. Smith and A. Hamilton, *Massive affordable computing using ARM processors in high energy physics*, J. Phys. Conf. Ser. **608** (2015) 012001.

[173] J. W. Smith and A. Hamilton, *ANA - A framework for building ATHENA on ARM*, J. Phys. Conf. Ser. **645** (2015) 012026.

[174] J. W. Smith et al., *ATLAS software stack on ARM64*, J. Phys. Conf. Ser. **898** (2017) 072001.

[175] G. Barrand et al., *GAUDI - A software architecture and framework for building HEP data processing applications*, Comput. Phys. Commun. **140** (2001) 45–55.

Appendices

APPENDIX A

Porting the ATLAS software stack to the ARM architecture

By using fewer instructions on a silicon chip, fewer transistors are needed, and thus the power consumption of a CPU is reduced. This makes such chips ideal for portable devices as the battery life is extended. Around 60% of mobile devices use the ARM (Advanced RISC (Reduced Instruction Set Computing) Machine) architecture. This includes smartphones, tablets, wearables and e-readers, with the percentage much higher for just smartphones. As of 2016, more than 86 billion ARM based chips have been shipped[1]. A large part of the appeal is that ARM does not produce any of their own CPUs and so has very few overhead costs. Instead, they sell their intellectual property. This gives companies more freedom to design their own CPUs.

Interest in ARM at the LHC began around 2013/14 with feasibility studies and the potential payoffs. Many standard benchmark studies were performed, centred around the 32-bit embedded ARM development boards. Experiments such as ATLAS, CMS, and LHCb also started work on porting their software stacks to the 32-bit ARM architecture [168–173]. However, it was found that while it is indeed very power conservative, ARM 32-bit did not have the available memory to make HEP computing feasible.

In 2016 ARM entered the server market and introduced their 64-bit architecture[2]. The latest results from porting and benchmarking a subset of the ATLAS software stack to the ARM 64-bit architecture (Aarch64) can be found in [174] and is summarised in this appendix. It should be noted that the fast paced environment of computing will see these types of benchmark results continuously changing as each architecture vie for the better efficiency. However, the message is the same; Aarch64 presents a feasible alternative to Intel x86.

A.1. Hardware

In early 2016 CERN openlab installed a cluster of Aarch64 evaluation prototype servers, which will be referred to as (Aarch64_Proto). Each server is comprised of a single-socket, ARM 64-bit system-on-a-chip, with 32 Cortex-A57 cores. In total, each server has 128 GB RAM connected with four fast memory channels. Another type of ARM server is also maintained, called HP Moonshot. Two types of Intel servers (Intel Atom and Intel Xeon) are also available. The Intel

---

[1] Taken from https://www.arm.com.

[2] The bottleneck is no longer CPU clock cycles, but rather the memory bandwidth. The 32-bit architectures limit the amount of random access memory (RAM).

Atom was the company's initial response to ARM and is fairly old, however it still provides interesting results. The features of each server are described in Table A.1.

Name	Processor	Cores	RAM	Cache	Fabrication (Release)	OS
HP Moonshot	X-Gene, 2.4 GHz	8 Armv8	64 GiB DDR3 (1600 MHz)	32 KiB L1/core, 256 KiB L2/core pair, 8 MiB L3	40 nm (2014)	Ubuntu 14.04
Aarch64_Proto	-, 2.1 GHz	32 Cortex-A57	128 GiB DDR3 (1866 MHz)	32 KiB L1, 1 MiB L2	16 nm (-)	Ubuntu 14.04
Intel Atom	Intel Atom Processor C2750, 2.4GHz	8	32 GiB DDR3 (1600 MHz)	24 KiB L1d, 32 KiB L1i, 1 MiB L2	22 nm (2013)	Fedora 21
Intel	Intel Xeon CPU E5-4650, 2.70 GHz	32	512 GiB DDR3 (1600 MHz)	32 KiB L1(d)(i)/core, 256 KiB L2/core, 20 MiB L3	32 nm (2012)	Scientific Linux CERN 6

Table A.1.: The hardware for Intel and Aarch64 servers [174].

A.1.1. AthSimulation

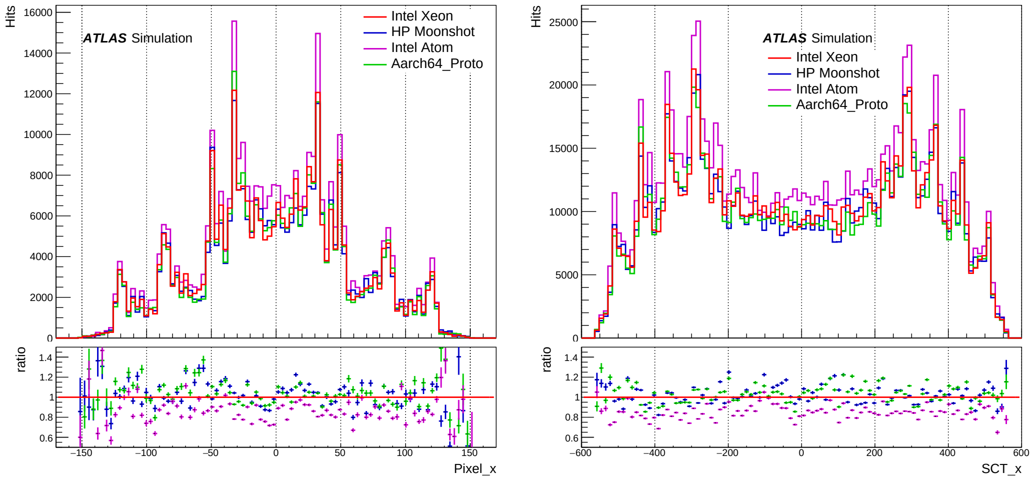
The ATLAS codebase (Athena) consists of around 2400 packages and 6.5 million lines of code. Due to its size and complexity, porting to alternative architectures is difficult. Thus, a project called AthSimulation was chosen, which consists of a subset of packages from a full Athena release. AthSimulation is capable of carrying out CPU intensive MC simulations needed for the experiment. At around 350 packages, the porting process becomes significantly easier and faster. Typical non-ATLAS specific packages include ROOT, GEANT4 and GAUDI [175]. ATLAS specific simulation code forms the peak of the pyramid of dependencies. In general, compilation options needed to be added for Aarch64, Intel specific compiler options removed and various build configurations made more general.

To enable multiple builds on various servers a continuous integration tool, Jenkins, was used. AthSimulation was compiled on both Aarch64 servers, while the Intel servers made use of the equivalent pre-existing versions built within ATLAS. The benchmark consists of simulating 100 $t\bar{t}$ events on the respective servers. Simulation took between 3.5 minutes (Intel Xeon) to just under 12 minutes (Intel Atom) per event depending on the architecture.

Validation

Validating the results from different architectures was made a high priority since this had not been checked before. Due to the nature of MC simulation, numerical identity is not expected. Some reasons can include and are not limited to random numbers being generated in an architecturally specific way, as well as the way floating point numbers are handled by each compiler. However, overall trends are expected to be similar. Figure A.1 shows the hits in the pixel and SCT detectors for different architectures compared to the Intel Xeon. When compared to the Intel Xeon, the ARM and Intel Atom servers give similar distribution shapes, but around 10-15% less hits on ARM and 15-20% more hits on Intel Atom. Further studies with larger simulated datasets needs to be done to understand why this occurs.

Figure A.2 shows a selected $t\bar{t}$ event simulated on Intel Xeon and Aarch64_Proto. There are subtle differences in energy deposits due to a different number of hits and also a very minor difference in tracks. However, the general topology of the event matches well.



(a) (b)

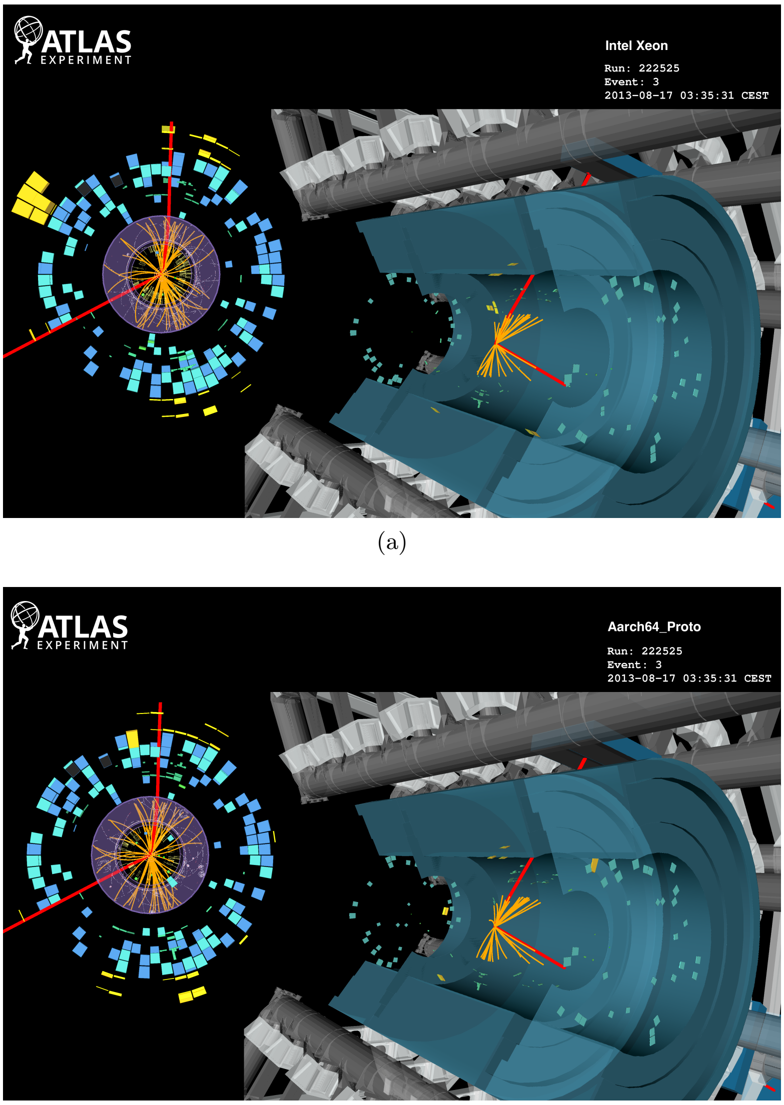
Figure A.1.: Results showing the hits in the a) pixel and b) SCT detectors in ATLAS. The ratio between the three servers and the Intel Xeon is shown in the ratio plots. The Intel Xeon is taken as the “accepted” distribution. [174]

A.1.2. Power measurements

To take memory bandwidth into account while performing power measurements, the benchmark was repeated on the Aarch64_Proto and the Intel Xeon server with multiple jobs running. In this benchmark, eight $t\bar{t}$ events are simulated on an increasing number (2,4,8,16,32) of cores. Figure A.3 shows the results. The top plot shows the net power usage during each test, while the bottom image shows the events/kWh for each test. Taking time and power usage into account, this shows that for these CPUs, the Aarch64 server is more efficient.

A.1.3. Outlook for ARM in HEP

It is clear that ARM servers have improved dramatically over the past few years. Their 64-bit architecture is now competitive with the traditional Intel machines. Studies like these show academia and industry alike that there is no need to be reliant on a single computing architecture. Various reasons ranging from cost-effectiveness to geopolitical issues could result in server farms being made available that might not necessarily be the standard Intel architecture. CERN needs to be ready to utilise whatever computing power is available at the most reasonable price, especially with the ever increasing computing demands. As an experiment ATLAS needs to be flexible in how they can deploy their software. Steps have been undertaken to make the ATLAS software more independent of operating system and CPU architecture.


ATLAS
EXPERIMENT
Intel Xeon
Run: 222525
Event: 3
2013-08-17 03:35:31 CEST
ATLAS
EXPERIMENT
Aarch64_Proto
Run: 222525
Event: 3
2013-08-17 03:35:31 CEST

(a)

(b)

Figure A.2.: Event displays for the same $t\bar{t}$ event simulated on the a) Intel Xeon and b) Aarch64_Proto server. Minor differences can be seen in the calorimeter energy deposits and tracks [174].

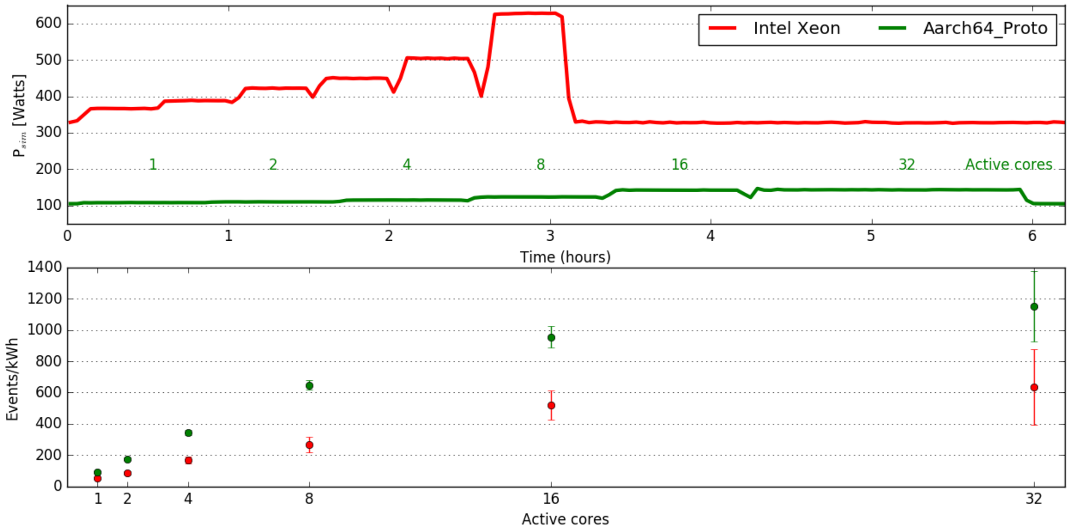
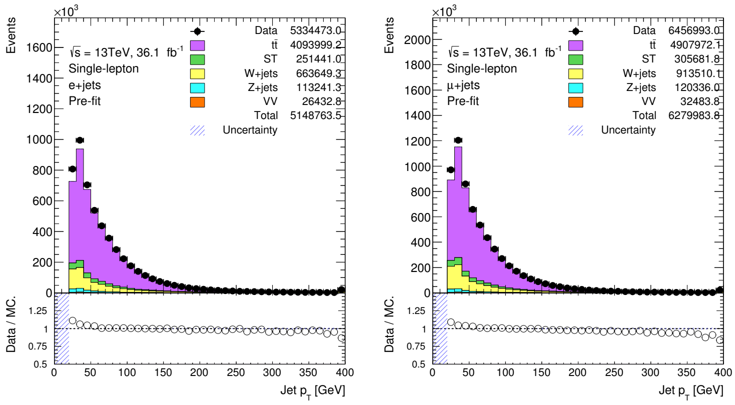


Figure A.3.: Power measurements for the benchmark running on an increasing number of cores. Top: total time and power for all tests. Bottom: Events/kWh calculated for each test [174].

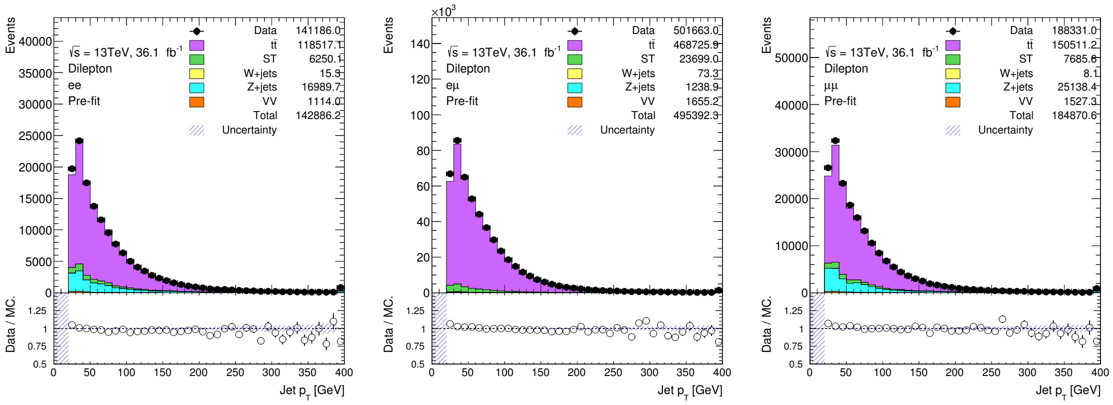
APPENDIX B

Support material for the $t\bar{t}\gamma$ process

B.1. Validation plots for $t\bar{t}$ selections



(a) e+jets (b) μ+jets



(c) ee (d) $e\mu$ (e) $\mu\mu$

Figure B.1.: Validation plots of the jet p_{T} for event selections before the requirement of a photon. Only statistical uncertainties are included. Not included is the fake lepton background. All processes shown are MC.

APPENDIX C

Object- and event-level neural network support material

C.1. Event-level Descriminator

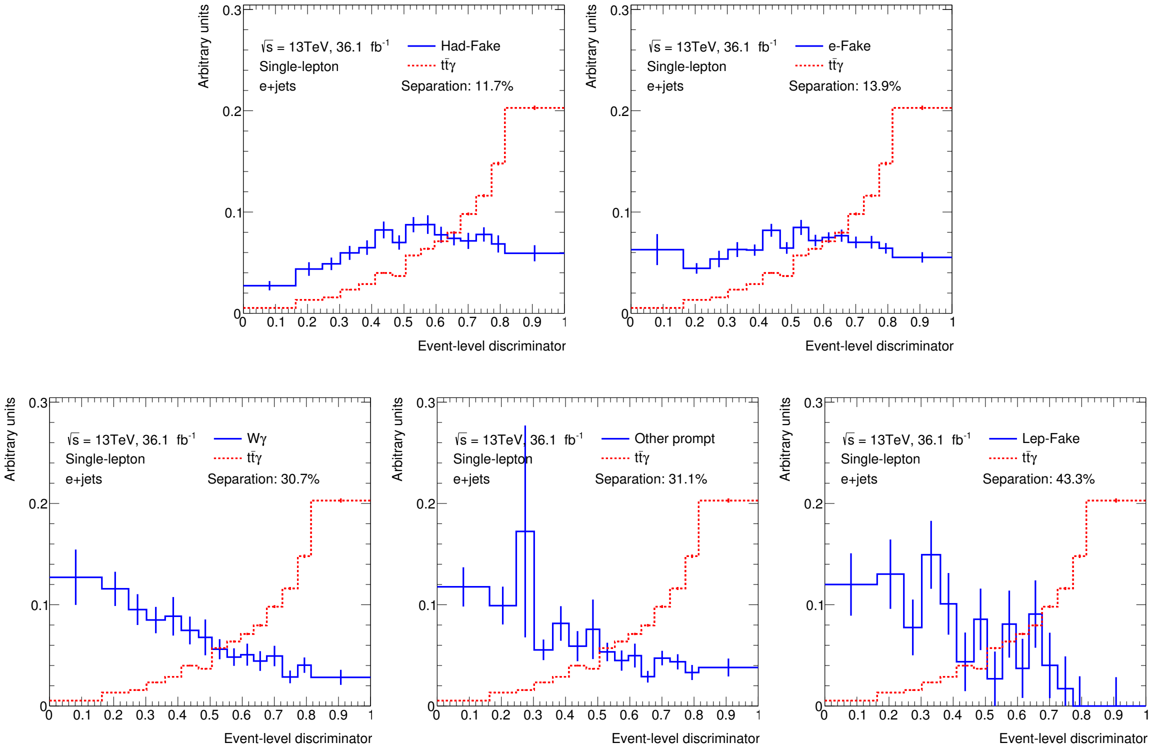


Figure C.1.: Event-level Discriminator variable for the e+jets channel with separation plots shown for signal and the different background components.

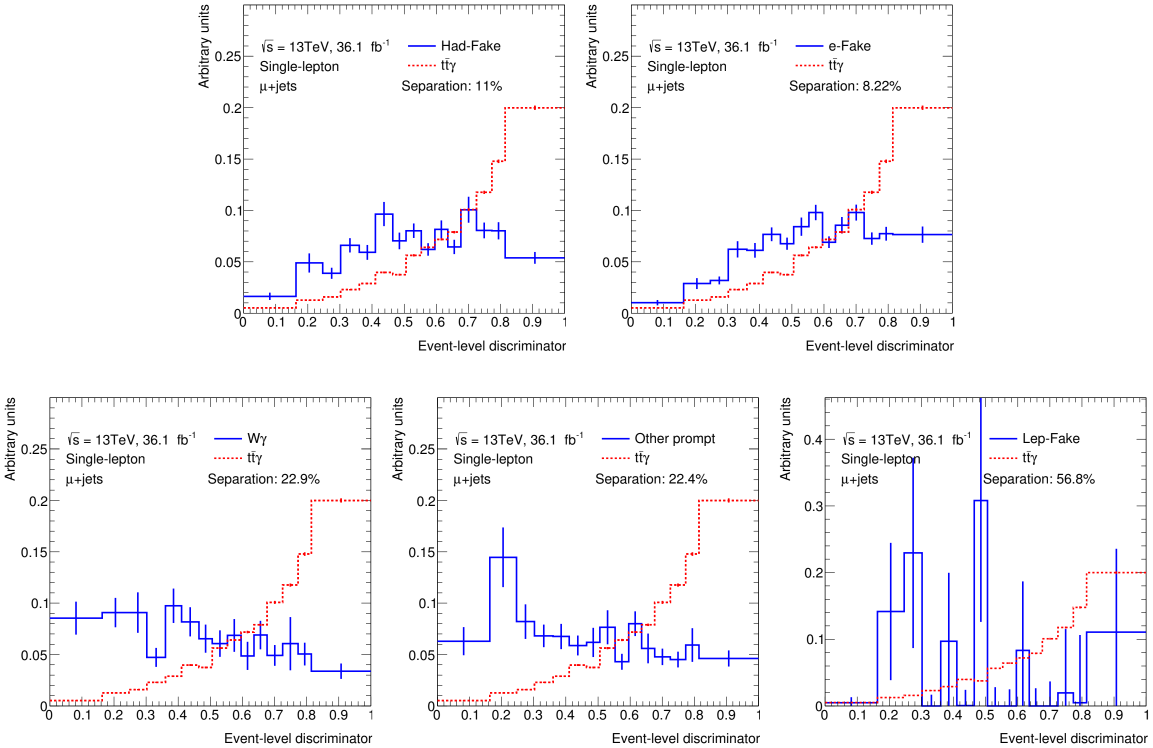


Figure C.2.: Event-level Discriminator variable for the μ+jets channel with separation plots shown for signal and the different background components.

APPENDIX D

Systematic uncertainties support material

D.1. Cross-section closure test for small MC backgrounds

There is currently a 50% normalisation uncertainty applied to each of the single top, diboson and $t\bar{t}V$ backgrounds This section checks the impact the systematic uncertainty has for these prompt backgrounds, grouped as Other prompt. For the single-lepton and dilepton channel two Asimov fits are performed with all systematic sources included. In the first fit a 50% uncertainty is assigned to each of these three backgrounds. In the second fit the systematic uncertainty is doubled, i.e. a 100% uncertainty is assigned. The results for the fits are shown in Table D.1. Doubling the uncertainty on the three backgrounds leads to an absolute increase in the error of ≈ 0.2% for both the single-lepton and dilepton channels. This impact is negligible and means these backgrounds do not play an important role in the fit. For this reason, the normalisation uncertainty is left at 50%.

Channel	Normalisation uncertainty	"+" error	"-" error
single-lepton	50%	0.0929	0.0914
	100%	0.0945	0.0927
dilepton	50%	0.0678	0.0629
	100%	0.0697	0.0647

Table D.1.: Two different Asimov fit scenarios for the single-lepton and dilepton channels where the normalisation uncertainty on the Other prompt backgrounds has been varied. All systematics are included in the fit. The error represents the absolute up/down error of μ.

APPENDIX E

Results section support material

E.1. Final fits

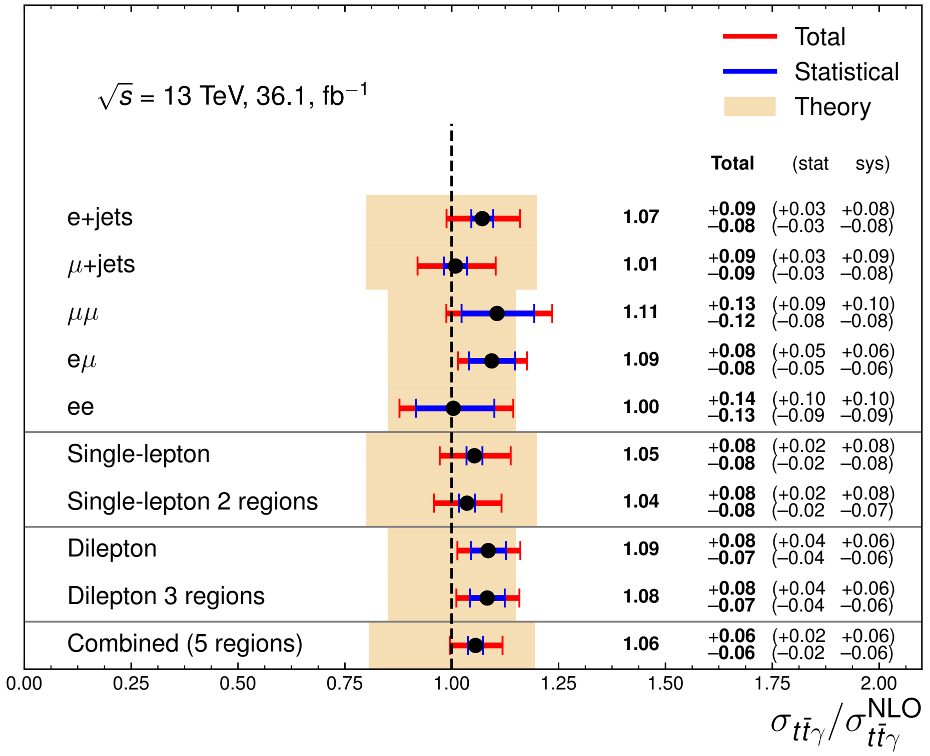
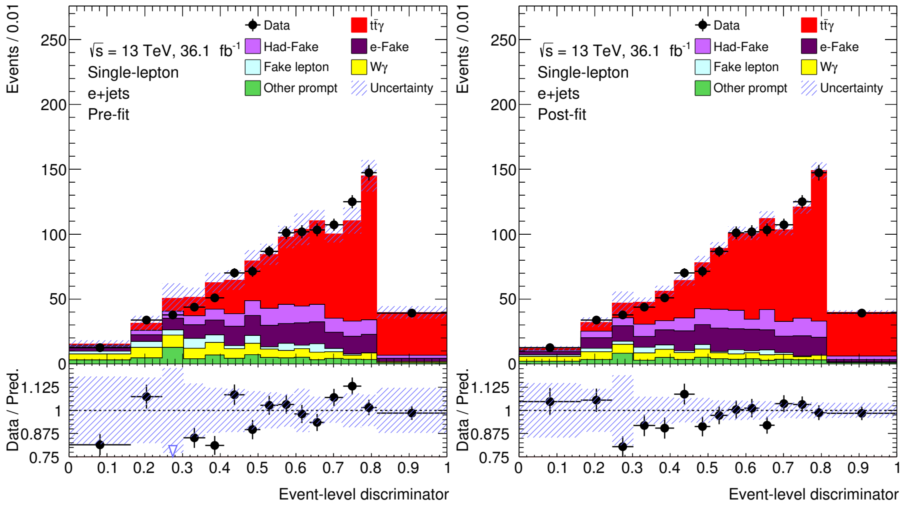
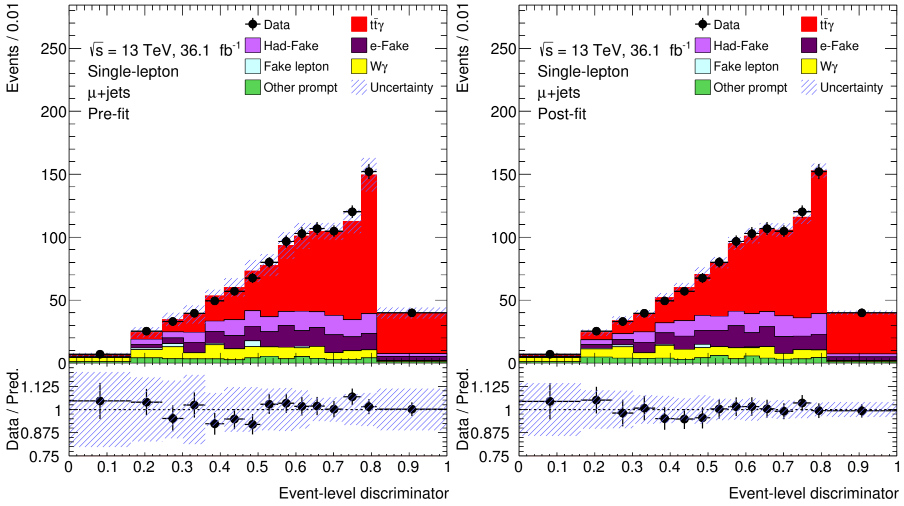


Figure E.1.: The observed signal strength for individual, merged and combined single-lepton and dilepton, and combined channels. The NLO SM prediction is represented by the dashed vertical line. The theoretical uncertainty for each fit is represented by the shaded region.

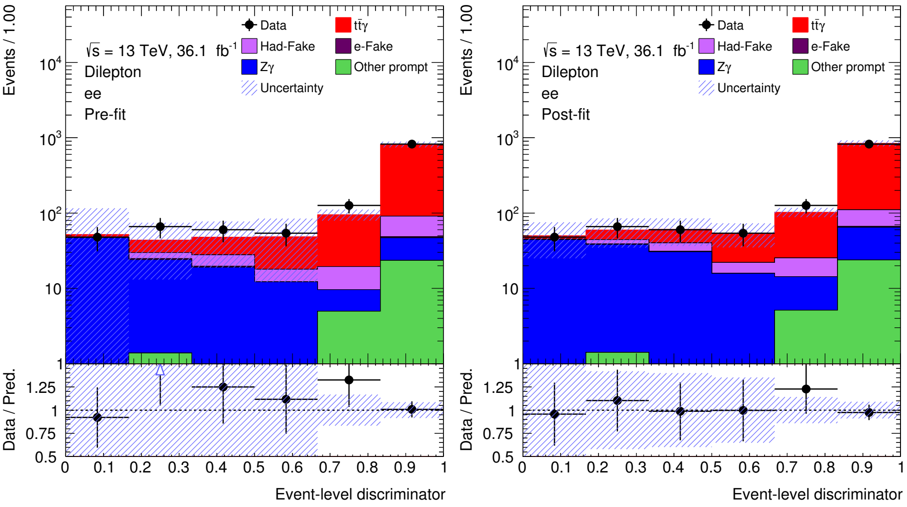


(a) Pre-fit e+jets (b) Post-fit e+jets

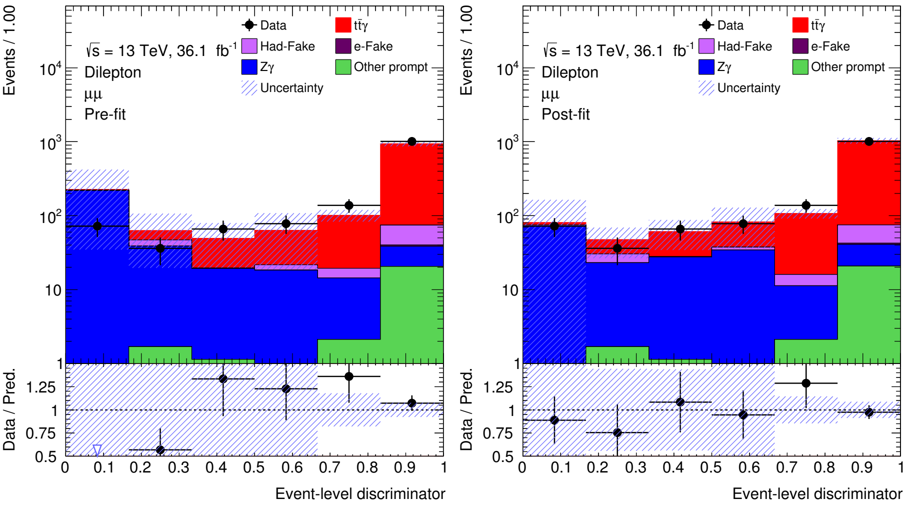


(c) Pre-fit μ+jets (d) Post-fit μ+jets

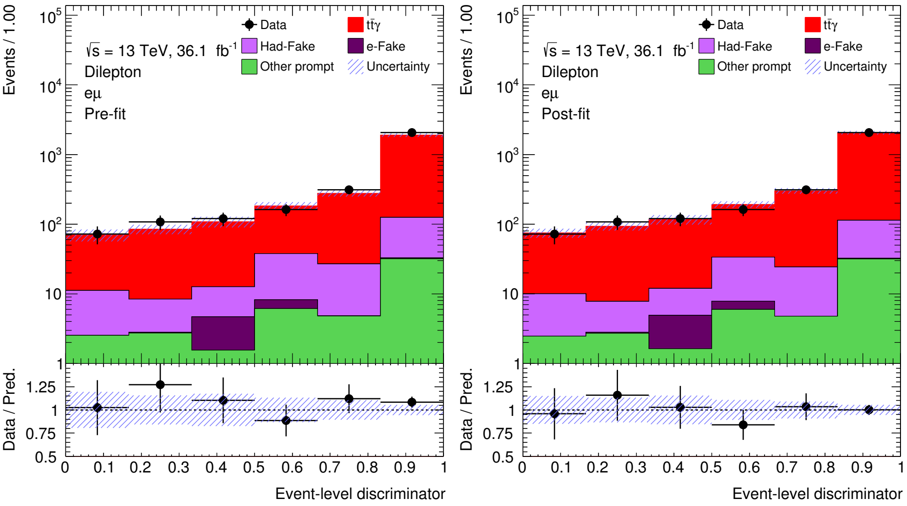
Figure E.2.: Pre- and post-fit plots for the individual single-lepton channels where the respective ELD distribution is used as the discriminating variable in the fit.



(a) Pre-fit ee (b) Post-fit ee



(c) Pre-fit $\mu\mu$ (d) Post-fit $\mu\mu$



(e) Pre-fit $e\mu$ (f) Post-fit $e\mu$

Figure E.3.: Pre- and post-fit plots for the individual dilepton channels where the respective ELD distribution is used as the discriminating variable in the fit.

	e+jets		μ+jets		*ee*		$\mu\mu$		$e\mu$	
	+ [%]	- [%]	+ [%]	- [%]	+ [%]	- [%]	+ [%]	- [%]	+ [%]	- [%]
Signal Modelling	1.36	1.21	0.50	0.61	7.40	5.80	3.20	2.88	3.60	3.53
Jets	5.30	4.60	5.60	5.00	3.34	2.96	2.91	2.66	1.61	1.45
Luminosity	2.33	2.03	2.35	2.01	2.47	1.99	2.35	2.01	2.28	2.02
Pileup	2.35	2.08	2.37	2.21	2.34	1.96	4.30	3.70	1.38	1.33
Photon Efficiencies	1.09	0.95	1.10	0.94	1.16	0.93	1.11	0.97	1.08	0.95
b-Tagging	0.86	0.97	0.80	0.84	0.58	0.57	0.27	0.31	0.14	0.16
Background modelling	4.50	4.50	6.00	5.90	3.53	3.41	3.04	3.24	2.98	3.08
Leptons	0.95	0.87	1.64	1.52	2.16	1.81	1.78	1.56	1.18	1.04
Prompt photon tagger (shape)	3.26	3.42	4.20	4.30	-	-	-	-	-	-
Eγ (Resolution and scale)	-	-	0.06	0.05	0.35	0.31	0.08	0.10	0.08	0.12
Template Statistics	2.40	2.28	2.84	2.66	3.80	3.25	4.10	3.70	1.76	1.61
Total systematic	7.9	7.5	9.0	8.6	10.5	8.8	8.5	7.7	5.6	5.3
Total statistical	2.1	2.1	2.2	2.2	9.4	8.9	7.9	7.5	5.0	4.9
Total	8.2	7.8	9.3	8.8	14.0	12.5	11.6	10.8	7.5	7.2

Table E.1.: Relative effects on μ due to the grouped systematic sources. Fits are performed with nuisance parameters in each group held constant with the rest floating. This new uncertainty is subtracted in quadrature from the total uncertainty to obtain $\Delta\mu$.

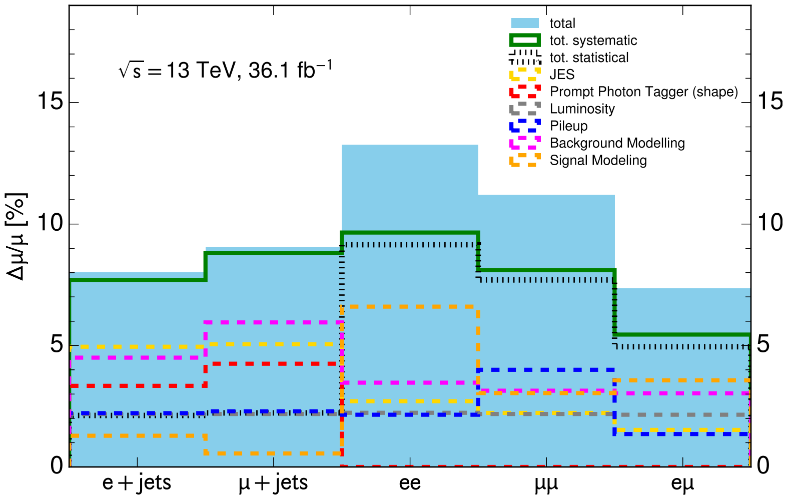


Figure E.4.: The relative contribution of uncertainties shown in groups. A fit is performed where all NPs in a group are kept fixed while the rest are left floating.The resulting uncertainty is subtracted in quadrature from the original uncertainty to obtain $\Delta\mu$.