Precision measurements of differential $t\bar{t}$ production cross sections as a function of kinematic event variables at 13 TeV at CMS

By

Douglas John Paul Burns

Department of Physics
University of Bristol

IIHE
University of Brussels

A dissertation submitted to the University of Bristol and Vrije Universiteit Brussel in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Science, School of Physics, Bristol and Faculteit Wetenschappen en Bio-ingenieurswetenschappen, Vakgroep Fysica, Brussel.

August 26, 2018

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University of Bristol and Vrije Universiteit Brussels regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE: .....

Author's contribution

The research conducted for this thesis has been done in collaboration with the Top Quark Group at the University of Bristol and the Vrije Universiteit Brussel, and all material presented here, unless stated otherwise, has been conducted by the author and collaborators. The author has spent time studying at both institutions as well as spending nine months working at CERN.

Contributions of the author include a major update of the pre-existing analysis performed at 7 and 8 TeV in order to expand and enhance it for 13 TeV data. The author contributed to a Physics Analysis Summary of differential top quark cross sections with respect to global event variables, based on the first $71\,\mathrm{pb}^{-1}$ of data collected by CMS in 2015.

> CMS Collaboration, "Measurement of differential top quark pair production cross sections in a fiducial volume as a function of event variables in pp collisions at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-TOP-15-013, 2015 [7 cites]

The summary incorporates several new event variables. The small data sample lead to measurements that were statisitically dominated.

A second analysis, for which the author was the primary contact person, was performed with much greater rigour. This includes, but is not limited to, the inclusion of correlations between the bins of a measurement and the addition of absolute cross section measurements. The author was also heavily involved in the update and implementation of systematic uncertainties, including the calculation of the b tagging efficiencies, the addition of a new set systematic uncertainties based on the modelling in simulation and the creation of the systematic covariance matrices used for the goodness-of-fit tests. The author has performed many validation tests both on data and simulation during the course of the research and has presented multiple times in working group meetings, including for the preapproval of the analysis. The author wrote significant parts of the documentation of the analysis, including for the paper, now published in the June 2018 volume of the Journal for High Energy Physics.

> CMS Collaboration, "Measurements of differential cross sections of top quark pair production as a function of kinematic event variables in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **06** (2018) 002, doi: 10.1007/JHEP06(2018)002, `arXiv:1803.03991` [2 cites]

The author presented the findings of this research at the International Conference for High Energy Physics in July 2018, as well as summarising the state of top quark physics at CMS at that time. All work beyond the scope of the paper has been performed by the author.

The author has contributed to the CMS experiment by performing Level-1 trigger shifts in the CMS control room during periods of data taking. The author was involved in efficiency studies for the High Level triggers used in 2016. Further experimental work is ongoing in relation to the so-called 'Highly Ionising Particle' effect, prevelant during data taking in 2016. This is where a build up of charge on the preamplifier of the readout chip, assumed to originate from a highly ionising particle, caused a loss in gain over subsequent bunch crossings and hence decreased the tracker hit efficiency. The problem was finally identified in a parameter controlling the rate at which charge bleeds away from the charge sensitive capacitor and the effect vanished. However, the effect of charge build up could become an issue again in the future at the much higher luminosities expected and so the author created a standalone simulation to model the response of the readout chip with respect to the charge buildup. In the future this will supply a correction based on measured gain distributions.

The author has performed outreach activities during the course of the research studies, in particular the CERN international Masterclass, the London Youth International Science Forum and giving lectures for high school students interested in Physics. The author has also been a demonstrator for degree students in the Univeristy of Bristol's experimental laboratory.

Abstract

Measurements of differential $t\bar{t}$ production cross sections are presented in the single-lepton decay channel, as a function of several kinematic event variables. These kinematic event variables do not require the reconstruction of the $t\bar{t}$ system and are N_{jets}, H_{T}, S_{T}, $p_{\text{T}}^{\text{miss}}$, p_{T}^{W}, p_{T}^{ℓ} and $|\eta^{\ell}|$. The measurements are performed with proton-proton collision data collected by the CMS experiment at the LHC during 2016, with an integrated luminosity of 35.9 fb^{-1}. The single top, V + jets and multijet QCD backgrounds are subtracted from the data yields to obtain the $t\bar{t}$ yield. The results are presented to particle level in a phase space similar to that of the detector. The required extrapolation is performed by unfolding the yield of $t\bar{t}$ events obtained. The unfolding is also used to correct for any detector acceptance, efficiency and bin-to-bin migration effects. Both normalised and absolute cross sections are calculated from the unfolded $t\bar{t}$ yields and are compared to state-of-the-art leading order and next-to-leading order $t\bar{t}$ simulations.

Contents

1 Introduction **1**

2 Theoretical overview **3**
2.1 The components of the standard model . . . 3
2.2 Representing particles and their interactions . . . 4
2.3 Quantum chromodynamics . . . 6
2.3.1 Colour confinement and hadronisation . . . 6
2.4 The weak interaction and electroweak unification . . . 7
2.4.1 Electroweak symmetry breaking . . . 8
2.4.2 The Cabbibo-Kobayashi-Maskawa mixing matrix . . . 9
2.5 The standard model and its shortcomings . . . 10
2.5.1 Gravity . . . 10
2.5.2 Massive neutrinos . . . 10
2.5.3 Dark matter . . . 10
2.5.4 Baryon asymmetry . . . 11
2.5.5 Hierarchy problem . . . 11
2.6 Top quark physics . . . 11
2.6.1 Top quark production . . . 12
2.6.2 Order of perturbation . . . 13
2.6.3 Top quark decay . . . 15
2.6.4 Top quark cross section measurements . . . 15

3 Modelling proton-proton interactions **17**
3.1 Hard interaction . . . 17
3.2 Additional interactions and soft processes . . . 19
3.3 Matching matrix-elements to parton showers . . . 20
3.4 Fragmentation and colour reconnection . . . 21
3.4.1 Lund string model . . . 21
3.4.2 Cluster fragmentation model . . . 23
3.5 Detector modelling . . . 24
3.6 Production of $t\bar{t}$ and background models . . . 24

4 The Large Hadron Collider and the Compact Muon Solenoid **28**
4.1 The Large Hadron Collider . . . 28
4.1.1 A brief history of the LHC and CMS . . . 28
4.1.2 Operating the LHC . . . 30
4.2 Luminosity measurements . . . 31
4.3 The Compact Muon Solenoid . . . 33
4.3.1 CMS pixel and strip trackers . . . 33
4.3.2 Electromagnetic calorimeter . . . 36
4.3.3 Hadronic calorimeter . . . 37
4.3.4 Superconducting magnet . . . 39
4.3.5 Muon detectors . . . 40
4.3.6 Triggers . . . 41
4.3.7 Computational power . . . 42
4.3.8 Data collection in 2016 . . . 43

5 Reconstructing the physics objects **45**
5.1 The particle flow algorithm . . . 45
5.1.1 Particle flow elements . . . 45
5.1.2 Linking algorithm . . . 46
5.1.3 PF muons . . . 47
5.1.4 PF electrons and photons . . . 47
5.1.5 PF jets . . . 48
5.1.6 PF missing transverse momentum . . . 50
5.1.7 PF isolation . . . 52
5.2 Analysis object selection . . . 52
5.2.1 Electron objects . . . 52
5.2.2 Muon objects . . . 53
5.2.3 Jet Objects . . . 54

6 Analysis formation, event selection and corrections **55**
6.1 Kinematic event variables and the event selection . . . 55
6.2 Jet energy corrections . . . 56
6.3 Reweighting from additional interactions . . . 58
6.4 Reweighting from b jet identification . . . 59
6.5 Reweighting from leptons . . . 60
6.6 Multijet QCD . . . 60
6.7 Data-simulation agreement in the signal region . . . 69
6.7.1 Calculating the $t\bar{t}$ yield . . . 69

7 Removing the CMS thumbprint **73**
7.1 Particle level measurement . . . 73
7.2 Choice of bins . . . 74
7.3 Unfolding . . . 82
7.4 Cross checking the unfolding . . . 83
7.4.1 Checking the uncertainties . . . 83
7.4.2 Checking for bias . . . 83
7.4.3 Calculating the cross sections . . . 86

8 Uncertainty **92**
8.1 Statistical uncertainty 92
8.2 Systematic uncertainties 92
8.2.1 Experimental uncertainties 93
8.2.2 Parton shower uncertainties 93
8.2.3 Other theoretical uncertainties 94
8.2.4 Reporting the uncertainties 94

9 Results **101**
9.1 Cross section measurements 101
9.2 The χ^2 goodness-of-fit tests 102
9.3 Effects of regularisation 102

10 Conclusions, prospects and outlook **113**
10.1 Double differential cross sections 114
10.2 Additional event variables 114
10.3 Effective field theory 114

Figures and Plots

1 A Feynman diagram showing the $t\bar{t}$ pair production by quark-antiquark annihilation and decay into the single lepton final state. Also shown are radiative processes in both the initial (ISR) and final state (FSR). Within the proton, there are also many QCD interactions occuring, such that the colliding partons in the initial state are described by the parton distribution function. All Feynman diagrams have been created using the feyntikz package. 5

2 A Feynman diagram showing the $t\bar{t}$ pair production and decay. An example QCD interaction is shown in bold red. Other QCD interactions are shown in light red. 6

3 The running of the strong coupling constant. At higher energy scales α_S tends to 0, resulting in a strengthening of the strong force. 7

4 A Feynman diagram showing the $t\bar{t}$ pair production and decay. An example weak interaction is shown in bold blue. Other weak interactions are shown in light blue. 8

5 Feynman diagrams for the production of single top quarks. The left panel shows single top quark production via the s-channel, the central panel by the t-channel and the right panel in association with a W boson. 12

6 The parton distribution function of the proton scaled by the fraction of momentum carried by the parton, $xf(x, \mu^2)$ with respect to the fraction of momentum carried by the parton. The left panel shows the distributions with respect to a lower energy scale and the right panel with respect to a higher energy scale. . . 13

7 The inclusive $t\bar{t}$ cross section, measured by ATLAS, CMS and the Tevatron in different decay channels at multiple $\sqrt{s}$. Also shown is the most precise theoretical prediction to date. 14

8 A set of Feynman diagrams showing LO and NLO processes. In the top left panel, the leading order diagram is shown. The top right panel shows a gluon-loop in the propagator. These loops are responsible for the running strong coupling constant. The two middle panels show NLO loops providing virtual particles to the diagrams and corrections in the renormalisation scale. Finally, the bottom two panels show the real additional partons that can be seen in the final state due to NLO matrix-element calculations. 14

9 Comparison of different $t\bar{t}$ models with the old CUETP8M1 tune, represented by dashed lines, and the new CUETP8M2T4 tune, given by solid lines. The left panel shows the comparison for the distribution of the p_T of the hadronically decaying top quark and the right panel for the additional jets distribution. . . . 16

10 An example simulation of a proton-proton collision showing all the various processes that must be modelled. Starting with the hard interaction in red calculated using perturbation theory, then dressing with additional soft processes in blue forming the parton shower. Additional parton interactions are added, shown in purple, each with a respective parton shower. Finally the parton shower is fragmented and hadronised into detectable, colourless hadrons, which can in turn decay and radiate. This fragmentation and hadronisation processes are shown in light green and green respectively. Photon radiation is shown in yellow. . . . 18

11 The four possible processes to strongly split a parton in the parton shower. The top left panel shows gluon emission, the top right shows $q\bar{q}$ pair production and the bottom two panels the colour configurations of gluon splitting. In addition, photon brehmsstrahlung from the quarks can occur but is not considered here. . 19

12 The left panel shows parton shower ordering by k_T^2 where the splitting strength $V_i(k_T^2)$ is ordered $V_1 > V_2 > V_3$. The right panel shows shower ordering by angle such that the opening angle decreases for every consecutive splitting $\theta_{max} > \theta_1 > \theta_2 > \theta_{3,4}$. 20

13 The left panel shows the string between the quark-antiquark colour dipole stretched by a gluon and the right panel shows for each string how fragmentation occurs with time. Lorentz boosts in the hadrons are shown by oscillation frequency. 22

14 A sketch of the cluster model of fragmentation and hadronisation. Colour singlet clusters are generated from the final state quarks and gluons forced to split into a q pair. These colour singlet clusters can be unnaturally massive and fission occurs until the clusters represent excited hadron resonances which then hadronise. . . 23

15 The full CERN complex showing schematically the location of all the experiments on the ring complex, including the location of the four main experiments at the LHC: CMS, ATLAS, LHCb and ALICE. 29

16 A cross section of a LHC dipole magnet. 31

17 The left panel shows the delivered and recorded integrated luminosity during data taking in 2016 and the right panel the integrated luminosity delivered by the LHC for all years of operation. 32

18 The distribution of the number of interactions in a bunch crossing. The minimum bias cross section is taken to be 80 mb. 32

19 A schematic of CMS, sliced to show the internal layout. It consists of layers around the central interaction point in the order of silicon pixel and strip detectors, electromagnetic and hadronic calorimeters, superconducting magnet and finally muon chambers. 34

20 A slice through the CMS tracker detector in the $r - z$ plane. The separate subsystem regions have been highlighted and are classified as the Tracker Inner Barrel and Disks (TIB and TID), the Tracker Outer Barrel (TOB) and the Tracker Endcaps (TEC). The Pixel subsystem is highlighted in red. Back-to-back stereo strip modules are highlighted in blue. 35

21 A geometrical quarter view of the CMS ECAL, highlighting the barrel, endcap and preshower detector regions. 36

22 The upper panel shows the energy resolution of the CMS ECAL measured in a beam test. The energy was measured in a 3×3 crystal lattice with the electron striking the central crystal. The lower panel shows the resolution extracted from an unbinned likelihood fit on $Z \to e^+e^-$ events. 38
23 A geometrical quarter view of CMS HCAL, highlighting the Barrel (HB), Endcap (HE), Forward (HF) and Outer (HF) subdetectors. 39
24 The transverse energy resolution of reconstructed jets in the CMS HCAL. It is split into the barrel jets ($|\eta| < 1.4$), endcap jets ($1.4 < |\eta| < 3.0$) and forward jets ($3.0 < |\eta| < 5.0$) 40
25 A geometrical one-quarter view of the CMS muon subsystems. The drift tube stations are indicated in yellow the cathode strip chambers in green and the resistive plate chambers in blue. 41
26 The muon p_{T} resolution as a function of p_{T} using the tracking only, muon subdetectors only or both. 42
27 A schematic representing the process flow of the L1T. On the left is the calorimeter trigger and on the right the muon trigger which operate on independent trigger primitives. The final decision is taken by the global trigger using all information available. 43

28 An example of clustering using the anti-k_{T} algorithm. Soft particles are preferentially clustered around high-p_{T} seeds forming conical jets. The jet boundary is defined by the respective hardness between neighbouring jets. 49
29 An illustration of a jet from a heavy flavour quark. A hadron is formed with a lifetime of $\sim 1 - 1.5$ ps which decays at a displaced secondary vertex (SV) with respect to the primary interaction vertex (PV) and hence has a large impact parameter (IP). 50
30 The flavour composition of jets from a $\mathrm{t\bar{t}}$ model simulation. A simple event selection is performed requiring an isolated lepton, four jets with $p_{\mathrm{T}} > 30$ GeV of which two are required to have CSVv2 > 0.8484. 51

31 A schematic showing the factorised steps to correct the energy scale of a jet. These consist of a pileup correction, detector response correction with additional residuals applied in data and optional flavour corrections. 56
32 The simulated jet energy response as a function of p_{T} and η. These corrections are based on the status of the detector after the end of Run I. 57
33 A plot showing the distribution of the number of reconstructed primary vertices in simulation and data. The left panel shows the distribution before pileup reweighting and the right panel after. 58
34 The left panel shows the b tagging efficiency with respect to jet p_{T} and the right panel with respect to jet η. Both panels show the efficiencies separated into b, c and light flavours. 59
35 The left panel shows the b jet multiplicity distribution before reweighting and the right panel after reweighting. 60
36 The electron identification efficiency for data and simulation calculated with events with a Z boson present using the tag-and-probe method. The ratio of the two efficiencies is the finer binned scale factors applied in the $|\eta^{\ell}|$ measurements. 61
37 The distributions of the event variables and I_{rel} in the non-isolated electron control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars. 63

38 The distributions of the event variables and I_{rel} in the conversion electron control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars. 64
39 The distributions of the event variables and I_{rel} in the non-isolated ($0.15 < I_{\mathrm{rel}} < 0.30$) muon control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars. 65
40 The distributions of the event variables and I_{rel} in the non-isolated ($0.30 < I_{\mathrm{rel}} < \infty$) muon control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars. 66
41 The prediction of multijet QCD in the alternate control region from the nominal control region for each event variable in the electron channel. Also shown as a comparison is the data-subtracted QCD estimate from the alternate control region. Below each of the distributions, the ratio of the two QCD predictions is shown. This ratio gives an estimation of the QCD normalisation uncertainty by the displacement from unity and was taken to be ±50% shown by the gold band. The QCD shape uncertainty is given by the spread of the ratio, taken to be ±30% shown by the grey band. 67
42 The prediction of multijet QCD in the alternate control region from the nominal control region for each event variable in the muon channel. Also shown as a comparison is the data-subtracted QCD estimate from the alternate control region. Below each of the distributions, the ratio of the two QCD predictions is shown. This ratio gives an estimation of the QCD normalisation uncertainty by the displacement from unity and was taken to be ±50% shown by the gold band. The QCD shape uncertainty is given by the spread of the ratio, taken to be ±30% shown by the grey band. 68
43 The distributions of N_{jets}, H_{T} and S_{T} after full event selection. The $\mathrm{t\bar{t}}$ simulation is normalised to the NNLO prediction. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in the data shown by the vertical error bars. The statistical uncertainty in the number of simulated events quadratically combined with the experimental uncertainties is shown by the hatched band. 70
44 The distributions of $p_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}^{\mathrm{W}}$, p_{T}^{ℓ} and $|\eta^{\ell}|$ after full event selection. The $\mathrm{t\bar{t}}$ simulation is normalised to the NNLO prediction. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in the data shown by the vertical error bars. The statistical uncertainty in the number of simulated events quadratically combined with the experimental uncertainties is shown by the hatched band. 71
45 The distribution of the H_{T} variable after full event selection and the application of top quark p_{T} reweighting. With the application of the reweighting the ratio is now flat. 72

46 The purity and stability of the bins in the e+jets channel, measured using the POWHEG+PYTHIA simulation sample. 75
47 The purity and stability of the bins in the μ+jets channel, measured using the POWHEG+PYTHIA simulation sample. 76

48 The residual distribution for the H_{T} event variable in the eJets channel on the left and the μ+jets channel on the right. The standard deviation of the half-gaussian, which is the resolution, is given as the vertical red line and compared to the bin width given as the vertical blue line. 77
49 The left panel shows the mapping between the reconstructed and particle-level distributions of the H_{T} event variable in the e+jets channel and the right panel in the μ+jets channel. The chosen common binning scheme passing all criteria is overlaid in red. 78
50 The set of response matrices, calculated using the POWHEG+PYTHIA simulation sample and normalised to one for all event variables in the e+jets channel using the common binning scheme. 79
51 The set of response matrices, calculated using the POWHEG+PYTHIA simulation sample and normalised to one for all event variables in the μ+jets channel using the common binning scheme. 80
52 A schematic view of the migration effects and statistical fluctuations. 82
53 A scan of the average global correlation coefficient with respect to the regularisation parameter τ in the e+jets channel. The optimal τ is shown as a red point at the minimum of the scan. 84
54 A scan of the average global correlation coefficient with respect to the regularisation parameter τ in the μ+jets channel. The optimal τ is shown as a red point at the minimum of the scan. 85
55 Total pull distribution when combining the 5000 pseudo experiments in each bins for the H_{T} variable in the e+jets channel on the left and the μ+jets channel on the right. 86
56 The pull mean and widths in relation to the bin numbers of the event variables in the e+jets channel. The 5000 pseudo experiments are generated from the POWHEG+PYTHIA response matrix. 87
57 The pull mean and widths in relation to the bin numbers of the event variables in the μ+jets channel. The 5000 pseudo experiments are generated from the POWHEG+PYTHIA response matrix. 88
58 The H_{T} event distribution of the POWHEG+PYTHIA sample with the top quark p_{T} reweighted up and down to cover differences to data in the e+jets channel. The distributions are normalised to one. 89
59 The cross sections for the reweighted models unfolded using the POWHEG+PYTHIA derived response matrix compared to the true model cross sections are shown for all event variables in the e+jets channel in the upper panels. The lower panels give the ratio of the two cross sections known as the bias. 90
60 The cross sections for the reweighted models unfolded using the POWHEG+PYTHIA derived response matrix compared to the true model cross sections are shown for all event variables in the μ+jets channel in the upper panels. The lower panels give the ratio of the two cross sections known as the bias. 91

61 The composition of the systematic uncertainties for the N_{jets}, H_{T} and S_{T} event variables. Dominant uncertainties are shown in bold. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty. 95

62 The composition of the systematic uncertainties for the $p_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}^{\mathrm{W}}$, p_{T}^{ℓ} and $|\eta^{\ell}|$ event variables. Dominant uncertainties are shown in bold. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty. 96
63 The composition of the systematic uncertainties for the N_{jets}, H_{T} and S_{T} event variables. Dominant uncertainties are shown in bold. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty. 97
64 The composition of the systematic uncertainties for the $p_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}^{\mathrm{W}}$, p_{T}^{ℓ} and $|\eta^{\ell}|$ event variables. Dominant uncertainties are shown in bold. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty. 98

65 The normalised cross section measurements for the N_{jets}, H_{T} and S_{T} kinematic event variables. They are compared against different $\mathrm{t\bar{t}}$ simulations. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below. 103
66 The normalised cross section measurements for the $p_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}^{\mathrm{W}}$, p_{T}^{ℓ} and $|\eta^{\ell}|$ kinematic event variables. They are compared against different $\mathrm{t\bar{t}}$ simulations. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below. 104
67 The absolute cross section measurements for the N_{jets}, H_{T} and S_{T} kinematic event variables. They are compared against different $\mathrm{t\bar{t}}$ simulations. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below. 105
68 The absolute cross section measurements for the $p_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}^{\mathrm{W}}$, p_{T}^{ℓ} and $|\eta^{\ell}|$ kinematic event variables. They are compared against different $\mathrm{t\bar{t}}$ simulations. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below. 106
69 The normalised cross section measurements for the N_{jets}, H_{T} and S_{T} kinematic event variables. They are compared against the POWHEG+PYTHIA simulation after varying the shower scales, the h_{damp} matching parameter and the application of top quark p_{T} reweighting. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below. 107
70 The normalised cross section measurements for the $p_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}^{\mathrm{W}}$, p_{T}^{ℓ} and $|\eta^{\ell}|$ kinematic event variables. They are compared against the POWHEG+PYTHIA simulation after varying the shower scales, the h_{damp} matching parameter and the application of top quark p_{T} reweighting. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below. 108
71 The absolute cross section measurements for the N_{jets}, H_{T} and S_{T} kinematic event variables. They are compared against the POWHEG+PYTHIA simulation after varying the shower scales, the h_{damp} matching parameter and the application of top quark p_{T} reweighting. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below. 109

72 The absolute cross section measurements for the $p_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}^{\mathrm{W}}$, p_{T}^{ℓ} and $|\eta^{\ell}|$ kinematic event variables. They are compared against the POWHEG+PYTHIA simulation after varying the shower scales, the h_{damp} matching parameter and the application of top quark p_{T} reweighting. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below. 110

73 Double differential $\mathrm{t\bar{t}}$ production cross section measurements with respect to S_{T} and N_{jets}. The upper two panels show the distributions after full event selection for the e+jets and μ+jets channels respectively. The central two panels show the response matrices in terms of global bin number and the lower two panels show the normalised and absolute cross section measurements in comparison to the POWHEG+PYTHIA model. 115

74 Double differential $\mathrm{t\bar{t}}$ production cross section measurements with respect to p_{T}^{ℓ} and $|\eta^{\ell}|$. The upper two panels show the distributions after full event selection for the e+jets and μ+jets channels respectively. The central two panels show the response matrices in terms of global bin number and the lower two panels show the normalised and absolute cross section measurements in comparison to the POWHEG+PYTHIA model. 116

76 The upper two panels show the Feynman diagrams for the modified $\mathrm{t\bar{t}g}$ vertices and the lower panel the additional $\mathrm{t\bar{t}gg}$ vertex from the inclusion of the $\mathcal{O}_{\mathrm{tG}}$ operator. The red dot indicates the effective vertex. 118

Tables

1 The fermionic, spin 1/2, particles of the SM and their properties. They are split into three generations of leptons and quarks. 4
2 The bosonic, integer spin, particles of the SM and their properties. The force mediators are spin 1 and the H boson spin 0. 4

3 The set of simulated samples used in this thesis. They include four different signal $t\bar{t}$ production samples and samples for the background production. The backgrounds are single top quark production, vector boson production in association with jets and multijet QCD. 26
4 The set of samples associated with the uncertainties in the POWHEG+PYTHIA modelling. Other uncertainties are modelled by the use of event weights in the central POWHEG+PYTHIA simulation. 27

5 The sets of single electron and single muon data collected during 2016 split into individual runs. A total luminosity of 35.9 fb^{-1} of data was collected. 44

6 The identification criteria for signal electrons (tight) and veto electrons (veto). The critera are split into the barrel and endcap regions at $|\eta_{SC}| = 1.479$ 53
7 The identification criteria for signal muons (tight) and veto muons (loose). . . . 54
8 The identification criteria for signal jets. 54

9 The transfer factors to the signal region from the four multijet QCD control regions. 62
10 The yield of $t\bar{t}$ estimated from the data for each event variable, after the background has been subtracted and fakes removed. 69

11 A comparison of the bin width and the resolution in each bin for all event variables in the electon channel. 81
12 A comparison of the bin width and the resolution in each bin for all event variables in the muon channel. 81

13 The upper and lower bounds, in %, from each source of systematic uncertainty in the normalised differential cross section, over all bins of the measurement for each variable. The bounds of the total relative uncertainty are also shown. . . . 99

14 The upper and lower bounds, in %, from each source of systematic uncertainty in the absolute differential cross section, over all bins of the measurement for each variable. The bounds of the total relative uncertainty are also shown. . . . 100

15 Results of a goodness-of-fit test between the normalised cross sections in data and several models, with values given as χ^2/number of degrees of freedom (ndf). 111

16 Results of a goodness-of-fit test between the absolute cross sections in data and several models, with values given as χ^2/number of degrees of freedom (ndf). . . 111

17 Results of a goodness-of-fit test between the unregularised normalised cross sections in data and several models, with values given as χ^2/number of degrees of freedom (ndf). 112

18 Results of a goodness-of-fit test between the unregularised absolute cross sections in data and several models, with values given as χ^2/number of degrees of freedom (ndf). 112

CHAPTER 1

Introduction

For hundreds of years, the people of this world have tried to answer the most fundamental questions we can ask. What is consciousness? Why do we dream? Are we alone in the Universe? How did life begin? How did the Universe begin? What is the Universe made of and how does it work? These are not easy questions. In truth, we can not definitively answer any of them. This does not, or ever will, dampen our spirits in the quest to answer them.

The standard model of particle physics is our best attempt to describe our Universe on the smallest scale and how it evolves. It is an extremely successful, well tested model that describes the existence and behaviours of all the observed particles which form the constituents of the visible Universe. Indeed, it predicts accurately how particles interact at energy scales over many orders of magnitude; from simple magnetic compasses and powerful lasers to the high energy collisions experienced at current particle colliders.

It is, nevertheless, incomplete. It fails to include the simple action of an apple dropping in a gravitational field, it fails to describe more than the baryonic matter in the universe, a total of approximately 4% and it fails to describe why we live in such a matter dominated Universe. But these failures are seen as opportunities by physicists, searching for new particles and new interactions that can be used to describe Nature.

Perhaps one of the most exciting areas of the the standard model to look at in detail is that orientated around the top quark. The top quark is the most massive particle of the standard model and will decay without forming a hadron, and as such it allows us to study its bare quark properties. Its large mass also means that the top quark is sensitive to new physics, either by direct production or from interference produced from particles produced at higher energy scales. Aside from searching for new physics, the top quark appears in, and can be a major background to, many rare standard model processes. Observing these processes will further validate the standard model. Many of the possible new physics processes will have a direct effect on the top quark pair production cross section.

For these reasons and more, it is vital to know the production cross section to the highest degree of precision possible, not only for the inclusive cross section, but for differential cross sections too. The breakdown of cross sections as a function of a particular variable provides valuable information on how well the top quark pair system is modelled.

This thesis sets out to measure the differential cross sections of top quark pair production with respect to several kinematic event variables. It does so in the single lepton decay channel

of the top quark and presents the measurements in a phase space similar to that accessible by the Compact Muon Solenoid detector and with respect to detectable particles.

In Ch. 2, a brief summary of the standard model, its shortcomings and top quark physics is discussed. Chapter 3 discusses the modelling of interactions and lists the simulated datasets used in this thesis. From simulated to real interactions, Ch. 4 discusses the Large Hadron Collider, the Compact Muon Solenoid (CMS) experiment and the data collection performed. Once the raw data is collected, it must be reconstructed into usable analysis objects. This is discussed in Ch. 5. Once the analysis objects have been created, an analysis can be formed. Chapter 6 discusses the event selection performed, the variable definitions and the corrections which need to be applied to simulated and real events. The yield of top quark pair events reconstructed for each variable is calculated. Chapter 7 describes the unfolding procedure used to remove the detector response imprinted on the data and to present the results to particle level in the visible phase space. The unfolded yields are used in the calculation of normalised and absolute cross sections. The uncertainties, both statistical and systematic, are discussed in Ch. 8. The final results are presented in Ch. 9 along with goodness-of-fit tests performed between the unfolded data and our top quark pair production models taking into account the correlations of uncertainties between bins of the measurement. A study on the effect of regularisation on the unfolding is shown. Finally, in Ch. 10 the future prospects beyond this thesis are discussed.

It is worth noting that natural units are used in this thesis where $c = \hbar = 1$. This means that mass, momentum and energy are all measured using eV.

CHAPTER 2

Theoretical overview

The *standard model of particle physics* (SM) is an elegant attempt to describe the fundamental nature of the Universe. At the heart of the model are the base constituents of matter, the *elementary particles* and the interactions between them, the *forces*. The SM describes the current experimental measurements well across a large range of energy scales. It manages to unify forces that we can see in the macroscopic world to those binding together the constituents of protons. The SM is an incomplete model, however. It does not, for example, describe the gravitational interactions between particles, nor does it fully explain the matter-antimatter imbalance seen in the Universe. Previous discrepancies in the SM have lead to predictions and discoveries in Nature, such as the existence of three generations of quarks due to a lack of CP-violation observed. In this way, the SM has evolved over the half-century since it was first postulated, to incorporate new experimental evidence to become a more complete description of nature and it will continue to do so for many years to come. This chapter will discuss the particles which make up the SM and how they interact with each other, forming the complete model. Some drawbacks of the SM will then be highlighted, before moving on to the physics of top quarks and their cross sections.

2.1 The components of the standard model

The SM [3–7] contains twelve types of fermion, split into leptons (ℓ) and quarks (q), whose interactions are mediated by four types of gauge boson. The set of leptons exists in three generations: electrons, muons and tau particles (e, μ, τ), together with their associated neutrinos, (ν_{e}, ν_μ, ν_τ). The set of quarks (q) also exist in three generations: up and down (u, d), charm and strange (c, s) and the top and bottom (t, b) particles. For each fermion (f) that exists there is an anti-fermion ($\bar{\mathrm{f}}$), with identical quantum properties except conjugated charges. The properties of the fermionic particles are shown in Tab. 1 as listed by the Particle Data Group [8].

Particles in the SM can interact via the electromagnetic (EM), weak and strong forces. In particle physics, these forces are described by quantum field theories, in which the interactions between fermions are mediated by the exchange of a virtual boson. The EM force is mediated by the photon (γ), the weak force by the massive W and Z bosons and the strong force by a set of eight gluons (g).

Table 1: The fermionic, spin $1/2$, particles of the SM and their properties. They are split into three generations of leptons and quarks.

	Leptons			Quarks		
Generation	**Particle**	**Mass (MeV)**	**Charge (e)**	**Particle**	**Mass (MeV)**	**Charge (e)**
1	ν_e	$< 2 \times 10^{-6}$	0	u	2.2	$2/3$
	e	0.511	-1	d	4.7	$-1/3$
2	ν_μ	< 0.19	0	c	1280	$2/3$
	μ	105.7	-1	s	96	$-1/3$
3	ν_τ	< 18.2	0	t	173100	$2/3$
	τ	1777	-1	b	4180	$-1/3$

Table 2: The bosonic, integer spin, particles of the SM and their properties. The force mediators are spin 1 and the H boson spin 0.

Force Mediator	**Particle**	**Mass (GeV)**	**Charge (e)**
EM	γ	0	0
Weak	W	80.385	± 1
	Z	91.188	0
Strong	g	0	0
—	H	125.09	0

The last particle of the current SM is the Brout-Englert-Higgs (BEH) boson (H), which is an excitation of the BEH field. The interaction with the BEH field, via the BEH mechanism, is responsible for imparting mass to the fundamental particles [9–11]. The properties of the bosons are shown in Tab. 2 [8].

While the SM is able to describe three of the four fundamental forces between particles, it is not a complete model. It is unable to reconcile the act of gravity at a fundamental particle level. This is one of the discrepancies seen between the SM and nature, and must be one of the future evolutions of the SM. Other problems include the baryon asymmetry seen in the Universe, the requirement for dark matter and the hierarchy problem. These are discussed in Sec. 2.5.

2.2 Representing particles and their interactions

Describing the dynamics of particles and their interactions with other particles through the excitations of relativistic quantum fields are not trivial. To give simple visualisations of these complex interaction processes, Richard Feynman developed a powerful diagrammatic tool in 1948 [12, 13]. This diagrammatic tool is now commonly called the Feynman diagram. All Feynman diagrams in this thesis are created using the feyntikz package [14]. A complete Feynman diagram showing a complex example of top quark-antiquark ($t\bar{t}$) production and decay is shown in Fig. 1. In all Feynman diagrams, time increases from the left-hand side to the right, such that all the initial state particles are represented on the left and final state particles towards the right. Each arrowed line in the Feynman diagram represents a free fermionic particle of the standard model, each sinusoidal line represents an electroweak mediator particle (W, Z, γ) and each coiled line a gluon. Anti-fermions are represented as fermions moving backwards in time. The free fermions and anti-fermions are described by the Dirac equation, which can be obtained by substituting the Lagrangian density for a Dirac field into the Euler-Lagrange equation. The Lagrangian density for free fermions is given by:

$$\mathcal{L}_{\text{free}}(x) = \overline{\Psi}(x)(i\gamma^\mu \partial_\mu - m)\Psi(x), \tag{2.1}$$

Figure 1: A Feynman diagram showing the $t\bar{t}$ pair production by quark-antiquark annihilation and decay into the single lepton final state. Also shown are radiative processes in both the initial (ISR) and final state (FSR). Within the proton, there are also many QCD interactions occuring, such that the colliding partons in the initial state are described by the parton distribution function.

where $\Psi(x)$ is the fermionic field depending on space-time, γ^μ are the gamma matrices, ∂_μ is the partial derivative and m the fermion mass. The $\mathcal{L}_{\text{free}}$ is invariant under the global phase transformation,

$$\Psi(x) \to e^{iq\chi}\Psi(x), \tag{2.2}$$

which is a global U(1) symmetry, and by Noether's theorem there must be a conserved current and charge which can be identified with the current and charge [15]. When a local phase transformation is applied however,

$$\Psi(x) \to e^{iq\chi(x)}\Psi(x), \tag{2.3}$$

where the phase $q\chi(x)$ depends on space-time, the Lagrangian is no longer invariant, but contains the additional term:

$$\mathcal{L}_{\text{free}}(x) = \mathcal{L}_{\text{free}}(x) - q\overline{\Psi}(x)\gamma^\mu\partial_\mu(\chi(x))\Psi(x). \tag{2.4}$$

This violation is solved by the introduction of a new gauge vector field, A_μ, and coupling constant q into the Lagrangian density, via the *covariant derivative*

$$D_\mu = \partial_\mu + iqA_\mu. \tag{2.5}$$

This gauge field interacts with the fermion field Ψ cancelling the additional term and maintaining gauge invariance, provided the photon field is massless. The combination of the free fermionic fields, photon fields and the EM interaction term is known as *Quantum Electrodynamics* (QED), with the full Lagrangian density given as:

$$\mathcal{L}_{\text{QED}} = \overline{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + q\overline{\Psi}\gamma^\mu\Psi A_\mu, \tag{2.6}$$

where $F^{\mu\nu}$ is the EM field strength tensor with the term describing free photons.

2.3 Quantum chromodynamics

The quantum field theory for the strong interaction, *Quantum Chromodynamics* (QCD), is similar in many ways to that of QED. The strong interactions are highlighted in red on Fig. 2. It is described by an SU(3) symmetry group, with three orthogonal states known as colour

Figure 2: A Feynman diagram showing the $t\bar{t}$ pair production and decay. An example QCD interaction is shown in bold red. Other QCD interactions are shown in light red.

charges (r, g, b). To maintain invariance under the local SU(3) gauge transformation

$$\Psi(x) \to e^{i\theta^a(x)t^a}\Psi(x), \tag{2.7}$$

eight new gauge fields, G^a_μ, must be introduced in a similar manner to QED. These fields come from the eight generators of the SU(3) symmetry group t^a, which can be represented by the Gell-Mann matrices as $t^a = -\frac{1}{2}\lambda^a$. As with QED, the extra fields are folded into the covariant derivative, this time with a strong coupling constant g_s

$$D_\mu = \partial_\mu + ig_s\frac{\lambda_a}{2}G^a_\mu. \tag{2.8}$$

Thus the strong force is mediated by eight massless gauge bosons (gluons), however, in difference to QED, these gluons carry a colour charge brought about by the non-commutation of the SU(3) generators, leading to gluon-gluon self-interactions. The full QCD Lagrangian density is given as

$$\mathcal{L}_{\text{QCD}} = \overline{\Psi}_i([i\gamma^\mu D_\mu]_{ij} - m\delta_{ij})\Psi_j - \frac{1}{4}G^{\mu\nu}_a G^a_{\mu\nu}, \tag{2.9}$$

where the gluon field strength tensor

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu, \tag{2.10}$$

is analogous to the electromagnetic field strength tensor and the SU(3) fine structure constants f^{abc}, are introduced by the gluon-gluon self-interaction.

2.3.1 Colour confinement and hadronisation

In Fig. 3, the running of the strong coupling constant with energy is shown. At higher energies

Figure 3: The running of the strong coupling constant. At higher energy scales α_S tends to 0, resulting in a strengthening of the strong force. [16].

the strong coupling constant g_s (or $\alpha_S = g_s^2/4\pi$) tends to 0. This is known as asymptotic freedom and means that the strong force becomes stronger at larger quark or gluon separations. The quarks and gluons are consequently bound in colourless states (hadrons and mesons) in a process known as *colour confinement*. Single quarks in the final state of Fig. 1 bind into baryons within a time scale of $\tau_{\text{had}} \simeq 1/\Lambda_{\text{QCD}} \approx 3 \times 10^{-24}$ s. These particles can be highly energetic and if the potential energy due to the strong force confining a $q\bar{q}$ pair is large enough then a new $q\bar{q}$ pair will be produced from the vacuum. If the new daughter hadrons produced also have enough energy, the process is repeated until the confinement is strong enough to hold them together. The splitting of hadrons and mesons in this way is known as *hadronisation* and results in a spray of particles known in experimental terms as a *jet*. Similarly, when a hard interaction occurs within the colliding protons, the protons cease to become colourless states, and the remnants hadronise forming more sprays of particles. These particles are known as the *underlying event*. The hadronisation process will be discussed in more detail in Sec. 3.4.

2.4 The weak interaction and electroweak unification

The weak interaction differs from QED and QCD in that it is possible to violate parity conservation $\Psi(x) \neq \Psi(-x)$, and to change the flavour of a quark. Example weak interactions are highlighted in blue on Fig. 4. The weak interaction is described by an SU(2) symmetry group, with the conserved charge being the weak isospin T. The third component of weak isospin T_3, is also conserved. Parity violation is seen in the weak interaction as only left-handed particles and right-handed anti-particles have been observed to interact via the weak force. At higher energies, the electromagnetic and weak interactions unify into the electroweak (EWK) interaction. This combination results in a $\text{SU}(2)_\text{L} \otimes \text{U}(1)_\text{Y}$ symmetry group, with three gauge fields, W^i_μ, introduced by the $\text{SU}(2)_\text{L}$ generators $\frac{1}{2}\sigma_i$ (the Pauli spin matrices) and one gauge field,

Figure 4: A Feynman diagram showing the $t\bar{t}$ pair production and decay. An example weak interaction is shown in bold blue. Other weak interactions are shown in light blue.

B_μ, from the $\mathrm{U(1)_Y}$ generator (weak hypercharge $Y = 2(Q - T_3)$). The EWK gauge bosons are formed from the interferences of these fields and are formed as:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \tag{2.11}$$

$$Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W, \tag{2.12}$$

and

$$A_\mu = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W, \tag{2.13}$$

where θ_W is the weak mixing angle [3], defined by the ratio of the two weak coupling constants (isospin and isocharge)

$$\tan(\theta_W) = \frac{g'}{g_W}. \tag{2.14}$$

The full EWK covariant derivative and Lagrangian density are given by

$$D_\mu = \partial_\mu + ig_W \frac{\sigma_i}{2} W_\mu^i + ig' \frac{Y_W}{2} B_\mu, \tag{2.15}$$

and

$$\mathcal{L}_{\mathrm{EWK}} = \overline{\Psi} i\gamma^\mu D_\mu \Psi - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} - \frac{1}{4} W_i^{\mu\nu} W_{\mu\nu}^i, \tag{2.16}$$

respectively. If this is assumed to be an accurate representation of the EWK interaction, it requires that gauge bosons and fermions be massless in order to preserve the gauge symmetry. As it is experimentally proven that fermions and the W and Z bosons have mass, a solution is needed. The problem is solved through a process called *electroweak symmetry breaking*.

2.4.1 Electroweak symmetry breaking

Electroweak symmetry breaking solves the mass problem by keeping the particles massless but introducing a new complex scalar field doublet and potential into the Lagrangian density.

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.17}$$

$$\mathcal{L}_{\mathrm{H}} = (D_\mu \phi)^\dagger (D^\mu \phi) - \frac{1}{2}\mu^2 \phi^\dagger \phi - \frac{1}{4}\lambda(\phi^\dagger \phi)^2 \tag{2.18}$$

This potential introduces an infinite set of degenerate non-zero ground states of ϕ. The vacuum expectation value v, is the magnitude of an arbitrary ground state provided μ^2 is negative

$$v = \sqrt{-\frac{\mu^2}{\lambda}} \approx 246 \text{ GeV}. \tag{2.19}$$

The choice of v, conventionally along the neutral, real component of ϕ, *spontaneously* breaks the symmetry of the Lagrangian [8]. The neutral component is chosen because the charged component of ϕ must be set to zero such that the addition of a vacuum expectation value will not break the conservation of electric charge. By taking this choice, ϕ can be rewritten as a perturbation around this point

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{2.20}$$

where $H(x)$ is the massive BEH scalar field with

$$m_H = v\sqrt{2\lambda}. \tag{2.21}$$

The other three degrees of freedom from the initial scalar field doublet form the three massive weak bosons, via the BEH mechanism [9–11] such that:

$$m_W = \frac{v g_W}{2}, \quad m_Z = \frac{v\sqrt{g_W^2 + g'^2}}{2}. \tag{2.22}$$

Fermions may also be permitted mass via the Higgs mechanism, by introducing additional terms to the Lagrangian density of the form

$$\mathcal{L}_{\mathrm{Y}} = g_y(\overline{\Psi}_L \phi \Psi_R + \overline{\Psi}_R \phi^\dagger \Psi_L), \tag{2.23}$$

where g_y (*Yukawa coupling*) is the coupling strength of the respective fermion to the BEH field. The Yukawa coupling strength is different for each fermion. This leads to fermion masses which are linearly proportional to and only dependent on g_y, given by

$$m_f = \frac{v g_y}{\sqrt{2}}. \tag{2.24}$$

2.4.2 The Cabbibo-Kobayashi-Maskawa mixing matrix

The weak interaction of quarks is described by the Cabbibo-Kobayashi-Maskawa (CKM) mixing matrix [17, 18]. It postulated the existence of a third family of quarks (t, b) to explain the observed CP violation in kaon decays [18]. It relates the weak, flavour eigenstates to the mass eigenstates by

$$\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \\ \mathrm{b}' \end{pmatrix} = \begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix}.$$

The relative strength of the weak interaction involving a flavour transition from $j \to i$ is given by the associated CKM matrix element V_{ij}. In the SM the CKM matrix is unitary ($V^\dagger V = I$) so that each row behaves as $|V_{\mathrm{ud}}|^2 + |V_{\mathrm{us}}|^2 + |V_{\mathrm{ub}}|^2 = 1$. This unitarity assumption leads to current experimental measurements of

$$\begin{pmatrix} |V_{\mathrm{ud}}| & |V_{\mathrm{us}}| & |V_{\mathrm{ub}}| \\ |V_{\mathrm{cd}}| & |V_{\mathrm{cs}}| & |V_{\mathrm{cb}}| \\ |V_{\mathrm{td}}| & |V_{\mathrm{ts}}| & |V_{\mathrm{tb}}| \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix},$$

given by [8]. The relative strength of the off-diagonal CKM elements means that flavour transitions between different generations of quarks are suppressed. There is a small complex phase that is present in the V_{ub} and V_{td} matrix elements, which allows for a small quantity of charge-parity (CP) violation in the SM.

2.5 The standard model and its shortcomings

The full SM combines the QCD and EWK field theories into the combined symmetry group $\mathrm{SU(3)_c} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$. This has the effect of making the full covariant derivative

$$D_\mu = \partial_\mu + ig_S \frac{\lambda_a}{2} G_a + ig_W \frac{\sigma_i}{2} W_\mu^i + ig' \frac{Y_W}{2} B_\mu, \tag{2.25}$$

and the full Lagrangian density the sum of all individual components

$$\mathcal{L}_{\mathrm{SM}} = \mathcal{L}_{\mathrm{QCD}} + \mathcal{L}_{\mathrm{EWK}} + \mathcal{L}_{\mathrm{H}} + \mathcal{L}_{\mathrm{Y}}. \tag{2.26}$$

It includes terms for free fermions and gauge bosons and the interactions between them, in addition to mass terms arising from the introduction of the BEH field and a non-zero v. It is the most complete model we have of our Universe, describing fundamental matter and its interactions to an unprecedented degree of precision, however as mentioned earlier, it is far from a complete description.

2.5.1 Gravity

Perhaps the largest omission from the SM is the fundamental force of gravity, which all massive particles experience. In order to include a description of gravity in the SM, a unification with the theory of general relativity at the quantum scale is needed. The formulation of this unification at present is unknown, however one popular way to do this is with string theory. Unfortunately, the effect of gravity is completely negligible at the current accessible energy scales and as such it is not experimentally feasible to directly test any unification theory.

2.5.2 Massive neutrinos

The current observation that only left-handed neutrinos interact weakly results in the Yukawa component of the Lagrangian density shown in Eq. 2.23, losing gauge invariance unless the neutrinos are massless. Experiments such as SNO which measured a third of the number of ν_{e} expected from production in the Sun [19], or Super-Kamiokande which measured half of the number of upward atmospheric ν_μ as downward ones [20], prove an oscillation in the flavour of the neutrino. This can be explained by the neutrinos acquiring a mass, such that the flavour states of the neutrinos are then linear superpositions of the mass states.

2.5.3 Dark matter

Other issues with the SM originate from cosmological observations. Firstly, the lack of enough mass in galaxies to explain the rotation curves of luminous matter, provides evidence for a type of additional matter which must be massive and electromagnetically inert [21]. This matter is known as dark matter and is predicted to contribute around 80% to the matter content of the Universe. There are other cosmological indicators for dark matter including galaxy clusters [22] and the cosmic microwave background [23]. While the current SM does provide a set of particles which behave in this way, the neutrinos, there are not enough to account for all

of the missing matter and they are too 'hot', i.e. their highly relativistic velocity is unable to explain observations.

The second issue is that the observation of the accelerating expansion of the Universe implies that there is a repulsive operation acting [24]. This process is enshrined under the term dark energy, which accounts for 70% of the mass-energy content of the Universe. The SM has no mechanism for describing dark energy.

2.5.4 Baryon asymmetry

Another striking problem in the SM is the matter-antimatter imbalance seen in the Universe. The Universe is predominantly matter based, however the SM predicts that matter and anti-matter are produced in equal quantities. Somehow, the antimatter present at the beginning of the Universe has been transformed or destroyed and therefore a process is required which favours matter over anti-matter [25]. While the SM does provide an explanation for some of this transformation in the lepton sector, from the CP violating, complex phases of the quark and neutrino mixing matrices, it contributes negligible amounts to the imbalance seen.

2.5.5 Hierarchy problem

The mass of the H boson depends on its bare mass and corrections from fermion loops depending on the square of the energy scale Λ.

$$m_{\mathrm{H}}^2 = m_{\mathrm{H,bare}}^2 + \mathcal{O}(\Lambda^2)$$

At the EWK scale (100 GeV), these quantum loop corrections are small, however as the energy scale approaches the Planck scale (10^{19} GeV), the point at which the SM is no longer valid, the corrections diverge. The very large divergences must be cancelled to keep the observed H boson mass light. This requires a significant amount of fine-tuning in the SM. One way this unnatural amount of fine-tuning can be alleviated, is to introduce an additional set of particles which, in part, cancel the loop corrections. *Supersymmetry* is one example of a theory that creates an additional set of particles by creating a symmetry of spin in the SM [26]. For every SM particle there exists a massive supersymmetric particle, differing by a half unit of spin, which can be present in quantum loop corrections to the H boson mass at higher energy scales, cancelling contributions from the SM particles.

2.6 Top quark physics

The discovery of the top quark was published in 1995 from a combination of the results from both the CDF and D0 experiments [27, 28]. It is the most massive particle in the SM to date, at 172.5 GeV. The large mass means that the lifetime of the top quark is very short at $\tau_{\mathrm{t}} = {}^{1}/_{\Gamma_{\mathrm{t}}} \simeq 5 \times 10^{-25}$ s [8]. In fact, it is shorter than the typical time it takes for a quark to hadronise, $\tau_{\mathrm{had}} \approx 3 \times 10^{-24}$ s so that it is possible for the bare quark properties to be measured directly.

Now, at the LHC, more top quarks are being produced than ever before, leading to a vast array of measurements being performed, such as precision measurements of the SM, searches for rare SM processes and searches for possible new physics present in top quark decays. Not only is top quark physics important in direct searches for new physics, it is often the process which forms the largest background in other searches for particles beyond the SM. Therefore, it is crucial to know the production cross section of the top quark as precisely as possible. The precise top quark production cross section measurements can also uniquely be used to directly

probe the V_{tb} element of the CKM matrix ($V_{tb} >> V_{ts} > V_{td}$), to set constraints on the gluon and quark *parton distribution functions* (PDFs) at high $M(\mathrm{t\bar{t}})$ and to measure the top quark Yukawa coupling strength.

2.6.1 Top quark production

Protons are made up of three valence quarks (uud), confined by gluons. At higher energies, the gluon multiplicity increases, as well as the number of pair-produced $\mathrm{q\bar{q}}$ pairs, known as sea quarks. This means that it is much more probable that an interaction will occur from collisions between gluons or non-valence quarks. Top quark production at the LHC at $\sqrt{s} = 13$ TeV proceeds primarily via $\mathrm{t\bar{t}}$ pair-production, where $\sim 90\%$ is from gluon-gluon fusion. The further $\sim 10\%$ comes from quark-antiquark annihilation. The quark-antiquark production mode is shown in Fig. 1. The top quark can also be produced singly via the s- and t- channels and in association with a W boson, as shown in Fig. 5. The production cross section for single top quarks is much smaller than that for $\mathrm{t\bar{t}}$ production because of the composition of the initial states required. Higher top multiplicity final states are also available in the SM but with a much smaller production cross section [29].

Figure 5: Feynman diagrams for the production of single top quarks. The left panel shows single top quark production via the s-channel, the central panel by the t-channel and the right panel in association with a W boson.

The production cross section for the interaction $\sigma_{pp\to\mathrm{t\bar{t}}}$, is predicted by the convolution of the partonic cross section $\hat{\sigma}_{\mathrm{t\bar{t}}}$ derived directly from the Feynman diagrams and the PDFs of the incident protons $f(x, \mu^2)$, summed over all possible parton flavours. This is given by

$$\sigma_{pp\to\mathrm{t\bar{t}}} = \sum_{a,b} \int_0^1 f(x_a, \mu^2) dx_a \; . \int_0^1 f(x_b, \mu^2) dx_b \; . \; \hat{\sigma}_{ab\to\mathrm{t\bar{t}}}. \tag{2.27}$$

The parton distribution function is defined as the number density of a parton carrying a fraction x of the total momentum of the proton in the longitudinal direction at the energy scale being probed, μ. The PDFs are measured from global fits to many different data measurements and sources, such as measurements of deep inelastic scattering, Drell-Yan production, $\mathrm{t\bar{t}}$ production and inclusive jet production, from many experiments at colliders such as Hadron-Electron Ring Accelerator and the LHC. The PDF is calculated at some arbitrary $\mu^2 = \mu_0^2$, which is taken as 1 GeV by the NNPDF collaboration in the NNPDF3.1 PDF set [30]. This is then evolved through the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [31–33] to higher energy scales. The $xf(x, \mu^2)$ distributions for the NNPDF3.1 PDF set at $\mu^2 = 10$ GeV^2 and $\mu^2 = 10^4$ GeV^2 are displayed in Fig. 6. When moving from low to high μ^2 values, it is clearly seen that a greater fraction of the momentum of the proton is carried by an increasing number of non-valence partons.

Figure 6: The parton distribution function of the proton scaled by the fraction of momentum carried by the parton, $xf(x, \mu^2)$ with respect to the fraction of momentum carried by the parton. The left panel shows the distributions with respect to a lower energy scale and the right panel with respect to a higher energy scale [30].

The prediction of the inclusive $t\bar{t}$ cross section from proton-proton collisions is shown in Fig. 7, together with current experimental measurements. The prediction from proton-anti-proton collisions is also shown. The cross sections converge at higher energies due to the increase in collisions between non-valence partons the distribution of which is identical in protons and anti-protons. At $\sqrt{s} = 13$ TeV, the inclusive $t\bar{t}$ production cross section was calculated to be $831.8^{+19.8}_{-29.2}(\text{scale}) \pm 35.1(\text{PDF} + \alpha_S)$ pb [34]. This $t\bar{t}$ cross section is calculated to next-to-next-to-leading-order (NNLO) accuracy in perturbative QCD, including resummation of next-to-next-to-leading logarithmic soft-gluon terms with Top++ (v2.0) [35–41].

2.6.2 Order of perturbation

The cross section of a process depends on a quantity known as the *matrix element*

$$\hat{\sigma}_{t\bar{t}} \propto |\mathcal{M}|^2$$

and the matrix element can be calculated using the Feynman rules from Feynman diagrams. Increasing the order of perturbation essentially adds an additional parton to each diagram which makes the matrix-element calculation much more complex but gives increasing accuracy. A comparison of the Feynman diagrams at leading order (LO) and next-to-leading order (NLO) is shown in Fig. 8, where in the top left panel a simple LO tree-diagram is shown. The top right panel shows loops in the propagator which can be related to the running of α_S. The rest of the panels show NLO processes, where the centre two panels are displaying loops corresponding to virtual particles and the lower two panels contribute real partons to the final state.

Perturbation theory is valid at high energy scales, however at lower energy scales QCD production becomes non-perturbative and the cross section is not able to be calculated using the matrix-element. The method of dealing with non-perturbative QCD is dealt with in Sec. 3.2.

Figure 7: The inclusive $t\bar{t}$ cross section, measured by ATLAS, CMS and the Tevatron in different decay channels at multiple $\sqrt{s}$. Also shown is the most precise theoretical prediction to date [42].

Figure 8: A set of Feynman diagrams showing LO and NLO processes. In the top left panel, the leading order diagram is shown. The top right panel shows a gluon-loop in the propagator. These loops are responsible for the running strong coupling constant. The two middle panels show NLO loops providing virtual particles to the diagrams and corrections in the renormalisation scale. Finally, the bottom two panels show the real additional partons that can be seen in the final state due to NLO matrix-element calculations.

2.6.3 Top quark decay

In the SM, the top quark will decay 99.9% [8] of the time to a b quark and W boson. The W boson then subsequently decays either into a $q'\bar{q}$ pair (hadronic) or a $\ell\bar{\nu}_\ell$ pair (leptonic). In the case of $t\bar{t}$ production, this leads to three classes of final states, the first where both W bosons decay leptonically, the second where both decay hadronically and lastly where one decays leptonically and the other hadronically. These are known as the *dilepton*, *all-hadronic* and *single lepton* decay channels respectively. The probability of decaying to a specific final state (*branching ratio*) is approximately 10% for the dilepton channel and 45% for the hadronic and single lepton channels.

Each final state category has merits and drawbacks for use in analyses. The dilepton final state provides the most clean experimental signature, however this comes at the cost of a reduced branching ratio. It is difficult to fully reconstruct the whole event as both neutrinos contribute to the missing transverse momentum. The all-hadronic final state in comparison has a large branching ratio and a large multijet QCD background. The high jet multiplicity can also cause problems from the combinatorics when reconstructing the $t\bar{t}$ system. While the all-hadronic final state is difficult at lower collision energies, it gains in sensitivity when analysed at high collision energies. This is because the decay products of the top quark collimate and merge into a single jet, which is able to be tagged as coming from a top quark, through jet-substructure techniques. Finally the single lepton decay channel has a large branching ratio, but with much reduced backgrounds with respect to the all-hadronic final state. The final states in which a tau particle is produced are often neglected however, due to the added complexity of reconstructing them. More details about the reconstruction of particles can be found in Ch. 5.

2.6.4 Top quark cross section measurements

Figure. 7 shows, in addition to the theoretical prediction of the inclusive $t\bar{t}$ production cross section, the corresponding experimental measurements, from both the CMS and A Toriodal Large ApparatuS (ATLAS) at different centre-of-mass energies ($\sqrt{s}$) in all of the channels previously stated. At CMS, the inclusive $t\bar{t}$ cross section measurements are performed at 13 TeV [43,44] at 7 and 8 TeV [45–56] and at 5 TeV [57]. Inclusive cross section measurements have been performed for $t\bar{t}$ production in association with a photon [58], vector boson [59] and most recently H boson [60].

Measurements of the $t\bar{t}$ production cross section with respect to some other variable are known as *differential* cross section measurements. Differential cross section measurements are especially useful for testing the understanding of state-of-the-art theoretical models. Ideally, these models should describe $t\bar{t}$ production with respect to all kinematic variables accurately, including contributions from additional jets, and should do so for all $\sqrt{s}$. In practice this is not the case, with the $t\bar{t}$ models being *tuned* to best describe the kinematic distributions. A tune is a complete set of simulation parameters describing physics processes within the simulation of the model, which have been varied to best describe differential data distributions. Figure. 9 shows the effect of this tuning on a set of $t\bar{t}$ models, where the new tune is shown in solid and the old in dashed, when comparing the cross sections as a function of the magnitude of the transverse momentum of the hadronically decaying top quark, $p_{\mathrm{T}}(\mathrm{t}_h)$, and the additional jet multiplicity. It is worth noting, that a tune applied to a specific model may not describe the differential data better than when applied to another independent model. The different models behind $t\bar{t}$ production are explained in more detail in Ch. 3. Indeed by changing the tune, it may cause the same model to describe some kinematic distributions to a worse degree, while improving other parts of the phase space.

Differential top quark cross section measurements can be presented to *particle level* where the kinematic distributions are constructed with respect to stable particles in the detector

Figure 9: Comparison of different $t\bar{t}$ models with the old CUETP8M1 tune, represented by dashed lines, and the new CUETP8M2T4 tune, given by solid lines. The left panel shows the comparison for the distribution of the p_T of the hadronically decaying top quark and the right panel for the additional jets distribution [61].

(mean lifetime longer than 30 ps). Alternatively, they can be presented to *parton level*, where the results are extrapolated with respect to the final state partons or to *detector level*, where no extrapolation is performed. Similarly, the results can be presented in a phase space similar to that accessible by the detector, the *visible phase space* or extrapolated to the *full phase space*. Measurements that are presented to parton level or that are extrapolated to the full phase space are influenced by large theoretical uncertainties due to that extrapolation and as such most analyses present measurements to particle level in the visible phase space. All the $t\bar{t}$ cross section measurements performed by the CMS (and ATLAS) experiment are completely complementary to each other and are presented at 7 and 8 TeV [62–65] and at 13 TeV [1,66–69].

This thesis looks at differential cross section measurements as a function of kinematic event variables in the single lepton channel presented to particle level in a visible phase space. The kinematic event variables are variables that do not require the reconstruction of the complete $t\bar{t}$ system. The event variables considered are the jet multiplicity, N_{jets}, the scalar sum of the jet p_T, H_T, the scalar sum of the p_T of all particles, S_T, the magnitudes of the transverse momentum imbalance, p_T^{miss}, the p_T of the leptonically decaying W boson, p_T^W, the p_T of the lepton, p_T^ℓ and lepton pseudorapidity $|\eta^\ell|$. These event variables are defined in Sec. 6.1. Measurements with respect to kinematic event variables have been performed at 7 and 8 TeV [70] and 13 TeV with a much smaller data set [2].

CHAPTER 3

Modelling proton-proton interactions

For any measurement in particle physics, sets of simulated samples need to be generated in order to compare the experimental data to theoretical models. These models include well known SM decay processes such as $t\bar{t}$ production, much rarer processes predicted by the SM such as $t\bar{t}H$ and in the case of a search for new physics a plethora of processes predicted by BSM theories.

These different processes are modelled by event generators in a similar manner, following a sequence of discrete steps. First of all, a possible hard scattering interaction is generated for a process at LO, or if possible NLO. If short-lived resonances are formed, their associated decays are also viewed as part of the hard process. Additional parton interactions can be included at LO and soft radiation added to the initial and final states, in a process known as the *parton shower*. The parton shower is then hadronised into colourless states and decayed, with the proton remnants forming the underlying event. Figure 10 shows diagrammatically an example simulated collision, where the hard scattering process is shown in red, the soft radiative processes in blue, the additional parton interactions in purple, the fragmentation in light green and hadronisation and decay in dark green. Photon radiation is shown in yellow. Finally, the interactions of the particles produced with the detector can be modelled and the detector response applied.

3.1 Hard interaction

The hard scattering process describes the interaction of two partons in the collider. It is modelled through Eq 2.27, with the PDFs taken from data convoluted with the partonic cross section. The partonic cross section is calculated from the matrix-element of the interaction, directly taken from the Feynman diagrams through the Feynman rules. If a short-lived resonance is formed as a result of the interaction, then its decay is also considered as part of the hard interaction. Many matrix-elements can already be calculated to NNLO or higher, however the generation of the hard scattering processes can currently only be calculated to LO or NLO. This is simply due to the combinatorics when taking into account all possible Feynman diagrams.

There are a variety of matrix-element generators available, two of the most popular in high energy particle physics being MadGraph5_aMC@NLO (MG5_aMC@NLO) [29] and POsitive Weight Hardest Emission Generator (POWHEG) [72–74]. The MG5_aMC@NLO matrix-

Figure 10: An example simulation of a proton-proton collision showing all the various processes that must be modelled. Starting with the hard interaction in red calculated using perturbation theory, then dressing with additional soft processes in blue forming the parton shower. Additional parton interactions are added, shown in purple, each with a respective parton shower. Finally the parton shower is fragmented and hadronised into detectable, colourless hadrons, which can in turn decay and radiate. This fragmentation and hadronisation processes are shown in light green and green respectively. Photon radiation is shown in yellow. Figure taken from [71].

element generator can model a process with additional partons. The number of additional partons in the process depends on the complexity of the process, e.g. for NLO $t\bar{t}$ production, an additional two partons can be generated. It does this by adding real diagrams based on LO or NLO contributions from lower order processes. In POWHEG, virtual processes are included in the matrix element, up to one loop. This means that for $t\bar{t}$ production, one less additional parton is able to be produced.

3.2 Additional interactions and soft processes

Once the hard process has been simulated from the matrix-element calculations, additional hard, low-p_T, interactions need to be included. All the hard interactions then present in the simulation must be *dressed* with soft emissions.

The most common hard process at the LHC is elastic gluon-gluon scattering, the cross section of which, diverges as the gluons become soft and collinear. This means that below some momentum transfer threshold (~ 5 GeV) the inclusive jet production cross section from elastic gluon-gluon scattering is larger than the total inclusive proton-proton cross section [75, 76], which infers that for every collision, more than one interaction is occurring. These are known as *multiple parton interactions* (MPI). The number of additional interactions present is proportional to the impact parameter of the colliding protons. A pair of protons colliding head on contains more additional interactions than a pair that collide obliquely.

The extra soft emissions come in the form of *initial* and *final state radiation* (ISR and FSR), where partons split from the hard interactions. Figure 11 shows the four possible ways in which a parton can be radiated. Firstly, a quark can radiate a gluon, secondly a gluon can pair-produce a $q\bar{q}$ pair and finally there are two colour configurations of gluon splitting.

Figure 11: The four possible processes to strongly split a parton in the parton shower. The top left panel shows gluon emission, the top right shows $q\bar{q}$ pair production and the bottom two panels the colour configurations of gluon splitting. In addition, photon brehmsstrahlung from the quarks can occur but is not considered here.

For FSR, the partons produced from the matrix-element calculation are split recursively until the energy of each remaining parton reduces to approximately 1 GeV, whereupon hadronisation effects become non-negligible. In contrast, the ISR is modelled differently. The longi-

tudinal momentum fractions, x_1 and x_2, of the incoming partons are taken from the PDF at the factorisation scale of the process and evolved backwards in time. The factorisation scale of an event is a cutoff where partons that carry a momentum less than the cutoff scale are absorbed into the PDF. If the momentum carried is above the scale then they participate in the interaction. This ensures no divergences in the matrix-element due to collinear or soft gluons. For every ISR emission, the parent parton must have the combined energy of the daughter parton and radiation. There are however, fewer available partons carrying the increased fraction of the proton momentum, seen in the PDF distribution, and so a suppression weight must be applied. The suppression weight is derived from the *Sudakov form factors* [77]. In practice, the application of ISR and MPI occur completely and in tandem, before the FSR is modelled.

The collective spray of partons produced by the hard process, MPI, ISR and FSR is called the parton shower. The evolution of the splittings in the parton shower is defined by ordering variables. One such shower ordering variable is the square of the p_{T} of the shower propagators (k_{T}^2), such that k_{T}^2 for a daughter splitting is less than the k_{T}^2 for a mother splitting [78]. The splittings between MPI can also be interleaved with one another. The k_{T}^2 ordering is illustrated in the left panel of Fig. 12, where the hard interaction occurs at Q_{max} and the splitting strengths, given by $V_i(k_{\mathrm{T}}^2)$, are ordered $V_1 > V_2 > V_3$ (The further in time from the hard interaction, the softer and more collinear the splitting).

Another way to evolve the shower is by using an angular ordering [79]. This angular ordering requires any future splitting to occur within a cone defined by the angular radius between the daughter partons from the previous splitting also illustrated in the right panel of Fig. 12. These showers often produce wide-angle soft splittings before the main hard splitting. Two widely used models for the parton shower are PYTHIA [80] and HERWIG++ [81], where PYTHIA uses a k_{T}^2-ordered shower and HERWIG++ an angular-ordered shower.

Figure 12: The left panel shows parton shower ordering by k_{T}^2 where the splitting strength $V_i(k_{\mathrm{T}}^2)$ is ordered $V_1 > V_2 > V_3$. The right panel shows shower ordering by angle such that the opening angle decreases for every consecutive splitting $\theta_{max} > \theta_1 > \theta_2 > \theta_{3,4}$.

3.3 Matching matrix-elements to parton showers

Parton showers and higher-order matrix-element calculations describe the same physics and so need to be combined consistently. The soft emissions added from the parton shower should interface with the matrix-element based radiation with no gaps or overlaps for a specific, fixed final state parton multiplicity. This problem is solved by a process known as *matching*.

One method of matching, known as the MLM-method [82], after its author M. L. Mangano, matches the final state partons produced by the matrix-element calculation to parton jets

created by clustering the showered partons. The angular distance from each final state parton to each parton jet, starting from the parton with the leading p_{T}, is calculated and is said to be matched if it is less than a threshold value. If the matching is successful the parton jet is removed from further calculations. All final state partons must be matched to parton jets, otherwise the event is rejected. When simulating the hard interaction process, including contributions with up to N additional partons, events in samples simulated with n additional partons, where $n < N$, are suppressed if they do not contain exactly n matched parton jets. This is because these events are present and more accurately modelled in the $n+1$ production sample and thus double counting is avoided. All the production samples generated using discrete additional parton multiplicities are subsequently *merged*.

When matching and merging at NLO the additional parton jet from the matrix-element calculation is the hardest emission and thus defines the upper shower scale level. While it is easy to interface this with showers ordered to some p_{T}, care must be taken with respect to an angular-ordered shower. Angular-ordered shower emissions with a p_{T} greater than the additional matrix-element parton jet need to be suppressed. The FxFx-method [83], extends the MLM-method of merging to matrix-elements calculated to NLO. The MLM-method is used when matching LO MG5_aMC@NLO to Pythia and the FxFx-method is used when matching NLO MG5_aMC@NLO to Pythia.

When Powheg models processes at a large factorisation scale, the number of high-p_{T} soft emissions is overestimated. It is resolved by introducing a p_{T} dependent damping factor,

$$D = \frac{h_{\mathrm{damp}}^2}{p_{\mathrm{T}}^2 + h_{\mathrm{damp}}^2},$$

where h_{damp} is the damping parameter, usually set to the factorisation scale (in the case for top quark production to m_t). D effectively sets the upper-p_{T} limit for the first additional emission and hence the upper shower scale level.

3.4 Fragmentation and colour reconnection

The fragmentation process models how final state partons from the parton shower hadronise into colourless states. There are two general methods in use for the fragmentation, the *Lund string model* [80, 84] and the *cluster model* [81].

3.4.1 Lund string model

Due to the self interactions of gluons, the strong field between two quarks can be represented by a confined tube of dimension $1+1$, otherwise known as a string. As the quarks move apart the potential energy in the string increases linearly by 1 GeV fm^{-1}. At some point the colour field contains enough potential energy that it is favourable to pair-produce light quarks, which combine with the initial quarks to form two new colourless particles, breaking the string. The string is stretched between quark endpoints by a number of gluons. Each gluon results in a kink to the string that depends on the kinematic properties of the gluon. The left panel of Fig. 13 shows the kink in a string produced by the addition of a gluon and the right panel shows the fragmentation of the string. As time evolves, if $\mathrm{q\bar{q}}$ pairs do not have enough potential in the colour field to pair-produce they end up bound to each other forming the colourless mesons of the final state. It is also possible for the potential in the colour field to be great enough to produce diquark-diantiquark pair instead of $\mathrm{q\bar{q}}$ with the end result being the formation of baryons instead of mesons. This approach is used in the Pythia parton shower simulation.

The string dynamics currently assume that the colour-flow modelled in the evolution of the parton shower from the hard interaction is independent from the colour-flow of the other

Figure 13: The left panel shows the string between the quark-antiquark colour dipole stretched by a gluon and the right panel shows for each string how fragmentation occurs with time. Lorentz boosts in the hadrons are shown by oscillation frequency. Figures taken from [85].

parton showers produced by the MPIs. Indeed, each individual MPI is assumed to operate in independent colour spaces. In reality, all the individual interactions are superimposed on each other, with soft gluons connecting all the final state partons from all sources, redistributing colour across the full parton shower. In the Lund string model this can have the effect of reorganising the strings produced in the parton shower. This is modelled in simulation by a process known as *colour reconnection* (CR), which reconfigures the colour strings after the parton shower, but before the fragmentation. Three different models which can be implemented using PYTHIA are described here.

MPI-based model

The *MPI-based* CR model [86] calculates the reconnection probability for all gluons from lower-p_{T} interactions to be inserted into the colour-flow dipoles of a higher-p_{T} interaction. Once all reconnections between MPIs have been identified, the configuration in which the total colour string length is minimised is selected. The PYTHIA parton shower simulation uses the MPI-based CR scheme as default, where it is assumed that the lifetime of the top quark is long enough to shield the top decay products from the effects of CR. A second MPI-based model exists where the top quark decays before the CR is applied and as such the top decay products are able to reconnect. This is known as MPI-based CR with *early resonance decays* (erdON).

QCD-inspired model

An alternative CR model is the *QCD-inspired* model [87]. In this model, instead of the SU(3) colour indices (r, g, b) being applied to individual partons, a set of nine indices, which describe the possible colour states of 2-parton and 3-parton combinations, are used. A corresponding set of anti-colour indices is used for anti-quarks and a gluon contains one of each index. Reconnections are possible when two partons have a matching colour and anti-colour index. The set of possible reconnections encompasses every possible reconnection from the default MPI-based CR model as well as reconnections possible between partons with accidentally matching colour indices. As with the MPI-based CR model, a reconnection is performed if it minimises the total QCD string length.

Gluon move/flip model

The *gluon move/flip* model [86], moves all final state gluons between colour strings so that the total string length is minimised. However, by moving only gluons, no quarks are able to

reconnect and so a subsequent procedure is applied in which two colour segments are flipped between colour strings. As with moving the gluons, the solution which minimises the total string length is used.

3.4.2 Cluster fragmentation model

In the cluster model of fragmentation implemented in the HERWIG++ parton shower simulation, all partons are showered until they reach an energy scale of ~ 4 GeV, where hadronisation effects become non-negligible. The final state gluons must have an energy at least twice that of the invariant mass of the lightest quark and are forced to split into a $q\bar{q}$ pair. The colour singlet states formed by the resulting colour connected partons are formed into clusters, as shown in Fig. 14. The cluster model assumes the principle of *colour-preconfinement* which states that the distribution of the mass of the clusters is independent of the centre-of-mass energy and the hard scattering process involved. Most clusters are distributed at low masses and can be regarded as excited hadron resonances and decayed into observed hadrons, however some clusters are too massive for this to be realistic and split in a process called *cluster fission*. If fission is required, a $q\bar{q}$ pair of u, d or s flavour, is taken from the vacuum and the cluster split such that one of the original quarks is contained in each new cluster. Fissions continue until all clusters are accepted by the mass parameters of the model. The final process of the fragmentation involves extracting a $q\bar{q}$ or $qq\bar{q}\bar{q}$ from vacuum in each cluster, forming a pair of hadrons, which in turn may decay into lighter hadrons.

Figure 14: A sketch of the cluster model of fragmentation and hadronisation. Colour singlet clusters are generated from the final state quarks and gluons forced to split into a q pair. These colour singlet clusters can be unnaturally massive and fission occurs until the clusters represent excited hadron resonances which then hadronise. Figure adapted from [88].

3.5 Detector modelling

So far, only the modelling of the physics process itself has been discussed, before any interaction with the detector has been performed. Indeed, it is very useful to compare measurements at this level, the *particle level*, as it allows direct comparisons between theoretical models and experimental measurements which have had the detector response removed. The removal of the detector response is discussed in Ch. 7. This means measurements performed to particle level with different detectors such as CMS and ATLAS can be directly compared with each other and to theoretical predictions.

In order to provide the mapping between experiment and particle level, the detector response needs to be modelled accurately. To do this a simulation toolkit called *GEometry ANd Tracking* (GEANT4) [89] is used. The GEANT4 package builds the physical layout of the experiment, including all detectors, absorbers, electronics and structural components surrounding the central interaction point. The magnetic field strength at all points in the detector is modelled. The particles are tracked from the collision point outwards through the active and non-active material, modelling respective interactions and decays. In the active material, the response of all individual components is modelled in a process known as *digitisation*. The digitisation takes into account any inefficiencies in the active material and returns an output consistent with that seen from measured data. A full description of the CMS detector is given in Ch. 4.

3.6 Production of $t\bar{t}$ and background models

Table 3 shows the complete set of simulations used for the differential cross section analysis performed. There are four different $t\bar{t}$ production models studied. Two $t\bar{t}$ samples are simulated with the POWHEG (v2) matrix element generator [90], where one sample is interfaced with PYTHIA (v8.212), using the CUETP8M2T4 tune [61], and the second with the HERWIG++ (v2.7.1) using the EE5C tune [91]. These are labelled POWHEG+PYTHIA and POWHEG+HERWIG++ respectively. Two additional simulated $t\bar{t}$ samples are produced with the MG5_aMC@NLO (v2.2.2) generator. One is used to generate events at LO accuracy with up to three additional partons and the second to NLO accuracy with up to two additional partons. Both are interfaced with PYTHIA using the CUETP8M1 tune [92] and are labelled MG5_aMC@NLO-LO and MG5_aMC@NLO-NLO respectively. The MLM jet-parton matching algorithm is used to match the matrix-element to the parton shower in the LO simulation and the FXFX algorithm in the NLO simulation. The POWHEG+PYTHIA model is taken as the central model used in this thesis.

All $t\bar{t}$ production samples are simulated with the top quark mass set to 172.5 GeV. For NLO samples the NNPDF30_nlo_as_0118 PDF set [93] is used, while for LO samples the NNPDF30_lo_as_0130 set is used. The samples are normalised to a NNLO $t\bar{t}$ cross section of 832 pb, as seen in Sec. 2.6.1. It is important to compare multiple $t\bar{t}$ generators in order to find the current most suitable description of top quark production and decay, and to identify any discrepancies in the models.

Table 3 also shows the dominant background samples from single top quark production, vector boson production with associated jets and multi-jet QCD production. All the single top quark processes produced via the t-channel and in association with a W boson are generated with POWHEG [94, 95] interfaced with PYTHIA and are normalised to cross sections that are calculated to NLO precision [96, 97]. Samples of W and Z boson production in association with jets, with leptonic final states, are generated at LO using MG5_aMC@NLO. Separate samples are generated with exactly one, two, three, and four additional jets to ensure a large set of background events. These samples are normalized to their NNLO cross sections [98]. Multijet QCD events are generated using PYTHIA for both the matrix-element calculations and parton

shower/hadronisation. Only multijet QCD events with large electromagnetic activity or that contain a muon are generated in order to maximise the available sample size.

The simulated samples need to contain enough events to ensure they are not affected by statistical fluctuations when comparing to the data. The luminosity, described in further detail in Sec. 4.2, produced for each simulation sample is shown in the final column of Tab. 3 and should ideally be comparable to or larger than the luminosity of the data collected. Looking ahead to Tab. 5, the total collected data luminosity is 35.9 fb^{-1} and so most of the signal and background samples do indeed contain enough events. The multijet QCD events are not particularly well modelled and can suffer from a lack of statistics and so a data-driven multijet QCD estimate is employed, which is described in Sec. 6.6. The multijet QCD samples are still required in the creation of the estimate however.

Samples are also generated for variations of some of the generator parameters in POWHEG+PYTHIA. These are shown in Tab. 4 and are used as part of the estimation of the theoretical uncertainties due to the modelling. Other theoretical uncertainties in the modelling are calculated as event weights and stored in the central POWHEG+PYTHIA sample. The CMS detector response for all simulated samples is modelled using GEANT4.

Table 3: The set of simulated samples used in this thesis. They include four different signal $t\bar{t}$ production samples and samples for the background production. The backgrounds are single top quark production, vector boson production in association with jets and multijet QCD.

	Matrix element	**Parton shower**	**Tune**	**Cross section** (pb)	**Events** ($\times 10^6$)	**Luminosity** (fb^{-1})
Models of $t\bar{t}$ production						
POWHEG+PYTHIA	POWHEG	PYTHIA	CUETP8M2T4	831.76	154.9	186.3
POWHEG+HERWIG++	POWHEG	HERWIG++	EE5C	831.76	59.2	77.1
MG5_aMC@NLO-LO	MG5_aMC@NLO	PYTHIA	CUETP8M1	831.76	10.1	12.2
MG5_aMC@NLO-NLO	MG5_aMC@NLO	PYTHIA	CUETP8M1	831.76	30.4	36.6
Models of single top background production						
Single top t-channel	POWHEG	PYTHIA	CUETP8M1	136.02	66.9	492.0
Single anti-top t-channel	POWHEG	PYTHIA	CUETP8M1	80.95	38.8	479.4
Single top tW-channel	POWHEG (v1)	PYTHIA	CUETP8M1	35.6	7.9	223.2
Single anti-top tW-channel	POWHEG (v1)	PYTHIA	CUETP8M1	35.6	7.9	222.8
Single top/anti-top s-channel	MG5_aMC@NLO	PYTHIA	CUETP8M1	6.35	0.6	98.1
Models of vector boson background production						
Drell-Yan + 1 jet	MG5_aMC@NLO	PYTHIA	CUETP8M1	1016	62.6	61.6
Drell-Yan + 2 jets	MG5_aMC@NLO	PYTHIA	CUETP8M1	331.4	20.0	60.3
Drell-Yan + 3 jets	MG5_aMC@NLO	PYTHIA	CUETP8M1	96.36	5.9	60.8
Drell-Yan + 4 jets	MG5_aMC@NLO	PYTHIA	CUETP8M1	51.4	4.2	81.7
W boson + 1 jet	MG5_aMC@NLO	PYTHIA	CUETP8M1	9493	45.4	4.8
W boson + 2 jets	MG5_aMC@NLO	PYTHIA	CUETP8M1	3120	60.2	19.3
W boson + 3 jets	MG5_aMC@NLO	PYTHIA	CUETP8M1	942.3	59.1	62.7
W boson + 4 jets	MG5_aMC@NLO	PYTHIA	CUETP8M1	524.2	30.0	57.2
Models of multijet QCD background production						
Muon enriched QCD (20-30)	PYTHIA	PYTHIA	CUETP8M1	558528000 * 0.0053	30.6	0.01
Muon enriched QCD (30-50)	PYTHIA	PYTHIA	CUETP8M1	139803000 * 0.01182	30.0	0.02
Muon enriched QCD (50-80)	PYTHIA	PYTHIA	CUETP8M1	19222500 * 0.02276	19.8	0.04
Muon enriched QCD (80-120)	PYTHIA	PYTHIA	CUETP8M1	2758420 * 0.03844	23.6	0.2
Muon enriched QCD (120-170)	PYTHIA	PYTHIA	CUETP8M1	469797 * 0.05362	8.0	0.3
Muon enriched QCD (170-300)	PYTHIA	PYTHIA	CUETP8M1	117989 * 0.07335	17.4	2.0
Muon enriched QCD (300-470)	PYTHIA	PYTHIA	CUETP8M1	7820.25 * 0.10196	49.0	61.4
Muon enriched QCD (470-600)	PYTHIA	PYTHIA	CUETP8M1	645.528 * 0.12242	19.0	240.1
Muon enriched QCD (600-800)	PYTHIA	PYTHIA	CUETP8M1	187.109 * 0.13412	10.0	397.7
Muon enriched QCD (800-1000)	PYTHIA	PYTHIA	CUETP8M1	32.3486 * 0.14552	19.8	4199.3
Muon enriched QCD (1000-Inf)	PYTHIA	PYTHIA	CUETP8M1	10.4305 * 0.15544	13.4	8264.9
Electron enriched QCD (20-30)	PYTHIA	PYTHIA	CUETP8M1	557600000 * 0.0096	9.2	0.002
Electron enriched QCD (30-50)	PYTHIA	PYTHIA	CUETP8M1	136000000 * 0.073	6.8	0.0007
Electron enriched QCD (50-80)	PYTHIA	PYTHIA	CUETP8M1	19800000 * 0.146	45.2	0.02
Electron enriched QCD (80-120)	PYTHIA	PYTHIA	CUETP8M1	2800000 * 0.125	76.5	0.2
Electron enriched QCD (120-170)	PYTHIA	PYTHIA	CUETP8M1	477000 * 0.132	77.8	1.2
Electron enriched QCD (170-300)	PYTHIA	PYTHIA	CUETP8M1	114000 * 0.165	11.5	0.6
Electron enriched QCD (300-Inf)	PYTHIA	PYTHIA	CUETP8M1	9000 * 0.15	7.4	54.6
bc to E QCD (30-80)	PYTHIA	PYTHIA	CUETP8M1	159068000 * 0.00255	15.3	0.04
bc to E QCD (80-170)	PYTHIA	PYTHIA	CUETP8M1	3221000 * 0.01183	14.9	0.4
bc to E QCD (170-250)	PYTHIA	PYTHIA	CUETP8M1	105771 * 0.02492	9.7	3.9
bc to E QCD (250-Inf)	PYTHIA	PYTHIA	CUETP8M1	21094.1 * 0.03375	9.8	13.7

Table 4: The set of samples associated with the uncertainties in the POWHEG+PYTHIA modelling. Other uncertainties are modelled by the use of event weights in the central POWHEG+PYTHIA simulation.

Variation	**Events** ($\times 10^6$)	**Luminosity** (fb^{-1})
ISR down	59.0	70.9
ISR up	59.0	71.0
FSR down	59.3	71.3
FSR up	59.2	71.2
Underlying event down	58.3	70.1
Underlying event up	59.0	70.9
Matching scale (h_{damp}) down	58.2	69.9
Matching scale (h_{damp}) up	58.9	70.8
CR MPI-based erdON	59.9	72.0
CR QCD-inspired	59.6	71.7
CR Gluon move	59.0	71.0
Top quark mass (169.5 GeV)	59.5	70.4
Top quark mass (175.5 GeV)	59.4	71.4

CHAPTER 4

The Large Hadron Collider and the Compact Muon Solenoid

The LHC is primarily a proton-proton circular collider, situated at the Conseil Européen pour la Recherche Nucléaire (CERN) on the French-Swiss border. It is built 100 m underground in the tunnels that housed the Large Electron-Positron (LEP) collider and is currently the largest machine in the world, with an accelerator ring 27 km in circumference. There are four main experiments operating at the LHC; two multipurpose general detectors, CMS and ATLAS; a precision b physics experiment (LHCb); and finally a detector dedicated to heavy ion physics, A Large Ion Collider Experiment (ALICE). The full accelerator complex at CERN is shown in Fig. 15. This chapter will highlight a brief history of the LHC, how it is operated to produce collisions in the experimental caverns and the current data-taking environment. A description is given of the subdetectors that form the CMS experiment.

4.1 The Large Hadron Collider

4.1.1 A brief history of the LHC and CMS

All dates from the LHC timeline are taken from [100]. The first public announcement for the construction of the LHC was given on the 16th December 1994. A couple of years earlier, on the 1st October 1992, the CMS experiment submitted its letter of intent to the LHC Experiments Committee and on the 31st January 1997, the CMS experiment was approved. Construction of the CMS site at Cessy began in July 1998, with the CMS cavern completed by the 1st February 2005. In the following three years, the detector was assembled in segments above ground and lowered down into the cavern. The final large detector segment was lowered on the 23rd July 2008. The final dipole magnet of the LHC was lowered into place on 26th April 2007, heralding the completion of the full accelerator ring. The 10th September 2008, was a very significant day in the history of the LHC marking the first circulation of proton beams around the collider. Shortly after, on the 19th, a fault occurred in an electrical connection between a dipole and quadrupole magnet, leading to a release of helium in the tunnels, causing significant damage to the LHC. In total 37 damaged magnets were replaced and 16 refurbished, with the final one

Figure 15: The full CERN complex showing schematically the location of all the experiments on the ring complex, including the location of the four main experiments at the LHC: CMS, ATLAS, LHCb and ALICE [99].

being lowered on the 30th April 2009. Beams were present again in the machine on the 20th November 2009, over a year after the malfunction.

Run I physics data collection at a centre-of-mass energy $\sqrt{s} = 7$ TeV, was started on the 30th March 2010 and ran through until the 18th October 2011. A second period of data collection at $\sqrt{s} = 8$ TeV was performed from the 5th April 2012 to the 6th February 2013. During this time, enough data was collected to confirm the existence of a SM-like H boson, officially announced on the 4th July 2012 [101]. With one of the primary mandates of the LHC fulfilled, attention has turned to discovering the existence of physics beyond the SM and measuring precisely the properties of the H boson and other SM particles.

On the 3rd June 2015, after a two year long shutdown for upgrades, the collider started up again for Run II at an unprecedented $\sqrt{s} = 13$ TeV. Four sets of data are scheduled to be taken during Run II, with a small set taken at the end of 2015 and major runs during 2016, 2017 and 2018. Each major run lasts from a commissioning period around March to the end-of-year technical stop in December. Run II data is scheduled to be collected until December 2018, from when it will enter a second long shutdown period before Run III scheduled to operate between 2020-2022. This thesis uses the data set collected during 2016.

4.1.2 Operating the LHC

The source for the proton beams is a small canister of hydrogen gas. The hydrogen gas is then ionised to create protons by applying a large electric field. The collection of protons is then accelerated, within ultra-high vacuums, through the Linear Accelerator 2 (LINAC2), which uses alternatingly charged conductors to accelerate the bunch of protons up to an energy of 50 MeV, before being injected into the Proton Synchrotron Booster (PSB). The PSB is an accelerator ring that boosts the protons to an energy of 1.4 GeV and passes them to the Proton Synchrotron (PS) which provides a further boost up to 25 GeV. From the PS the proton bunches are accelerated in the Super Proton Synchrotron (SPS) to an energy of 450 GeV. The SPS splits the single beam into two counter-rotating ones and injects them into the LHC. In the LHC, each proton beam can be accelerated up to a maximum energy of 7 TeV per beam. During Run I, this beam energy was operated at 3.5(4) TeV and has been increased to 6.5 TeV for Run II.

To reach these operational collision energies the proton bunches are accelerated, confined and squeezed by magnetic fields of up to 8.3 T from over 9300 superconducting magnets. The most common type of magnet used in the LHC is the dipole magnet, of which there are 1232. Each dipole magnet is 15 m long, weighs 35 tonnes and has a magnetic field strength of 8.3 T which is used to bend the beam around the collider. Figure 16 shows a cross section through one of the LHC dipole magnets. Each dipole contains two beam pipes for the counter-circulating beams and along each of these beam pipes are coiled niobium-titanium alloy wires which form the superconducting dipole magnet. Superconductivity in the magnet is achieved by a closed circuit of liquid helium, which cools the iron yoke to 1.9 K.

In addition to the dipole magnets, the LHC uses 858 quadrupole magnets. Of these, 392 are lattice quadrupoles used to stop the beam from defocussing over time due to the similarly charged beam constituents. Quadrupole magnets are also used to confine the proton beams as they are entering the experimental caverns. A set of three, known as an inner triplet, squeezes each beam to a diameter of 16 μm to ensure a maximal number of collisions at the interaction point. There are additional multipole magnets used to correct for other, smaller, effects present in the beam, such as the gravitational force on the protons.

Figure 16: A cross section of a LHC dipole magnet [102].

4.2 Luminosity measurements

The instantaneous luminosity is defined as:

$$\mathcal{L} = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{4.1}$$

$\mathcal{L}$ depends on the bunch crossing rate, f, the number of protons in each colliding bunch, n_1 and n_2, and the root-mean-square of the transverse beam sizes in the horizontal (σ_x) and vertical (σ_y) directions. The beams are assumed to have a Gaussian profile, of width $\sigma = 16\ \mu$m, and to be colliding head-on. For nominal running during Run II, each beam contains 3564 bunch spaces of which typically 2808 are filled during data taking, leading to a minimum bunch spacing of 25 ns and a collision rate $f = 40$ MHz. Each bunch contains $\approx 1.1 \times 10^{11}$ protons, producing a peak $\mathcal{L} = 1.50 \times 10^{34}$ cm^{-2} s^{-1}. $\mathcal{L}$ varies with respect to time due to the depletion of the protons in each bunch after every collision. The decrease in luminosity is mitigated to a small extent by the reduction of the beam widths σ_x and σ_y, however each beam has an approximate physics taking lifetime of ten hours before they need to be refilled. The total number of inelastic collisions (known as minimum bias events) produced at the LHC, over a time t, can be estimated from the luminosity by:

$$\mathrm{N} = \sigma_{\mathrm{minbias}} \times \mathcal{L}_{\mathrm{int}}, \tag{4.2}$$

where $\sigma_{\mathrm{minbias}}$ is the minimum bias cross section. $\mathcal{L}_{\mathrm{int}}$ is the integrated instantaneous luminosity over time t

$$\mathcal{L}_{\mathrm{int}} = \int_{t_0}^{t} \mathcal{L}(t)\mathrm{d}t. \tag{4.3}$$

The $\mathcal{L}_{\mathrm{int}}$ distribution during the 2016 data taking period is shown in the left panel of Fig. 17, where a total of 35.9 fb^{-1} of certified data was collected by the CMS experiment. The right panel shows the delivered luminosity for all other years of data taking. Using the current $\sigma_{\mathrm{minbias}}$

Figure 17: The left panel shows the delivered and recorded integrated luminosity during data taking in 2016 and the right panel the integrated luminosity delivered by the LHC for all years of operation. Figures taken from [103].

measurement of 69.2 mb, this means that approximately 2.5×10^{15} collisions were processed in this period. More than one proton-proton interaction typically occurs in each bunch crossing and these additional interactions are referred to as *in-time pileup*. As the collision rate at the LHC is so high, it is also possible for remnants of previous bunch crossings to affect the current one, particularly in calorimeters where there is a latency higher than a typically bunch spacing. These remnants are called *out-of-time pileup*. The average number of interactions per bunch crossing during 2016 data taking was approximately 30, as can be seen in Fig. 18, using a minimum bias cross section of 80 mb. Particles from pileup are included in the particle reconstruction of the event, leading to misidentification. Reconstruction and pileup mitigation techniques are described in Secs. 5.1.5 and 6.2.

Figure 18: The distribution of the number of interactions in a bunch crossing. The minimum bias cross section is taken to be 80 mb. Figure taken from [103]

4.3 The Compact Muon Solenoid

CMS is a general purpose detector, designed to be able to perform precision measurements of the standard model and searches for new physics [104]. It is a cylindrical, hermetic detector composed of a barrel region and two endcaps, each formed of layers of subdetectors around the beam collision point, as shown in Fig. 19. These subdetectors, in order from the collision point, are the pixel and strip tracker, electromagnetic calorimeter (ECAL), hadronic calorimeter (HCAL) and muon chambers. The superconducting magnet is situated in between the hadronic calorimeter and muon chambers. To operate in a high luminosity environment the CMS experiment must have excellent resolution in each of its subdetectors and a low latency to manage the high collision rate. The materials used must be reliable while having maximal longevity when operating in a high-radiation environment.

CMS uses a right-handed coordinate system, where the x-axis points towards the centre of the accelerator ring, the y-axis points vertically upward and the positive z-axis lies parallel to the anti-clockwise beam axis. The azimuthal angle ϕ, is measured from the x-axis in the transverse $x - y$ plane. The polar angle θ, is measured from the positive z-axis in the $z - y$ plane. At the CMS experiment, collisions are viewed in the centre-of-mass frame of the colliding protons, which means that the colliding partons usually have a Lorentz boost along the beam direction. For this reason, the angles of the particles are normally expressed in terms of the rapidity (y) or pseudorapidity (η), in which the production is roughly constant. The rapidity is defined as

$$y \equiv \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{4.4}$$

where E is the energy of a measured particle and p_z is the z-component of the momentum. Rapidity is advantageous to use as rapidity differences between two particles is invariant under a Lorentz boost along the z-axis. At the LHC, most particles are highly relativistic such that the rest mass is small compared to the energy ($p_z \approx E \cos\theta$). This can be substituted into Eq. 4.4 and using trigonometric relations, the pseudorapidity is defined as

$$\eta \equiv - \ln \left(\tan \left(\frac{\theta}{2} \right) \right). \tag{4.5}$$

Jet production in pp collisions is roughly constant in η, reflecting the forward nature of the production. The subdetectors at the CMS experiment extend to at least a pseudorapidity magnitude of 2.4, covering more than 95% of the phase space. Individual subdetectors can extend to a higher $|\eta|$, for example the forward detectors of the HCAL, to study very boosted jets or the underlying event. The acceptance within $|\eta| < 2.4$ can be hampered by issues such as the gap between barrel and endcap detectors.

To completely describe the kinematic properties of a particle, its energy and momentum must be known. While it is not possible to detect the complete momenta spectrum of an event, due to many particles produced in the collision being lost down the beam pipe and the unknown boost of the partons, it is possible to measure the momenta in the transverse plane, where the sum of the momenta is conserved. To measure the transverse momenta and energy of particles, CMS uses a combination of hits in the tracking systems from charged particles and energy deposits in the calorimeters. These subsystems are described in more detail in the following sections.

4.3.1 CMS pixel and strip trackers

Figure 20 shows the layout of the silicon tracking detectors in CMS. Closest to the interaction point are the pixel detectors consisting of three barrel layers situated at radial distance $r =$

Figure 19: A schematic of CMS, sliced to show the internal layout. It consists of layers around the central interaction point in the order of silicon pixel and strip detectors, electromagnetic and hadronic calorimeters, superconducting magnet and finally muon chambers [105].

4.4, 7.3 and 10.2 cm in the transverse plane from the interaction point with two endcap disks lying at $z = \pm 34.5$ and ± 46.5 cm. There are 65 million pixels in the inner tracker, each $100 \times 150\ \mu\text{m}^2$, giving a total active area of 1.06 m^2. Surrounding the pixel detectors are the silicon strip tracker sensors. These are arranged into four subsystems: the Tracker Inner Barrel and Disks (TIB and TID), the Tracker Outer Barrel (TOB) and the Tracker Endcaps (TEC). The TIB consists of four barrel layers between $20 < r < 55$ cm with each end capped by 3 disks ($|z| < 118$ cm) forming the TID. Wrapping around the TIB and TID, is the TOB consisting of six barrel layers between $55 < r < 116$ cm. Finally, the TEC consist of nine disks between $124 < |z| < 282$ cm, capping the TIB, TID and TOB. There are 15 148 silicon strip modules in total that make up the outer tracker, giving a total active surface area of 198 m^2. Some strip layers in the tracker, coloured blue in Fig. 20, have back-to-back modules, rotated with respect to each other by a small stereo angle ($\approx 5°$) in order to improve the 3-dimensional point resolution by providing a measurement of the z co-ordinate in the barrel and r in the disks.

Figure 20: A slice through the CMS tracker detector in the $r-z$ plane. The separate subsystem regions have been highlighted and are classified as the Tracker Inner Barrel and Disks (TIB and TID), the Tracker Outer Barrel (TOB) and the Tracker Endcaps (TEC). The Pixel subsystem is highlighted in red. Back-to-back stereo strip modules are highlighted in blue. Figure taken from [106].

Each tracking sensor operates in a similar manner. For the pixel sensors, an n^+-n junction is utilised where the silicon has been doped with phosphorous to different degrees. For the strip sensors, a p^+-n junction is used where a silicon strip doped with boron has been laid on the bulk n-type silicon. Doping with phosphorous causes the silicon to become an electron donor (n-type) and doping with boron to become an electron acceptor (p-type). The n(p)^+-type silicon is where silicon has been doped to such an extent that the resistivity is very low.

In the case of the silicon strip sensor, electrons diffuse from the bulk n-type silicon to a strip of p^+-type silicon, creating a small opposing electric field from the formation of ions either side of the p^+-n junction. The field stops further electron-hole movement and forms a potential step. The region with no excess electrons or holes is known as the depletion zone, which causes the junction to act as a diode. A reverse bias voltage is applied to ensure the silicon is fully depleted by attracting the free electrons and holes away from the p^+-n junction. When a charged particle ionises the n-type bulk silicon, the electrons and holes produced ($\approx 30\,000$) drift in the large electric field such that the holes move towards the p^+-type silicon producing

a small signal current which is then amplified and shaped. This method can be applied to any semiconductor junction where each side has different doping concentrations.

The pixel detector receives a charged particle flux of 1 MHz mm^{-2}. At this high rate of particles the pixels must have a high granularity to keep the fraction of channels hit (occupancy) reasonable. A low occupancy in the pixel detector is necessary to reconstruct individual particle tracks back to an interaction vertex, in a high-track environment. Lower occupancies present in the TOB and TEC mean that a longer strip can be used, however this incurs an increase in noise, which is alleviated by increasing the thickness of the silicon sensor from 320 μm to 500 μm. Tracks reconstructed in the pixel and strip detectors can then be used in vertex reconstruction, particle identification and charged particle momenta measurements. The track reconstruction is detailed in Sec. 5.1.1.

4.3.2 Electromagnetic calorimeter

The CMS ECAL, as shown in Fig. 21, is made of three subsystems: the barrel, the endcaps and the preshower detector. Together, they form a compact coverage around the interaction point, up to $|\eta| < 3$. A total of 75 848 lead tungstate ($PbWO_4$) scintillating crystals are used in

Figure 21: A geometrical quarter view of the CMS ECAL, highlighting the barrel, endcap and preshower detector regions. Figure taken from [107].

the ECAL barrel and endcaps. In the barrel section, the crystals are laid out in an $\eta - \phi$ grid extending to $|\eta| < 1.479$. The inner surface of the barrel is at $r = 1.29$ m, with each crystal having a front surface area of 2.2×2.2 cm^2 and length 23 cm. In each endcap, the crystals are instead laid out in a $x - y$ grid, with the front surfaces at $z = \pm 3.14$ m, extending through $1.479 < |\eta| < 3.0$. Each endcap crystal has a front surface area of 2.9×2.9 cm^2 and a length of 22 cm. All the crystals are slightly off-centre with respect to the primary interaction point to avoid particles being lost down the small gaps between adjacent crystals.

$PbWO_4$ is an ideal choice for the CMS ECAL due to its short radiation and Molière lengths of 0.89 and 2.2 cm respectively. The radiation length gives a measure of the penetration of the EM shower into the crystal and the Molière length how confined the EM shower is, such that a cascade will mostly be contained within one crystal. In addition, $PbWO_4$ crystals are very fast with 80% of the blue-green scintillation light being emitted in the time it takes for

another collision to occur. The light emission from the $PbWO_4$ crystals is relatively small at 30γ per MeV of energy deposited and therefore needs to be amplified using silicon avalanche photodiodes (barrel) and vacuum phototriodes (endcaps). $PbWO_4$ is also radiation resistant tolerating up to 6.2×10^{17} MeV kg^{-1} of radiation energy.

The purpose of the ECAL preshower detector is to distinguish between a high energy single photon and a π^0 decaying into two closely spaced photons. To do this the preshower consists of two layers of lead, which initiate an electromagnetic shower for photons (and electrons), each closely followed by a layer of silicon strip detectors. The silicon strips are placed orthogonal to each other to measure shower positions precisely and are fine enough in granularity to differentiate between a single photon shower and multiple close-spaced photons. A π^0 decaying in the barrel typically has low enough energy not to need the additional resolution provided by the preshower detector. The preshower degrades the resolution of the ECAL endcap due to the dense lead plates, however as the energy deposited within the lead is proportional to that deposited in the silicon, a correction can be applied mitigating the effect.

These attributes mean that the ECAL can be both compact enough to fit within the magnet and granular enough to perform with high energy resolution. The resolution performance is measured by fitting a Gaussian function to the reconstructed energy distributions at different test beam energies and is parametrised by

$$\left(\frac{\sigma_{\mathrm{ECAL}}}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{4.6}$$

where S is a stochastic term (number of photoelectrons), N is a noise term (electronics and digitisation) and C is a constant term (leakage from the back of the ECAL). The left panel of Fig. 22 shows the energy resolution with respect to the energy deposited in a 3×3 crystal lattice around incident electrons at a test beam measurement before the beginning of Run I with the values for S, N and C shown. The right panel shows the energy resolution with respect to $|\eta|$ from the most recent public measurement [108]. It is extracted from an unbinned likelihood fit on $Z \to e^+e^-$ events collected in 2015 for electrons with low brehmstrahlung radiation and $E_{\mathrm{T}} \approx 45$ GeV. The resolution was measured to be $< 2\%$ for $|\eta| < 1.0$ and between $2 - 5\%$ elsewhere.

4.3.3 Hadronic calorimeter

The majority of the hadronic calorimeter, shown in Fig. 23, sits within the superconducting solenoid. It contains sampling calorimeters in the barrel (HB) and in the two endcaps (HE). The HB operates to an $|\eta| < 1.4$ and the HE covers the range to $1.4 < |\eta| < 3.0$. The HB is radially restricted between the ECAL and the superconducting magnet at 1.77 m $< r < 2.95$ m and as such there is only a limited amount of absorber present necessitating the requirement for scintillator plates situated outside the solenoid (HO). The HB is divided into two half-barrels each composed of 18 identical azimuthal wedges consisting of alternating layers of non-magnetic, short radiation-length brass absorber and plastic scintillator tiles. Each wedge is capped by steel absorber for structural strength. The HE are formed of a similar structure. Around 70 000 scintillating tiles of dimensions $\Delta\eta \times \Delta\Phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and 0.17×0.17 for $|\eta| > 1.6$ are used in the construction of the HCAL.

The forward hadronic detectors (HF) are situated ± 11.2 m from the primary interaction point within a pseudorapidity range of $3.0 < |\eta| < 5.2$. The HF, which is very close to the beam line, receives a large flux of particles and therefore must be very radiation resistant. To this end it is made of steel and quartz fibres, where the emitted Cherenkov light in the quartz fibres is passed to photomultipliers and read as signal [109]. The HF allows for the reconstruction of forward jets and also allows for a better description of $p_{\mathrm{T}}^{\mathrm{miss}}$.

Figure 22: The upper panel shows the energy resolution of the CMS ECAL measured in a beam test. The energy was measured in a 3×3 crystal lattice with the electron striking the central crystal. The lower panel shows the resolution extracted from an unbinned likelihood fit on $Z \to e^+e^-$ events. Figures taken from [104, 108].

Figure 23: A geometrical quarter view of CMS HCAL, highlighting the Barrel (HB), Endcap (HE), Forward (HF) and Outer (HO) subdetectors. Figure taken from [104].

As hadrons pass through the ECAL they leave only minimal deposits before entering the HCAL. When they continue through the dense absorption layers they interact and produce cascades of particles. These particle showers cause the plastic scintillator to emit blue-violet light which is passed by very small wavelength-shifting fibres as green light to the readout where the signal is amplified using hybrid photodiodes. At small $|\eta|$ the stopping power of the HB is not enough to fully contain the hadronic cascade so the HO is used to catch the tails of the hadronic cascades improving the $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution of the detector. The resolution of the HCAL is parametrised as

$$\frac{\sigma_{\mathrm{HCAL}}}{E} = \frac{S}{\sqrt{E}} + C, \tag{4.7}$$

where S is measured to be 0.85 and C to be 0.07 in [110]. This is similar to the expected resolution of hadronic cascades shown in Fig. 24 which are given in terms of reconstructed jet transverse energy. Jet clustering will be discussed in Sec. 5.1.5.

4.3.4 Superconducting magnet

The solenoid at the heart of CMS is the largest superconducting magnet that has ever been built at 13 m long and 6 m in diameter. It is composed from coils of a niobium-titanium alloy wire that are cooled to an operating temperature of 4.5 K. It operates at a field strength 3.8 T reduced from a design operation field strength of 4 T in order to maximise longevity. The magnet and steel return yokes, which are used to contain the magnetic field, are the most massive components in CMS weighing approximately 12,000 tonnes. The high field strength allows for very precise measurements of the curvature of charged tracks leading to precise measurements of the momenta of charged particles.

Figure 24: The transverse energy resolution of reconstructed jets in the CMS HCAL. It is split into the barrel jets ($|\eta| < 1.4$), endcap jets ($1.4 < |\eta| < 3.0$) and forward jets ($3.0 < |\eta| < 5.0$) Figure taken from [104].

4.3.5 Muon detectors

The muon chambers are the outermost set of detectors at CMS, interleaved with the steel return yoke of the superconducting solenoid. There are three types of muon chamber present at CMS: resistive plate chambers (RPC), cathode strip chambers (CSC) and drift tubes (DT). The layout of the muon chambers is presented in Fig. 25.

The relatively low muon rates and uniform local magnetic field strengths in the barrel make DT stations a good option for measuring muons in this phase space region covering up to $|\eta| < 1.2$. Each cell is 13×42 mm^2 containing a gaseous mixture composed of 15% CO_2 and 85% Ar, with an anode central wire running the length of the cell. As a charge is left in the cell, liberated electrons drift towards the central wire forming a signal. The cells arranged into a set of four layers, each offset by half a cell-width to reduce non-sensitive detector regions, in what is know as a superlayer. A set of three of the superlayers form a DT muon station, where the outer two are aligned parallel to the beam line and the central one perpendicular. This means the central superlayer can be used to calculate the z co-ordinate and the outer two to measure the co-ordinates in r, ϕ space. The DT cells have a high spatial resolution but also a slow drift time of up to 380 ns.

In the endcaps where the muon rates are higher, 540 CSCs are employed in the range $0.9 < |\eta| < 2.4$. They have trapezoidal shapes to provide contiguous ϕ coverage in four disks in each endcap. The CSCs have a quicker response time, a finer granularity and are less affected by variations in the magnetic field making them preferable to the DTs. Each CSC station consists of seven layers of cathode panels, each with strips milled running lengthwise at constant $\Delta\phi$ width, interleaved with six layers of azimuthally running anode wires. The formation of six gas gaps are filled with 40% Ar, 50% CO_2 and 10% CF_4. As a charged muon passes through the gas, electrons travel to the anode wires and ions to the cathode strips providing the co-ordinates in r, ϕ space.

Finally, 610 RPCs are found in both the barrel (six layers) and endcaps (four layers) up to $|\eta| < 1.8$. They are composed of gaseous parallel-plate detectors with two gas gaps filled with a mixture of 96.2% $C_2H_2F_4$, 3.5% iso-C_4H_{10} and 0.3% SF_4, separated by a common charge

Figure 25: A geometrical one-quarter view of the CMS muon subsystems. The drift tube stations are indicated in yellow the cathode strip chambers in green and the resistive plate chambers in blue. Figure taken from [111].

read-out strips. The RPCs have adequate spatial resolution, however they are complimentary to the DTs and CSCs due to their very fast response time (≈ 2 ns) which is much faster than the time it takes for the bunch crossing. This makes the RPCs ideal detectors for triggering muons, able to unambiguously identify the relevant bunch crossing a muon track is associated to. Each component is discussed in more detail in [104, 111].

Figure 26 shows the expected momentum resolution of the muons based on tracker information only, muon chamber information only and a combination of tracker and muon chamber information. At low muon p_{T} the resolution is driven from the reconstruction of muons in the tracker, however as the p_{T} of the muon increases then the combination of information from the tracker and muon chambers results in improved resolution.

4.3.6 Triggers

The CMS detector produces events at an approximate rate of 40 MHz. If each reconstructed event has a size of ~ 1 MB, then data would be produced at a rate of 40 TB s^{-1}. It is not feasible to fully reconstruct and store the full rate of data production so a method of data reduction, called triggering, is used. Triggering decides on an event-by-event basis whether a collision is likely to be highly-energetic and interesting, or just an elastic collision or low-energy QCD process. There are two types of triggering systems at CMS, the Level-1 hardware triggers (L1T) and the High-Level software triggers (HLT).

The L1T must decide within 3.2 μs of an event occurring whether it should be rejected of tentatively accepted. The schematic workflow of the L1T is shown in Fig. 27. It uses only basic information from the detector known as *trigger primitives* to make a decision. These trigger primitives are formed from the calorimeter deposits and the hits in the muon detectors.

Figure 26: The muon p_{T} resolution as a function of p_{T} using the tracking only, muon subdetectors only or both. Figure taken from [104].

There are two trigger subsystems to the L1T, the calorimeter trigger and the muon trigger. These access the trigger primitives from different subdetectors and operate independently, with the exception of information passed from the calorimeter trigger to the muon trigger for the computation of muon isolation. The outcome of both trigger components is then passed to the L1 *global trigger* which makes the final decision on the event. The L1T reduces the event rate from 40 MHz to 100 kHz, which provides enough latency for the events to be fully reconstructed before being processed by the HLT.

The L1T runs approximately 200 trigger algorithms, the outcomes of which form the *seeds* for the HLT. The HLT with access to the L1 seeds, full collision kinematics and composition, reduces the final physics production rate to 1 kHz and classifies the events with respect to specific topologies. The events are collected into data sets containing similar topologies, for example sets with a single isolated electron or sets that have high jet multiplicities.

4.3.7 Computational power

The amount of raw data produced by the LHC and the number of large simulations generated for analyses is far too much for a single storage site to be viable. Therefore the Worldwide LHC Computing Grid (WLCG) was produced to share the burden. The WLCG is a tiered system with all raw data being stored at a central Tier-0 computing facility and subsequently distributed to Tier-1 and Tier-2 sites in the countries and institutes affiliated to CERN.

The complete set of raw data is stored at the CERN Tier-0 data centre, where the combined permanent archive stored on tape breached 200 Pb on the 29th June 2017. In terms of online storage the data centre at CERN holds 45 Pb with a further 5.5 Pb extension at the Wigner Research Centre for Physics in Budapest. The 100,000 processor cores at the data centre deal with 1 Pb of data every day.

The large Tier-1 facilities store a large fraction of the data and simulation sets so that there are multiple redundancies for every data set. The 13 Tier-1 facilities are the main workhorses

Figure 27: A schematic representing the process flow of the L1T. On the left is the calorimeter trigger and on the right the muon trigger which operate on independent trigger primitives. The final decision is taken by the global trigger using all information available [112].

of the WLCG performing large-scale reprocessing of the raw data sets and the generation of simulated events.

Tier-2 sites, of which there are 155, store data sets distributed to them from Tier-1 facilities and specific data sets that are useful for locally based analysis teams. They provide a moderate number of processor cores for computational tasks, most generally for the production of simulated events. The distributed nature of the file system allows for any person to access data sets from any site which has it stored and to use free computational power anywhere that is part of the WLCG, increasing the overall computational efficiency.

4.3.8 Data collection in 2016

As already stated in Sec. 4.2, an integrated luminosity of 35.9 fb^{-1} was collected by CMS in 2016. To further classify the data used in this thesis, Tab. 5 shows the data collected with respect to the run period. A run period refers to periods of similar data taking throughout the year. Run period A is used for the commissioning of the detector. Run periods often change after a short technical stop has occurred. The initial data sets used in this thesis are classified as containing either a single electron or a single muon. On these data sets a preselection is made via the HLT selection. In the e+jets channel the trigger requires there to be an electron with a $p_{\mathrm{T}} > 32$ GeV, a $|\eta| < 2.1$ and to pass a set of tight identification requirements (HLT_Ele32_eta2p1_WPTight_Gsf). In the μ+jets channel an isolated global muon or isolated tracker muon is required with a $p_{\mathrm{T}} > 24$ GeV and a $|\eta| < 2.4$ (HLT_IsoMu24 $\cup$ HLT_IsoTkMu24).

Table 5: The sets of single electron and single muon data collected during 2016 split into individual runs. A total luminosity of 35.9 fb^{-1} of data was collected.

Channel	**Data set**	**Run ranges**	**Luminosity** (fb^{-1})
SingleElectron(Muon)	Run2016B-03Feb2017_ver2-v2	272007–275376	5.79
	Run2016C-03Feb2017-v1	275657–276283	2.57
	Run2016D-03Feb2017-v1	276315–276811	4.25
	Run2016E-03Feb2017-v1	276831–277420	4.01
	Run2016F-03Feb2017-v1	277772–278808	3.10
	Run2016G-03Feb2017-v1	278820–280385	7.54
	Run2016G-03Feb2017-ver2(3)-v1	280919–284044	8.61
Total integrated luminosity			35.87

CHAPTER 5

Reconstructing the physics objects

Whether the components of an event are measured from a detector or simulated from a model, the deposits need to be reconstructed into physics objects. In CMS, the reconstruction of objects from the calorimeter energy deposits and charge deposits in the tracker and muon chambers is done by an algorithm called *particle flow* (PF) [113]. These collections of deposits are reconstructed into likely electrons, photons, muons, charged and neutral hadronic jets and missing p_{T} and are known as *PF candidates*. The PF candidates undergo further, stricter identity cuts depending on the needs of the analysis, into high-purity collections of object candidates.

5.1 The particle flow algorithm

The PF algorithm is used to reconstruct objects to particle candidates, from the raw deposits in the event. The first step is to reconstruct the *PF elements* to be used in the algorithm. These elements are the charged particle tracks and vertices along with the calorimeter clusters and muon system hits. Each final state particle can, in general, be matched to multiple PF elements. The combination of the correlated measurements, allows for the reconstruction and identification of particles.

5.1.1 Particle flow elements

Tracks are built in an iterative process. Firstly, a track seed is generated for a triplet of hits in the pixel detector, with a few hits compatible with a charged-particle trajectory. Next, additional hits are added, layer-by-layer, to the current track candidate based on the track seed trajectory, extrapolated via a Kalman filter [114]. New track candidates are formed in this process when multiple hits match the extrapolation and track candidates with missing hits are also kept. Once the outer tracker layer has been reached, possible duplicates are avoided by requiring each pair of track candidates to be composed of at least 50% of unique hits. If not, then the candidate with the most missing hits is dropped and if they are the same length then the one with the highest χ^2 is dropped. The final track fitting is then performed, using all associated track hits to determine the charged-particle properties (vertex, p_{T}, etc...). Additional track quality cuts are imposed which include a minimum p_{T} cut of 0.9 GeV, the

number of missing hits, the goodness-of-fit χ^2 of the track fit and the distances to the beamline and a reconstructed primary vertex. Once a track candidate is accepted its component hits are removed from the collection of tracker hits, and the process repeats, until no more track candidates can be found.

Primary vertices are reconstructed by clustering the reconstructed tracks according to the z-coordinate at the point of closest approach to the beam line. An adaptive vertex fit is used on these sets of tracks to find the position of the vertex [115]. The primary interaction vertex is defined as the primary vertex with the largest sum of p_{T}^2 from the objects associated to it. A dedicated algorithm is used to identify secondary displaced vertices within the tracker volume, using the complete set of reconstructed tracks [116, 117]. These displaced vertices mainly originate from nuclear interactions with the tracker material or from conversion photons.

Muon track elements are identified using two different methods. The first matches each track element present in the muon chambers to a track reconstructed in the inner tracking system and a track fit performed using a Kalman filter. These track are called *global* muon tracks. The second method takes all tracks in the tracker and looks for compatible signatures in the calorimeters and in the muon chambers. If a compatible signature if found then the inner track is labelled as a *tracker* muon track.

Calorimetric cluster candidates are constructed individually, but with a similar methodology, for the ECAL and HCAL. Initially, cluster seeds are created which are the local energy deposit in a cell which is maximal with respect to adjacent cells and above a seed energy threshold. Next, bordering cells are added to the seed if they contain a deposit above a given cell energy threshold. All thresholds can be found in [113]. Calorimeter clusters not associated to an extrapolation of a track candidate are clear indicators of a neutral particle. Care must be taken however, when a neutral-particle deposit overlaps a charged-particle deposit. These are detected as an excess in cluster energy with respect to the sum of the p_{T} of the associated charged particles. The method of obtaining an accurate calorimeter calibration in order to maximise the identification of the neutral clusters, while minimising the misreconstructed energy excesses and to get the correct energy scale for all neutral particles is also given in [113].

5.1.2 Linking algorithm

The next step in reconstructing a particle is to use a *link algorithm* to connect all the requisite PF elements together. The probability for the link algorithm to connect two elements together depends on the granularity of the CMS subdetectors and the number of particles to resolve. For a single particle, the probability to link all elements together depends predominantly on the volume of material traversed before the calorimeters due to nuclear interactions introducing trajectory kinks and secondary particles.

A link between the central track and calorimeter cluster is created if the extrapolation of the track candidate from the last hit position to a depth of one interaction length in the calorimeters is within a cluster candidate area. Neutral clusters in the ECAL are linked to electron brehmsstrahlung radiation if a tangential extrapolation (starting from tracker hit position) lies within the boundaries of a cluster candidate and the η displacement between the cluster and original track candidate extrapolation is less than 0.05. The brehmstrahlung photons have a high chance to form a pair of *conversion electrons* ($\mathrm{e}^+\mathrm{e}^-$). A conversion finder is used to find links between compatible track candidates originating from a photon conversion and if found is subsequently linked to the originating track candidate. Links between ECAL, HCAL and preshower cluster candidates are formed if they are outside of tracker acceptance. The cluster position in the more granular subdetector must lie within the cluster envelope of the coarser subdetector. Charged particle tracks associated to a common secondary vertex are linked. Finally, links can be created between a central track candidate and information from the muon

detectors either as a global or a tracker muon track. All PF elements linked together form a *PF block*.

The PF algorithm uses the PF blocks as an input and performs reconstruction and identification in the following order. Muon candidates are identified and reconstructed before removing the corresponding PF elements form the PF block. Next, electrons are processed including brehmsstrahlung radiation concurrently with isolated photons and again associated PF elements removed. Finally, the remaining PF elements are associated to charged hadrons, neutral hadrons and photons eminating from the fragmentation and hadronisation occurring within parton jets. The reconstruction and identification of each *PF object* is described in the following sections.

5.1.3 PF muons

PF muons are identified from the global and tracker muon track properties. First, isolation requirements are imposed on the global muon tracks. Isolation is a measure of the activity of the event around the object of interest. In this case, additional tracks and calorimeter deposits are taken within a cone of

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3, \tag{5.1}$$

and if

$$\sum p_{\mathrm{T}}^{\mathrm{track}} + \sum E_{\mathrm{T}}^{\mathrm{calo}} < 0.1 \times p_{\mathrm{T}}^{\mathrm{muon}},$$

then no further selection is required as hadrons misidentified as muons are rejected adequately. Above this value, charged hadrons are susceptible to be misreconstructed as muons, due to *punch-through* from the back of the HCAL. In this regime, the global muons must pass a tight-muon identification criteria described in Sec. 5.2.2 and either at least three matching hits found in the muon chambers or the calorimeter deposit is compatible with the muon hypothesis. If the muon does not pass the tight-muon ID due to a poor inner or global fit, then it is retained if the tracker muon fit is of high quality and the calorimeter deposit is compatible with the muon hypothesis.

For muons with $p_{\mathrm{T}} < 200$ GeV, the muon track resolution is better for tracker muons than for global muons and so the tracker muon track is used to determine the muon properties. Otherwise they are taken from the track fit with the minimum χ^2.

5.1.4 PF electrons and photons

Due to the similarity in the basic properties of electrons and isolated photons (electrons emit brehmstrahlung photons and photons convert to e^+e^- pairs), the PF algorithm reconstructs isolated PF photons and PF electrons simultaneously. PF electrons are seeded from a cluster candidate linked to a GSF (Gaussian sum filter) track [118], given that it is not additionally linked to more than three other tracks. Isolated PF photons are seeded from clusters > 10 GeV with no link to a GSF track.

GSF tracks are based on a combination of two different methods of track reconstruction. The first is the *ECAL-based* method, which extrapolates to the inner track hits depending on the ECAL cluster energy and position. To seed the ECAL-based track all clusters of $E_{\mathrm{T}} > 4$ GeV are used with the assumption that they are produced by either an electron or positron. A large proportion of the energy of the electron or positron is lost to bremhstrahlung and so the performance of the method depends strongly on the ability to collect all the associated energy depositions and without depositions from other particles. To collect the energy, the cluster is extended in ϕ around the electron direction to account for the bending in the magnetic field. The group of clusters is known as a *supercluster*. The ECAL-based track seeds suffer at low-p_{T}

due to the large bending in the magnetic field and in jets due to the non-isolated environment. The second is the *tracker-based* method and recovers electron tracks missed by the ECAL-based approach. It starts from the iterative-based track candidates with a $p_{\mathrm{T}} > 2$ GeV, with more specific requirements of the number of missing hits and the χ^2 of the track fit. A GSF fit (which is reduced) is used on the track hits as it is more suitable to electron fitting than a Kalman filter. A final cut on a boosted-decision-tree that combines the discriminating power of a number of distributions, e.g. the χ^2 of the GSF fit, the distance between the extrapolated track and the nearest cluster and the energy lost along the GSF track, is used to select tracker-based track seeds. The track seeds obtained by the two methods are combined together and a final full GSF fit processed.

PF electrons are required to satisfy additional identification criteria on both the track seed and cluster properties. These criteria include the track fit χ^2, the ratio of energy in the HCAL to the ECAL and the number of hits. A full set of parameters is given in [113]. PF photons are selected depending on compatibility with the photon hypothesis of the isolation with respect to GSF tracks and clusters, the shower shape and the ECAL to HCAL energy ratio.

5.1.5 PF jets

Once the electrons, isolated photons and muons have been reconstructed and removed from the PF blocks, the remaining particles must be from neutral hadrons, charged hadrons or non-isolated photons formed from the hadronisation of the final state. Hadronic jets are composed of these hadrons and photons, where the photon content deposits primarily in the ECAL, the neutral hadron content in the HCAL and charged hadron content in both calorimeters.

All calorimetric PF elements that are not linked to a tracking element are considered to be either neutral hadrons or photons. Within the tracking acceptance the charged and neutral hadrons can be identified separately and as the neutral hadrons only leave $\approx 3\%$ of the jet energy in the ECAL, the ECAL deposits are considered to originate solely from photons and the HCAL deposits from neutral hadrons. Outside the tracking acceptance however, there is no way to separate charged hadrons from neutral hadrons. As the charged hadrons also leave deposits in the ECAL, it is not possible to assign the deposits to photons alone. Now the links between ECAL deposits and HCAL deposits become important. ECAL deposits not linked to HCAL are considered to be photons, while linked deposits are considered to arise from a neutral or charged hadron.

Once the neutral hadrons and non-isolated photons have been identified and removed, there are only calorimeter elements linked to one or more track elements remaining. These make up the charged hadrons. PF jets are created by using the anti-k_{T} jet clustering algorithm on particles reconstructed from PF algorithm, with a distance parameter $\Delta R < 0.4$.

Anti-k_{T} jet clustering algorithm

The charged and neutral hadron deposits are clustered using the anti-k_{T} jet clustering algorithm [119]. It is an extension to the Cambridge/Aachen and k_{T} jet clustering algorithms. These algorithms work by calculating the distance between two entities, either particles or proto-jets d_{ij} and the distance between an entity and the beam line, d_{iB}. If the smallest distance is between two entities then they are recombined into a new proto-jet, whereas if d_{iB} is smaller, then i is called a jet and removed from the list of entities. All distances are then recalculated and the clustering repeated until no entities remain. The maximum cone size is given by the double of the radius parameter, R. The differences between the anti-k_{T} algorithm occur in the definition of the distance parameters

$$d_{ij} = \min(k_{Ti}^{2p}, k_{Tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \tag{5.2}$$

and

$$d_{iB} = k_{Ti}^{2p}. \tag{5.3}$$

The parameter p gives the strength of the relative contributions to d_{ij} and d_{iB} from the hardness of the entity and the angular separation. The angular separation is defined as

$$\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}.$$

For $p = +1$ the k_T algorithm is recovered and the Cambridge/Aachen algorithm when $p = 0$. The anti-k_T algorithm uses $p = -1$ which means that precedence is given to the hardness of the entities, e.g. a soft particle is much more likely to cluster to a hard particle than another soft one. This leads to perfectly conical jets if no two initial hard particles are within $2R$ of each other. If there are two hard particles within $R < \Delta_{12} < 2R$ but $k_{T1}^2 >> k_{T2}^2$ then the hardest jet will be conical and the softer a cone missing the overlap. If $k_{T1}^2 \sim k_{T2}^2$ then the jet boundary will lie between the two jet axes. For closely spaced hard particles within $\Delta_{12} < R$, clustering will occur, but for $k_{T1}^2 \sim k_{T2}^2$ the shape becomes a union of the cones of both constituent hard particles as well as the cone centred on the final jet. This means that the anti-k_T algorithm is resilient against soft radiation but adaptive to hard emission. An example anti-k_T jet clustering is shown in Fig. 28.

Figure 28: An example of clustering using the anti-k_T algorithm. Soft particles are preferentially clustered around high-p_T seeds forming conical jets. The jet boundary is defined by the respective hardness between neighbouring jets. Figure taken from [119].

As the particle flow algorithm attempts to reconstruct every individual particle in an event, before the jet clustering, it is possible to assign charged hadrons to either a high-p_T interaction vertex or an additional interaction vertex. If definitely matched to a pileup vertex then the associated jet constituents are removed. This is known as *charged hadron subtraction*. If there is no clear association then the charged hadrons remain in the jet. Pileup subtraction is discussed further in Sec. 6.2.

Identification of b quark jets

The properties of a jet arising from a b quark are different enough to those of a jet formed from a uds quark or gluon that it is possible to discriminate between them. The ability to do this is very important for many analyses, especially those involving top quark physics because it reduces the impact from background processes considerably. The crucial aspect of this discrimination arises from the formation of b hadrons from the b parton, which have a longer lifetime and typically travel a few millimetres in the detector before decaying. Figure 29 shows an illustration of a heavy-flavour jet associated to a secondary vertex. The discriminating

Figure 29: An illustration of a jet from a heavy flavour quark. A hadron is formed with a lifetime of $\sim 1 - 1.5$ ps which decays at a displaced secondary vertex (SV) with respect to the primary interaction vertex (PV) and hence has a large impact parameter (IP). Figure taken from [120]

variables include the displaced tracks from the secondary vertex, the impact parameter, the increased fragmentation due to the higher initiating mass which can lead to both increased p_{T} measurements with respect to the jet axis and a greater proportion of charged leptons in the jet constituents.

The *combined secondary vertex* algorithm (CSVv2) [120] takes this information and assigns a discriminator value between 0 and 1. The closer the discriminating value to 1 the more likely the jet is to be a b jet. The flavour composition of jets with respect to CSVv2 discriminant is shown in Fig. 30, using a basic event selection applied to a $\mathrm{t\bar{t}}$ simulation. It is immediately clear that applying a requirement for two high CSVv2 values, enhances the number of events with two jets emanating from b quarks and thus the number of signal $\mathrm{t\bar{t}}$ events with respect to background processes.

5.1.6 PF missing transverse momentum

Conservation of momentum states that the sum of all transverse momentum is zero. The presence of a momentum imbalance $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, implies an undetectable particle in the event, e.g. a neutrino. In terms of particle flow, $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is given by the negative vector sum of all the present particle flow objects

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_{i}^{\text{PF objects}} \vec{p}_{\mathrm{T}}^{\,i}.$$

Figure 30: The flavour composition of jets from a $t\bar{t}$ model simulation. A simple event selection is performed requiring an isolated lepton, four jets with $p_T > 30$ GeV of which two are required to have CSVv2 > 0.8484. Figure taken from [120].

More generally, the magnitude of this missing momentum is used p_T^{miss}. In most events p_T^{miss} is small, however there is a small probability that it is reconstructed to be artificially large. The cause of this is most often the misreconstruction or misidentification of a high-p_T muon. If an event has a large p_T^{miss} then the PF algorithm applies a post processing step. The first issue investigated are genuine muons from the cosmic ray muon background coincident with a bunch crossing. They are recognised as such if their trajectory is more than 1 cm away from the beam line and removed from the particle flow collection is the p_T^{miss} is reduced by at least one half. The second issue relates to bad reconstruction of the momentum of the muon, notified by significant differences in the global and tracker muon p_T estimates. The bad reconstruction can be caused by interactions in the steel return yoke, in-flight decays and incorrectly assigned hits in the inner tracker. If the p_T^{miss} is reduced by at least one half when recalculating with respect to the different muon p_T estimates then the muon estimate with the lowest p_T^{miss} is taken. The third issue comes from the misidentification of a charged hadron punching through into the muon chambers. A muon is reconstructed together with an energetic neutral hadron from the calorimeter deposits. A similar issue is that of an overlapping muon and neutral hadron deposit being misreconstructed as a charged hadron deposit. The particles are substituted with each other respectively if the p_T^{miss} is reduced by at least one half.

The p_T^{miss} is corrected accordingly when jet energy corrections (JEC), explained in Sec. 6.2, are applied to the PF jets.

$$\vec{p}_T^{\,miss} = -\sum_i^{\text{PF objects}} \vec{p}_T^{\,i} - \sum_j^{\text{PF jets}} (\vec{p}_T^{\,j,\,corr} - \vec{p}_T^{\,j}).$$

5.1.7 PF isolation

The requirement of leptons to be isolated has already been used in the reconstruction of the PF objects. Often, more stringent isolation selections are required and so the *PF relative isolation* I_{rel} is employed. It is defined as the ratio of the additional p_{T} within a cone of set distance around the lepton to the p_{T} carried by the lepton. The additional p_{T} includes all charged and neutral hadrons as well as photons present within the cone.

$$I^{\ell}_{\text{rel}} = \frac{\sum p_{\text{T}}^{\text{charged}} + \sum p_{\text{T}}^{\text{neutral}} + \sum p_{\text{T}}^{\gamma}}{p^{\ell}_{\text{T}}}. \tag{5.4}$$

In addition, contributions from additional interactions can be taken into account. Charged particles from these extra interactions are not associated to the primary vertex and are removed. Direct removal is not possible for neutral particles produced by the pileup. One possible correction for this neutral pileup is the $\Delta\beta$ correction. This uses the ratio of charged hadrons to neutral hadrons and photons in inelastic collisions to estimate the contribution of non-charged pileup from charged pileup. The correction $\Delta\beta$ is set as 0.5. The relative isolation then becomes

$$I^{\ell}_{\text{rel}} = \frac{\sum p_{\text{T}}^{\text{charged}} + \max(0, \sum p_{\text{T}}^{\text{neutral}} + \sum p_{\text{T}}^{\gamma} - \Delta\beta \sum\limits_{\text{pileup}} p_{\text{T}}^{\text{charged}})}{p^{\ell}_{\text{T}}}. \tag{5.5}$$

5.2 Analysis object selection

While the PF algorithm is good at resolving particles, it is sometimes necessary to impose stricter identification criteria. By increasing the probability that an analysis object is correctly reconstructed, the selection efficiency for that object decreases correspondingly. Where the statistical uncertainty is not an issue or where an analysis requires a high object purity *tighter* cuts tend to be used. This is the case for $\text{t}\bar{\text{t}}$ events produced at the LHC, having both a high cross section and large data set sizes. In the following the identification criteria for all the objects used in this thesis will be detailed.

5.2.1 Electron objects

As stated in Sec. 4.3.8, events that pass the HLT_Ele32_eta2p1_WPTight_Gsf trigger are selected as the basis of the e+jets channel in this thesis. A harder p^{ℓ}_{T} cut is chosen at $p^{\ell}_{\text{T}} > 34$ GeV to operate in a region optimal with respect to the trigger efficiency. The $|\eta^{\ell}| < 2.1$ cut is maintained, however electrons falling into the gap between the barrel and endcap ($1.4442 < |\eta^{\ell}| < 1.566$) are discarded. The signal electron object must pass a set of strict identification requirements (*tight* ID). An additional collection of veto electrons are selected using a lower $p^{\ell}_{\text{T}} > 15$ GeV and a looser set of identification criteria (*veto* ID).

Electron identification

Further identification requirements are placed on the PF electrons to be used as analysis objects. The additional requirements are combined into an electron identification which provides more discrimination for isolated prompt signal electrons against hadronic jets, fake electrons and other background sources of electrons. The electron identification is provided by the E/gamma *physics object group* (POG) for different working points in the barrel and endcap regions of the detector. The tight and veto IDs are shown in Tab. 6. The shower shape variable $\sigma_{i\eta i\eta}$ is the root-mean-square of the position of the energy deposited in a 5×5 lattice centred in the

seed crystal. An electromagnetic shower will be much more confined than a hadronic shower and therefore have a lower value of $\sigma_{i\eta i\eta}$. The absolute differences in η and ϕ between the supercluster position and the track direction at the primary vertex extrapolated to the ECAL assuming no brehmstrahlung radiation is given by $|\Delta\eta_{in}|$ and $|\Delta\phi_{in}|$. Additional particles present in hadronisation processes bias the position of the supercluster due to the increased bending in low-p_{T} radiations. The ratio of the hadronic energy to the electromagnetic energy H/E is smaller for deposits from electrons than from hadrons. A tighter requirement on the PF isolation with effective area correction, using a cone size of $\Delta R < 0.3$, is included. The effective area method subtracts the product of the effective area of the cone and the offset energy density, visited in Sec. 6.2. It is more suitable than the $\Delta\beta$ correction due to being insensitive to the electron radiations. The variable $|1/E - 1/p|$ is proportional to the radius of curvature of the electron track and gives additional discrimination against electrons from conversion photons. The missing inner hits requirement provides a stricter track quality acceptance than present in the PF algorithm. A selected electron is not allowed to originate from an identified conversion photon. In addition to the electron identification, the electron must originate close to the primary vertex d_0 and close to the beam line d_z.

Table 6: The identification criteria for signal electrons (tight) and veto electrons (veto). The critera are split into the barrel and endcap regions at $|\eta_{SC}| = 1.479$

		Tight		**Veto**	
Cut		**Barrel**	**Endcap**	**Barrel**	**Endcap**
full $5 \times 5\sigma_{i\eta i\eta}$	$<$	0.00998	0.0292	0.0115	0.037
$\lvert\Delta\eta_{in}\rvert$	$<$	0.00308	0.00605	0.00749	0.00895
$\lvert\Delta\phi_{in}\rvert$	$<$	0.0816	0.0394	0.228	0.213
$\frac{H}{E}$	$<$	0.0414	0.0641	0.356	0.211
PF I_{rel} with EA correction	$<$	0.0588	0.0571	0.175	0.159
$\lvert\frac{1}{E} - \frac{1}{p}\rvert$	$<$	0.0129	0.0129	0.299	0.15
Missing Inner Hits	$<=$	1	1	2	3
Conversion Veto		Y	Y	Y	Y
Additional to ID					
d_0 (cm)	$<$	0.05	0.10	0.05	0.10
d_z (cm)	$<$	0.10	0.20	0.10	0.20

5.2.2 Muon objects

Signal analysis muons must have a $p_{\mathrm{T}}^{\ell} > 26$ GeV cut, similarly imposed from the efficient operation region of the HLT_IsoMu24 $\cup$ HLT_IsoTkMu24 trigger selection and must pass tight ID requirements. Veto analysis muons have a $p_{\mathrm{T}}^{\ell} > 15$ GeV, aligned with that in the e+jets channel and must pass a *loose* muon ID.

Muon identification

Analysis muons are selected from the PF muons by applying a muon identification in an identical manner to the analysis electrons. The criteria of the identity are listed in Tab. 7 and are supplied by the muon POG. For signal muons, hits are required along the entire path of the muon from the silicon pixel detector to the muon chambers. A good χ^2/ndf of the track fit is also required. Constraints are placed with respect to the primary vertex and beam line as with the electron

identification. Finally, a more stringent cut on the PF I_{rel} with $\Delta\beta$ correction using a cone size of $\Delta R < 0.4$ is required.

Table 7: The identification criteria for signal muons (tight) and veto muons (loose).

Cut		**Tight**	**Loose**
is PF muon		Y	Y
is global or tracker muon		Y	Y
Pixel hits	$>$	0	—
Tracker layer hits	$>$	5	—
Muon stations hit	$>$	2	—
Muon chambers hit	$>$	1	—
χ^2/ndf	$<$	10	—
d_0 (cm)	$<$	0.2	—
d_z (cm)	$<$	0.5	—
Additional to ID			
PF I_{rel} with $\Delta\beta$ correction	$<$	0.15	0.25

5.2.3 Jet Objects

On top of the charged hadron subtraction in the PF algorithm, analysis jets are required to have $p_{\text{T}} > 30$ GeV, $|\eta| < 2.4$ and pass a loose ID supplied by the Jet/MET POG, shown in Tab. 8. To ensure a jet is not additionally classified as a lepton it must have a minimum distance to the nearest lepton $\Delta R > 0.4$. If the jet is close to a lepton then it is assumed to have originated from that lepton and is consequently removed from the jet collection. A jet is additionally classified as a b jet, if it also passes the medium working point of the CSVv2 algorithm.

Table 8: The identification criteria for signal jets.

Cut		**Loose**
Neutral hadron fraction	$<$	0.99
Charged hadron fraction	$>$	0
Neutral electromagnetic fraction	$<$	0.99
Charged electromagnetic fraction	$<$	0.99
Number of constituents	$>$	1
Number of charged constituents	$>$	0
Muon fraction		—

CHAPTER 6

Analysis formation, event selection and corrections

With the complete reconstruction of events in both data and simulation it is possible to begin forming an analysis. Several variables are defined to be measured and an appropriate event selection must be formed, dependent on the event topology that is being investigated. Corrections still need to be applied to the analysis objects to resolve any remaining differences between data and simulation. Before the event selection, data objects are corrected depending on the response of the subcomponents in the detector, e.g. there should be no angular dependence on the energy of a particle. Then the simulation is then corrected to match what is observed in the data, e.g. due to differences in the efficiencies of the various reconstruction algorithms applied. Once the $t\bar{t}$ yields have been calculated after all corrections are applied, then the cross sections can be measured.

6.1 Kinematic event variables and the event selection

As stated in Sec. 2.6.4, this thesis measures seven event variables which are N_{jets}, H_{T}, S_{T}, $p_{\text{T}}^{\text{miss}}$, p_{T}^{W}, p_{T}^{ℓ} and $|\eta^{\ell}|$. They are defined in the following way:

- N_{jets} is the total multiplicity of jets in the event with $p_{\text{T}} > 30$ GeV and $|\eta^{\ell}| < 2.4$.
- H_{T} is the scalar sum of the p_{T} of these jets.
- $p_{\text{T}}^{\text{miss}}$ is defined as the magnitude of the missing transverse momentum $\vec{p}_{\text{T}}^{\text{miss}}$, as defined in Sec. 5.1.6.
- p_{T}^{ℓ} is the magnitude of the transverse momentum of the signal lepton.
- $|\eta^{\ell}|$ is the magnitude of the pseudorapidity of the signal lepton.
- S_{T} is the sum of the H_{T}, p_{T}^{ℓ} and $p_{\text{T}}^{\text{miss}}$ event variables.

- $p_{\mathrm{T}}^{\mathrm{W}}$ is the magnitude of the transverse momentum of the leptonically decaying boson defined as

$$p_{\mathrm{T}}^{\mathrm{W}} = \sqrt{(p_x^{\ell} + p_x^{\mathrm{miss}})^2 + (p_y^{\ell} + p_y^{\mathrm{miss}})^2}.$$

The event selection is performed using the collections of particles created following the methods described in Sec. 5.2. The signal event selection requires a single isolated tight lepton, either electron or muon, with zero additional veto leptons. A minimum of four jets is needed, with at least two being tagged as coming from a b quark.

6.2 Jet energy corrections

The response from the detector calorimeters to particles is not linear which means that a simple translation between reconstructed jet energy and particle-level jet energy is not feasible. Jet energy corrections (JEC) are used to provide the mapping between the measured and particle-level jet energies. The JEC are applied sequentially with each step correcting for a different effect. These steps are shown in Fig. 31. Each level applies a scale factor correcting the jet

Figure 31: A schematic showing the factorised steps to correct the energy scale of a jet. These consist of a pileup correction, detector response correction with additional residuals applied in data and optional flavour corrections. Schematic taken from [121].

energy with respect to its associated parameters (p_{T}, η, flavour, etc...).

The first correction is applied to counter additional energy in the jet, known as *offset energy*, coming from pileup and the underlying event. In-time pileup is already partially removed by the charged hadron subtraction algorithm, which accounts for $\sim 50\%$ of the offset energy. The rest of the diffuse neutral energy and out-of-time pileup is removed on an event-by-event basis using the *hybrid jet area method* dependent on the offset energy density ρ and the effective jet area A^{jet}. The variable ρ is taken as the median of the energies present in a grid of $\eta - \phi$ cells. The use of the median creates an insensitivity to hard jets in the event. The correction is applied as

$$p_{\mathrm{T}}^{\mathrm{corr}} = p_{\mathrm{T}}^{\mathrm{uncorr}} - A^{\mathrm{jet}}\Big(\rho_0(\eta) + \rho\beta(\eta)[1 + \gamma(\eta)\log(p_{\mathrm{T}}^{\mathrm{uncorr}})]\Big), \tag{6.1}$$

where ρ_0, β and γ are η dependent variables which are estimated from simulated QCD dijets processed with and without pileup overlay and parametrise the shape of the offset energy distribution. Residual corrections applied to data which stem from differences between the data and the detector simulation are determined using the random cone method on zero-bias events. Zero-bias events are events in which there is no hard-scattering process and so any jet energies should come from pileup provided the electronic noise is small. The random cone method estimates ρ by clustering many jets around random cones mapping all $\eta - \phi$ space and taking the average jet p_{T}. This correction removes any dependence of the jet energy on the luminosity for the subsequent corrections

Next, a second set of corrections are applied to both data and simulation to create a uniform jet energy response with respect to p_{T} and η. This is done using a simulated QCD dijet sample to compare the p_{T} of the reconstructed jets to the particle-level jets. The response is defined as the ratio of the average reconstructed jet p_{T} to the average particle-level jet p_{T}. The simulated jet response at the end of Run I, which is used in the JEC applied in this thesis, can be seen in Fig. 32. Preliminary measurements of the simulated jet energy response for detector conditions during 2016 are given in [122].

Figure 32: The simulated jet energy response as a function of p_{T} and η. These corrections are based on the status of the detector after the end of Run I [121].

Residual corrections are calculated from data and applied to correct small differences seen in the jet energy response between simulation and data. These consist of a η dependent correction determined from QCD dijet events relative to a barrel reference jet of similar p_{T} and a correction on jet energy scale with respect to p_{T}. The p_{T} correction is based on Z/γ+jets and QCD multijet events. Finally, optional corrections based on the response of different particle flavours can be applied. These corrections are derived from simulation and are not applied in this thesis.

As well as variations present in the jet energy scale, variations are also present in the jet energy resolution (JER). The JER scale factors $\mathrm{SF}_{\mathrm{JER}}$ are determined from the ratio of data to simulated QCD dijet events with respect to p_{T}, η and additional interaction vertices. The JER of jets in simulation $\sigma_{\mathrm{JER}}^{\mathrm{sim}}$ is also better than that seen in data and so must be smeared to match. A hybrid of two approaches is used to correct and smear the resolution in simulation. If a reconstructed jet is well matched to a particle-level jet such that

$$\Delta R < \frac{R_{\mathrm{cone}}}{2}$$

and

$$|p_{\mathrm{T}} - p_{\mathrm{T}}^{\mathrm{ptcl}}| < 3\,\sigma_{\mathrm{JER}}^{\mathrm{sim}}\,p_{\mathrm{T}}, \tag{6.2}$$

then the jet p_{T} is rescaled by

$$1 + (\mathrm{SF}_{\mathrm{JER}} - 1)\frac{p_{\mathrm{T}} - p_{\mathrm{T}}^{\mathrm{ptcl}}}{p_{\mathrm{T}}}.$$

However, if the jet is not matched then a stochastic smearing is applied

$$1 + \mathcal{N}(0, \sigma_{\mathrm{JER}}^{\mathrm{sim}})\sqrt{\max(\mathrm{SF}_{\mathrm{JER}}^2 - 1, 0)}.$$

$\mathcal{N}(0, \sigma_{\mathrm{JER}}^{\mathrm{sim}})$ represents a random number sampled from a Gaussian distribution centred on 0 with a standard deviation given by the JER in simulation.

6.3 Reweighting from additional interactions

The number of additional interactions in simulation differs to that seen in data. The left panel of Fig. 33 shows the difference seen in the number of reconstructed vertices after full event selection. The true vertex multiplicity distribution in simulation is reweighted to match that seen in data. This is done by creating a scale factor from the number of additional interactions seen in data and simulation. The true distribution of the number of additional interactions in data $N_{\mathrm{PU}}^{\mathrm{Data}}$ is estimated every bunch crossing using

$$N_i^{\mathrm{True}} = \sigma_{\mathrm{minbias}} \times \mathcal{L}_i.$$

The parameter $\sigma_{\mathrm{minbias}}$ is the total inelastic cross section taken as 69.2 mb and $\mathcal{L}_i$ the instantaneous luminosity in the luminosity section i. The true distribution of additional interactions in simulation $N_{\mathrm{PU}}^{\mathrm{Sim}}$ also includes interactions from out-of-time pileup. The number of reconstructed vertices in simulation is then reweighted by

$$\frac{N_{\mathrm{PU}}^{\mathrm{Data}}}{N_{\mathrm{PU}}^{\mathrm{Sim}}}, \tag{6.3}$$

and the application is shown in the right panel of Fig. 33. While a discrepancy is seen in the reweighting, the distributions of the vertex multiplicities matches when considering the uncertainties on $\sigma_{\mathrm{minbias}}$ ($\pm 4.6\%$).

Figure 33: A plot showing the distribution of the number of reconstructed primary vertices in simulation and data. The left panel shows the distribution before pileup reweighting and the right panel after.

6.4 Reweighting from b jet identification

The b jet identification efficiencies for a given configuration of jets in data and simulation are different. The simulation is corrected to match data by applying a weight

$$w = \frac{P(\text{Data})}{P(\text{Sim})},$$

where

$$P(\text{Sim}) = \prod_{i\,=\,\text{tagged}} \epsilon_i \prod_{j\,=\,\text{not tagged}} (1 - \epsilon_j)$$

and

$$P(\text{Data}) = \prod_{i\,=\,\text{tagged}} \text{SF}_i \epsilon_i \prod_{j\,=\,\text{not tagged}} (1 - \text{SF}_j \epsilon_j).$$

The parameters ϵ_i and SF_i are the b quark tagging efficiency and scale factor ($\epsilon_{\text{DATA}}/\epsilon_{\text{MC}}$) for the jet i, respectively. The scale factors applied to the b and light flavour jets in this thesis are derived in [120].

Both the b tagging scale factors and efficiencies are calculated and applied as functions of p_{T}, η and flavour (b, c, udsg). The b jet tagging efficiency and light jet mis-tagging efficiencies are calculated from the POWHEG+PYTHIA $\text{t}\bar{\text{t}}$ simulation. They are taken as the fraction of tagged jets over the total number of b jets. Figure 34 shows the b quark jet tagging efficiencies, which for genuine b jets is $\approx 70\%$ and for misidentified jets from light quarks is $\approx 1\%$.

Figure 34: The left panel shows the b tagging efficiency with respect to jet p_{T} and the right panel with respect to jet η. Both panels show the efficiencies separated into b, c and light flavours.

Figure 35 shows the b jet multiplicity after full event selection before (left panel) and after (right panel) applying the b jet reweighting. The discrepancy in the first bin is covered by the uncertainty in multijet QCD, taken from simulation. The size of the multijet QCD normalisation uncertainty is $\approx 60\%$ motivated in Sec. 6.6.

Figure 35: The left panel shows the b jet multiplicity distribution before reweighting and the right panel after reweighting.TODO add QCD error bin 1

6.5 Reweighting from leptons

The reconstruction, identification and triggering efficiencies of leptons are measured in data and corrected in simulation to match data. The electron efficiencies and scale factors are derived using the tag-and-probe method following prescriptions defined in [123] using the clustering algorithm described in [113] and are presented in [124]. Muon scale factors are measured in a similar manner and are presented in [125, 126]. In both channels the efficiencies are measured from events containing a Z boson and the corrective scale factors are applied as a function of p_{T} and supercluster pseudorapidity η_{SC}. The total lepton scale factors vary between 0.95 and 1.

The electron ID scale factors are produced in bins of η_{SC} coarse enough to have significant effects on the distribution of $|\eta^{\ell}|$. This is because the high resolution of $|\eta^{\ell}|$ allows for fine-bin measurements and the application of the scale factors would cover several bins. It is therefore not appropriate to use these electron ID scale factors for the $|\eta^{\ell}|$ event variable. They are rederived with respect to the finer binning used in the production of the electron reconstruction scale factors for an identical p_{T} range. An identical method used for the original scale factors is used in the rederivation. Figure 36 shows the efficiency of the electron ID in data (points) and simulation (dashes) and the scale factor to be applied is the ratio. This new scale factor is only applied for the $|\eta^{\ell}|$ differential measurements.

6.6 Multijet QCD

The multijet QCD background is notoriously hard to model and as such a data driven approach is used to estimate its contribution in the signal region from an orthogonal control region. The control region is designed to be enhanced in multijet QCD and to achieve that in the e+jets channel the QCD event selection inverts the isolation requirement in the electron ID and requires exactly zero b jets. For the μ+jets channel the control region event selection requires $0.15 < I_{\mathrm{rel}} < 0.30$ and zero b tagged jets. In these control regions, the contribution of $\mathrm{t\bar{t}}$, single top quark, and V+jets background is estimated from simulation to be $\approx 15 - 20\%$.

Figure 36: The electron identification efficiency for data and simulation calculated with events with a Z boson present using the tag-and-probe method. The ratio of the two efficiencies is the finer binned scale factors applied in the $|\eta^{\ell}|$ measurements.

These estimations are subtracted from the data to extract the shape of the multijet QCD in data. The QCD distributions are then scaled to the signal region by a transfer factor derived from simulation

$$t_f = \frac{N^{\text{Signal}}_{\text{QCD, MC}}}{N^{\text{Control}}_{\text{QCD, MC}}},$$

where $N^{\text{Signal}}_{\text{QCD, MC}}$ and $N^{\text{Control}}_{\text{QCD, MC}}$ are the total normalisation of the QCD simulation in the signal and control regions respectively. Transfer factors for all nominal control regions are shown in Tab. 9.

Table 9: The transfer factors to the signal region from the four multijet QCD control regions.

Control region	**Transfer factor**
Non-isolated electron	0.27
Conversion electron	0.03
Non-isolated muon $0.15 < I_{\text{rel}} < 0.30$	0.11
Non-isolated muon $0.30 < I_{\text{rel}} < \infty$	0.08

To validate the use of the nominal control region, the QCD contribution is estimated in an alternate control region and compared to background subtracted data in that control region. The comparison can also yield the shape and normalisation uncertainties on the data driven QCD estimate. In the e+jets channel the alternate control region uses an electron ID where the conversion veto or missing hits criteria has been inverted and in the μ+jets channel the $0.30 < I_{\text{rel}} < \infty$ region is used. Both still require exactly zero b tagged jets. The data-simulation comparison plots of the event variables in the multijet QCD control regions for the e+jets channel are shown in Figs. 37 and 38 and for the μ+jets channel in Figs. 39 and 40.

Figures 41 and 42 show the data-driven QCD estimates in the alternate control region compared to the QCD calculated by subtracting backgrounds from the data in the alternate control region. The differences between the two predictions show the accuracy of the prediction, which is expected to be similar for the prediction in the signal region. The uncertainty of the QCD prediction in the signal region can therefore be taken from the discrepancies seen in this control region. Any offset from unity in the ratio of the two QCD yields gives the normalisation uncertainty, taken as ±60% (gold band) and the spread of the ratio gives the shape uncertainty, taken as ±30% (grey band). The regions where the ratio is outside these uncertainty bands the expected contribution of QCD is negligible, e.g. at high $p_{\text{T}}^{\text{miss}}$.

Figure 37: The distributions of the event variables and I_{rel} in the non-isolated electron control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars.

Figure 38: The distributions of the event variables and I_{rel} in the conversion electron control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars.

Figure 39: The distributions of the event variables and I_{rel} in the non-isolated ($0.15 < I_{\mathrm{rel}} < 0.30$) muon control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars.

Figure 40: The distributions of the event variables and I_{rel} in the non-isolated ($0.30 < I_{rel} < \infty$) muon control region. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in data given by the vertical error bars.

Figure 41: The prediction of multijet QCD in the alternate control region from the nominal control region for each event variable in the electron channel. Also shown as a comparison is the data-subtracted QCD estimate from the alternate control region. Below each of the distributions, the ratio of the two QCD predictions is shown. This ratio gives an estimation of the QCD normalisation uncertainty by the displacement from unity and was taken to be ±50% shown by the gold band. The QCD shape uncertainty is given by the spread of the ratio, taken to be ±30% shown by the grey band.

Figure 42: The prediction of multijet QCD in the alternate control region from the nominal control region for each event variable in the muon channel. Also shown as a comparison is the data-subtracted QCD estimate from the alternate control region. Below each of the distributions, the ratio of the two QCD predictions is shown. This ratio gives an estimation of the QCD normalisation uncertainty by the displacement from unity and was taken to be ±50% shown by the gold band. The QCD shape uncertainty is given by the spread of the ratio, taken to be ±30% shown by the grey band.

Table 10: The yield of $t\bar{t}$ estimated from the data for each event variable, after the background has been subtracted and fakes removed.

N_{jets}		H_{T}		S_{T}		$p_{\text{T}}^{\text{miss}}$		p_{T}^{W}		p_{T}^{ℓ}		$\lvert\eta^{\ell}\rvert$	
e+jets	μ+jets	e+jets	μ+jets	e+jets	μ+jets	e+jets	μ+jets	e+jets	μ+jets	e+jets	μ+jets	e+jets	μ+jets
119093	199104	24766	43648	17251	34976	93510	148677	36616	73313	24136	95549	46964	68369
62708	102831	43925	76303	46290	82681	90590	154097	81633	141297	56492	86796	47170	71588
24527	39321	46982	78608	50965	85795	26419	45211	60153	91306	44433	61162	42412	62465
8032	12904	36136	59113	38548	60701	5383	8450	27720	38543	31077	40540	35970	52702
2392	3668	24622	38966	26213	39828	1292	2050	8535	11006	20509	26049	22080	43279
975	1392	16473	25672	15801	23594	588	825	2280	2818	13232	16857	12358	29249
—	—	9941	15164	9751	13932	—	—	870	1040	9155	10745	7988	14065
—	—	6159	9113	5671	7876	—	—	—	—	5938	6966	3009	17546
—	—	3700	5401	3115	4335	—	—	—	—	3889	4565	—	—
—	—	2151	3100	1863	2453	—	—	—	—	2653	3013	—	—
—	—	1107	1670	998	1327	—	—	—	—	1899	2100	—	—
—	—	715	887	589	742	—	—	—	—	1228	1398	—	—
—	—	861	1222	565	755	—	—	—	—	902	988	—	—
—	—	—	—	—	—	—	—	—	—	585	696	—	—
—	—	—	—	—	—	—	—	—	—	595	614	—	—
—	—	—	—	—	—	—	—	—	—	494	559	—	—
—	—	—	—	—	—	—	—	—	—	538	642	—	—

6.7 Data-simulation agreement in the signal region

Figures 43 and 44 show the agreement between the data and the sum of the signal and background events from simulation in the combined e+jets and μ+jets channels. All corrections previously described have been applied and the multijet QCD contribution estimated from data. A total of 662 381 events are measured in data, the composition of which is predicted to be 92.1% $t\bar{t}$ production, 4.4% single top quark production, 2.1% V+jets production and 1.4% from multijet QCD. The total level of agreement between simulation and data ($< 0.2\%$) indicates the total cross section is compatible to that stated in Sec. 2.6.1. The statistical uncertainty in the data is given by vertical error bars. The hatched bands shows the statistical uncertainty in the number of simulated events and experimental uncertainties combined in quadrature. The set of experimental uncertainties include those from the JEC and $p_{\text{T}}^{\text{miss}}$, the luminosity, the pileup, lepton and b tagging scale factors and the background predictions. These uncertainties are explained in more detail in Sec. 8.2.1. The ratios of the event variables that are correlated to the p_{T} of the top quark show that the $p_{\text{T}}^{\text{top}}$ is softer in data than in simulation. This is a well known issue, with other analyses reporting the same effect. By reweighting the top quark p_{T} distribution to that measured by [66, 67], the slope observed in the ratio of these variables is corrected, as shown in Fig. 45.

6.7.1 Calculating the $t\bar{t}$ yield

The $t\bar{t}$ yield can be extracted from data by subtracting off the simulated backgrounds and data-driven multijet QCD estimate. Events that are reconstructed but do not enter the visible phase space at particle level, known as fakes and detailed in Sec. 7.1, are also subtracted. These extracted yields, shown in Tab. 10, are still dependent on the detector resolution and would make any cross section measurement comparison only valid where both are smeared by the CMS response. This is time consuming in terms of simulation and does not allow like-for-like comparisons with other experiments. The effects from the detector acceptance, efficiency and bin-to-bin migrations is corrected for by the process known as *unfolding*.

Figure 43: The distributions of N_{jets}, H_{T} and S_{T} after full event selection. The $\text{t}\bar{\text{t}}$ simulation is normalised to the NNLO prediction. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in the data shown by the vertical error bars. The statistical uncertainty in the number of simulated events quadratically combined with the experimental uncertainties is shown by the hatched band.

Figure 44: The distributions of p_T^{miss}, p_T^W, p_T^ℓ and $|\eta^\ell|$ after full event selection. The $t\bar{t}$ simulation is normalised to the NNLO prediction. The ratio of the number of events in data to that in simulation is shown below each of the distributions, with the statistical uncertainty in the data shown by the vertical error bars. The statistical uncertainty in the number of simulated events quadratically combined with the experimental uncertainties is shown by the hatched band.

Figure 45: The distribution of the H_T variable after full event selection and the application of top quark p_T reweighting. With the application of the reweighting the ratio is now flat.

CHAPTER 7

Removing the CMS thumbprint

The CMS detector naturally smears the kinematic properties of the particles it detects, the effects of which are known as the response of the detector. When comparing to measurements from other experiments or to simulations without the modelling of the detector, the detector response, derived from simulation, needs to be removed. This process is known as unfolding. There are two common approaches to the level of unfolding. Firstly, one can unfold to parton level which extrapolates the results with respect to the constituent partons. This has large theoretical uncertainties stemming from the modelling of the parton shower. In addition, these measurements are usually also extrapolated to the full kinematic phase space. The second option is to present the results to particle level, i.e. with respect to the stable particles produced after the shower modelling where stable particles are for this purpose defined as particles with a lifetime longer than 30 ps.

7.1 Particle level measurement

This thesis presents measurements at particle level in a phase space chosen to closely resemble that used to select events in data. Particle-level objects are used to define the phase space which is identical for both the e+jets and μ+jets channels which allows for a consistent combination after unfolding. Particle-level objects used to define the phase space are constructed from stable particles produced in simulation by the event generator but before the detector interactions are modelled.

The generator-level descriptions of the particles are based on the RIVET framework [127] following prescriptions given in [128]. Simulated electrons and muons not originating from a hadron or a quark are used to define electrons and muons at particle level. Photons close to the lepton are assumed to have radiated from the lepton, and are clustered with it using the anti-k_{T} algorithm with $\Delta R = 0.1$. Particle-level jets are constructed by clustering all stable particles, excluding those used in defining the leptons, with the anti-k_{T} algorithm with $\Delta R = 0.4$. If a particle-level jet originated from a b quark, b hadrons are included in the clustering of jets but with the magnitude of the four momenta scaled to negligible values. They are known as *ghost* particles, because they do not affect the kinematic properties of the jet. A jet with a constituent b hadron is considered to have originated from a b quark. The particle-level $p_{\mathrm{T}}^{\mathrm{miss}}$ is calculated from all stable visible particles.

The visible phase space is given by a single particle-level electron or muon with $p_{\mathrm{T}}^{\ell} > 26$ GeV and $|\eta^{\ell}| < 2.4$. Events with additional leptons of $p_{\mathrm{T}}^{\ell} > 15$ GeV and $|\eta^{\ell}| < 2.4$ are not permitted. The particle-level jet selection differs slightly from that in data. Three particle-level jets must have $p_{\mathrm{T}} > 30$ GeV and the fourth a relaxed cut of $p_{\mathrm{T}} > 20$ GeV. Two of the particle-level jets need to be tagged as originating from a b quark. The variables H_{T}, S_{T} and N_{jets} are subsequently calculated with respect to all jets with $p_{\mathrm{T}} > 20$ GeV.

As stated in Sec. 6.7.1, the yield of $\mathrm{t\bar{t}}$ events for each bin in data is obtained by subtracting the contribution of each background process. In addition, the contribution of $\mathrm{t\bar{t}}$ events which satisfy the selection criteria, but do not enter the visible phase space at particle level, is estimated from simulation and subtracted from the data. This accounts for $\sim 7\%$ of all $\mathrm{t\bar{t}}$ events and are predominately those in which one of the jets fails the particle-level jet selection, but passes the reconstructed jet selection because of the resolution of the detector. The relaxed particle-level jet selection reduced this fraction from $\sim 20\%$ to create the largest possible data sample. The effect on the total uncertainty from this additional extrapolation is negligible. No selection is applied on the decay channel of the top quarks, so the phase space does not exclusively contain single electron or muon $\mathrm{t\bar{t}}$ events. In particular, there are contributions from events where one top quark decays to a tau lepton and subsequently to an electron or muon, or where both top quarks decay leptonically but one lepton is not within the particle-level acceptance.

7.2 Choice of bins

Events can migrate between bins of the measurement because of the finite resolution of the CMS detector, i.e. the reconstructed variable can have a different value and so can be in a different bin compared to the particle-level variable. The choice of binning can be used to minimise migrations. The binning scheme for each of the event variables, with the exception of the jet multiplicity whose binning is naturally defined, is created by iteratively adding together fine bins of that variable until a set of binning criteria is met.

The primary criteria to reduce the level of migration can be represented by the purity p and stability s of the bin which are defined as

$$p^i = \frac{N^i_{\mathrm{rec\&gen}}}{N^i_{\mathrm{rec}}}$$

and

$$s^i = \frac{N^i_{\mathrm{rec\&gen}}}{N^i_{\mathrm{gen}}},$$

where $N^i_{\mathrm{rec\&gen}}$ is the number of events generated and reconstructed in bin i and $N^i_{\mathrm{rec}}(N^i_{\mathrm{gen}})$ is the number of events reconstructed(generated) in bin i. Purity measures the effect of migrations into bins and stability migrations out of bins. By requiring p^i and $s^i > 0.6$ an acceptable balance between the minimisation of the migration of events between bins and the retention of information is obtained. This purity and stability cuts are reduced to 0.5 for $p_{\mathrm{T}}^{\mathrm{miss}}$, due to its naturally low resolution. Figure 46 shows the purity and stability of bins in the e+jets channel and Fig. 47 in the μ+jets channel.

On top of purity and stability, a few other restrictions are required for the choice of binning. The width of the bins must be at least as wide as the typical resolution of the variable in that bin. The resolution is calculated as the standard deviation of the distribution of the residuals ($|\mathrm{Reco} - \mathrm{Gen}|$). Figure 48 shows an example of the resolution, as calculated from the one-sided residual distribution, compared to the bin width. The left panel shows the e+jets channel and the right panel the μ+jets channel for the H_{T} bin spanning $220 - 275$ GeV. Tables 11 and 12

Figure 46: The purity and stability of the bins in the e+jets channel, measured using the POWHEG+PYTHIA simulation sample.

Figure 47: The purity and stability of the bins in the μ+jets channel, measured using the POWHEG+PYTHIA simulation sample.

Figure 48: The residual distribution for the H_T event variable in the eJets channel on the left and the μ+jets channel on the right. The standard deviation of the half-gaussian, which is the resolution, is given as the vertical red line and compared to the bin width given as the vertical blue line.

show the resolution and bin width comparisons for each bin of each variable in the e+jets and μ+jets channels respectively.

The number of reconstructed $t\bar{t}$ events in each bin as estimated from simulation needs to be at least 500, which corresponds to a maximum statistical uncertainty of $\approx 5\%$. If the final bin contains less than 500 events it is merged with the previous bin. This requirement is to ensure the response matrices used in the unfolding do not introduce bias based on the statistical fluctuations of simulation.

Figure 49 shows an example of the fine bin mapping between the reconstructed distribution and the particle-level distribution for the H_T event variable in the e+jets channel on the left panel and μ+jets channel on the right panel. The best binning scheme calculated is overlaid in red.

The response matrices for the e+jets and μ+jets channels, calculated using the common binning scheme are shown in Figs. 50 and 51 respectively. The binning requirements result in a close-to-diagonal response matrix with an acceptable statistical uncertainty.

Figure 49: The left panel shows the mapping between the reconstructed and particle-level distributions of the H_T event variable in the e+jets channel and the right panel in the μ+jets channel. The chosen common binning scheme passing all criteria is overlaid in red.

Figure 50: The set of response matrices, calculated using the POWHEG+PYTHIA simulation sample and normalised to one for all event variables in the e+jets channel using the common binning scheme.

Figure 51: The set of response matrices, calculated using the POWHEG+PYTHIA simulation sample and normalised to one for all event variables in the μ+jets channel using the common binning scheme.

Table 11: A comparison of the bin width and the resolution in each bin for all event variables in the electon channel.

N_{jets}			H_{T}			S_{T}			$p_{\text{T}}^{\text{miss}}$			p_{T}^{W}			p_{T}^{ℓ}			$\lvert\eta^{\ell}\rvert$		
Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$	Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$	Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$	Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$	Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$	Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$	Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$
1.0	1.0	1.0	110.0	26.4	0.2	179.0	33.4	0.2	50.0	17.1	0.3	50.0	21.0	0.4	14.0	0.9	< 0.1	0.30	< 0.1	< 0.1
1.0	1.0	1.0	55.0	27.1	0.5	75.0	34.2	0.5	55.0	28.5	0.5	55.0	22.6	0.4	15.0	1.1	< 0.1	0.30	< 0.1	< 0.1
1.0	1.0	1.0	65.0	29.4	0.5	85.0	38.4	0.5	70.0	32.0	0.5	60.0	26.4	0.4	15.0	1.3	< 0.1	0.30	< 0.1	< 0.1
1.0	1.0	1.0	70.0	31.3	0.4	90.0	41.9	0.5	70.0	32.5	0.5	75.0	30.2	0.4	15.0	1.4	< 0.1	0.30	< 0.1	< 0.1
1.0	1.0	1.0	75.0	33.8	0.5	100.0	45.2	0.5	70.0	33.8	0.5	85.0	33.6	0.4	15.0	1.6	0.1	0.30	< 0.1	< 0.1
2.0	1.0	0.5	85.0	36.0	0.4	105.0	48.8	0.5	250.0	34.3	0.1	90.0	36.6	0.4	15.0	1.8	0.1	0.30	< 0.1	< 0.1
—	—	—	90.0	38.5	0.4	115.0	52.1	0.5	—	—	—	430.0	39.7	< 0.1	15.0	1.9	0.1	0.20	< 0.1	< 0.1
—	—	—	100.0	41.1	0.4	125.0	56.2	0.4	—	—	—	—	—	—	15.0	2.1	0.1	0.40	< 0.1	< 0.1
—	—	—	110.0	43.7	0.4	130.0	60.0	0.5	—	—	—	—	—	—	15.0	2.4	0.2	—	—	—
—	—	—	120.0	47.0	0.4	145.0	63.6	0.4	—	—	—	—	—	—	15.0	2.5	0.2	—	—	—
—	—	—	125.0	50.0	0.4	155.0	68.9	0.4	—	—	—	—	—	—	15.0	2.7	0.2	—	—	—
—	—	—	135.0	52.8	0.4	175.0	73.2	0.4	—	—	—	—	—	—	15.0	2.9	0.2	—	—	—
—	—	—	675.0	58.0	< 0.1	875.0	80.7	< 0.1	—	—	—	—	—	—	15.0	3.1	0.2	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15.0	3.3	0.2	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	20.0	3.5	0.2	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	30.0	3.8	0.1	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	150.0	4.6	< 0.1	—	—	—

Table 12: A comparison of the bin width and the resolution in each bin for all event variables in the muon channel.

N_{jets}			H_{T}			S_{T}			$p_{\text{T}}^{\text{miss}}$			p_{T}^{W}			p_{T}^{ℓ}			$\lvert\eta^{\ell}\rvert$		
Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$	Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$	Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$	Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$	Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$	Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$	Bin width	Resolution	$\frac{\text{Resolution}}{\text{Bin width}}$
1.0	1.0	1.0	110.0	26.3	0.2	179.0	33.6	0.2	50.0	17.3	0.3	50.0	19.8	0.4	14.0	0.5	< 0.1	0.30	< 0.1	< 0.1
1.0	1.0	1.0	55.0	26.9	0.5	75.0	34.4	0.5	55.0	27.4	0.5	55.0	23.3	0.4	15.0	0.8	< 0.1	0.30	< 0.1	< 0.1
1.0	1.0	1.0	65.0	29.2	0.4	85.0	38.8	0.5	70.0	32.3	0.5	60.0	27.2	0.5	15.0	1.0	< 0.1	0.30	< 0.1	< 0.1
1.0	1.0	1.0	70.0	31.5	0.4	90.0	42.2	0.5	70.0	33.0	0.5	75.0	30.8	0.4	15.0	1.4	< 0.1	0.30	< 0.1	< 0.1
1.0	1.0	1.0	75.0	33.6	0.4	100.0	45.5	0.5	70.0	33.6	0.5	85.0	33.9	0.4	15.0	1.8	0.1	0.30	< 0.1	< 0.1
2.0	1.0	0.5	85.0	35.9	0.4	105.0	48.6	0.5	250.0	34.8	0.1	90.0	36.4	0.4	15.0	2.2	0.1	0.30	< 0.1	< 0.1
—	—	—	90.0	38.2	0.4	115.0	52.4	0.5	—	—	—	430.0	39.7	< 0.1	15.0	2.7	0.2	0.20	< 0.1	< 0.1
—	—	—	100.0	40.9	0.4	125.0	56.1	0.4	—	—	—	—	—	—	15.0	3.3	0.2	0.40	< 0.1	< 0.1
—	—	—	110.0	43.6	0.4	130.0	59.4	0.5	—	—	—	—	—	—	15.0	3.9	0.3	—	—	—
—	—	—	120.0	46.4	0.4	145.0	63.9	0.4	—	—	—	—	—	—	15.0	4.6	0.3	—	—	—
—	—	—	125.0	49.0	0.4	155.0	68.8	0.4	—	—	—	—	—	—	15.0	5.3	0.4	—	—	—
—	—	—	135.0	52.5	0.4	175.0	72.7	0.4	—	—	—	—	—	—	15.0	5.9	0.4	—	—	—
—	—	—	675.0	58.8	< 0.1	875.0	82.0	< 0.1	—	—	—	—	—	—	15.0	6.0	0.4	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15.0	6.5	0.4	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	20.0	7.5	0.4	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	30.0	8.6	0.3	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	150.0	11.3	< 0.1	—	—	—

7.3 Unfolding

The unfolding is performed using the TUnfold algorithm [129]. A schematic of the unfolding process is shown in Fig. 52. It shows the relation between the true distribution $\vec{x}$ and the measured distribution $\vec{y}$. The matrix of probabilities $\mathbf{A}$ describes the migrations from a bin of the true distribution into any of the reconstructed bins. The average expected count of events y differs from the observed event counts due to statistical fluctuations.

Figure 52: A schematic view of the migration effects and statistical fluctuations. Figure taken from [129].

The TUnfold algorithm uses a least squares method of estimating the true distribution $\vec{x}$ from the reconstructed distribution $\vec{y}$. Statistical fluctuations are smoothed using Tikhonov regularisation. The algorithm works to minimise

$$\mathcal{L}(x) = \mathcal{L}_1 + \mathcal{L}_2$$

by finding the stationary point, where

$$\mathcal{L}_1 = (\vec{y} - \mathbf{A}\vec{x})^{\mathrm{T}} \mathbf{V}_{yy}^{-1} (\vec{y} - \mathbf{A}\vec{x})$$

and

$$\mathcal{L}_2 = \tau^2 (\vec{x} - f_b \vec{x}_0)^{\mathrm{T}} (\mathbf{L}^{\mathrm{T}} \mathbf{L}) (\vec{x} - f_b \vec{x}_0).$$

$\mathcal{L}_1$ represents the least squares minimisation and $\mathcal{L}_2$ describes the regularisation. $\mathcal{L}_1$ involves the measured observable distribution $\vec{y}$ composed of m bins and the inverse of its covariance matrix $\mathbf{V}_{yy}^{-1}$. The covariance matrix is diagonal and holds the square of the statistical uncertainty. The response matrix $\mathbf{A}$ is derived from the simulated POWHEG+PYTHIA sample.

$\mathcal{L}_2$ damps fluctuations in $\vec{x}$ stemming from statistical fluctuations in $\vec{y}$. The parameter τ^2 defines the strength of the regularisation. If it is too small then the unfolding often yields results with large fluctuations leading to large negative correlations between neighbouring bins. If it is too large then the result will be biased towards the input model $\vec{x}_0$. The parameter f_b decides whether to use a bias distribution based on a model ($f_b = 1$), suppressing differences between unfolded data and model, or to suppress the differences between the unfolded data and 0 ($f_b = 0$). This thesis sets $f_b = 1$. Regularisation can be done by suppressing the differences between either the size, the derivative or the curvature (second derivative) of the

unfolded distribution. These are given by the choice of $\mathbf{L}$. This thesis regularises by curvature, approximated by $(x_{i+1} - x_i) - (x_i - x_{i-1})$ leading to an $\mathbf{L}$ matrix of order $(m-2) \times m$, where non-zero elements are present in elements $L_{i,i} = 1$, $L_{i,i+1} = -2$, $L_{i,i+2} = 1$.

The regularisation parameter τ^2 is found by minimising the average global correlation coefficient. The components of the global correlation coefficient, ρ_i, are taken from the covariance matrix $\mathbf{V}_{xx}$

$$\rho_i = \sqrt{1 - \frac{1}{(\mathbf{V}_{xx}^{-1})_{ii}(\mathbf{V}_{xx})_{ii}}}.$$

The average global correlation is defined by

$$\sum_i \frac{\rho_i}{n},$$

where n is the number of bins at particle level. Scans of the average global correlation coefficient for a range of regularisation strengths are shown in Figs. 53 and 54, with the best τ highlighted at the minimum.

7.4 Cross checking the unfolding

The unfolding is checked to ensure that negligible bias is introduced or mistreatment of the uncertainties on the reconstructed data occurs. The checks are performed using the particle-level truth information.

7.4.1 Checking the uncertainties

The effect of unfolding on the transformation of statistical uncertainties from the reconstructed data to the unfolded data is checked by the distribution of *pulls*. A pull is defined as the ratio of the difference between number of unfolded events in a bin and the true number of events in a bin to the uncertainty on the number of unfolded events

$$\text{Pull}^i = \frac{|x^i_{\text{truth}} - x^i_{\text{unf}}|}{\sigma^i_{\text{unf}}}.$$

The measured and truth distributions for 5000 pseudo experiments are generated by taking variations from the profile of the POWHEG+PYTHIA response matrix. Each pseudo experiment is unfolded using another varied response matrix and the pull calculated for each bin. An example pull distribution in the e+jets channel for the H_{T} event variable in shown in Fig. 55. The resulting mean and width of the pull distribution are close to zero and one respectively, so the normalisation and statistical uncertainties of the unfolded distributions are treated correctly by the unfolding. Figures 56 and 57, show the means and widths of the pull distributions for each bin of the global event variables, which also show correct treatment of the normalisation and statistical uncertainties.

7.4.2 Checking for bias

The response matrix is model dependent because it is constructed directly from a simulated $\text{t}\bar{\text{t}}$ model. A model which poorly describes the data could introduce a bias in the unfolded distributions. To test the size of any bias that could be introduced, the top p_{T} spectrum in the simulated POWHEG+PYTHIA sample is reweighted up and down to cover any discrepancy between the POWHEG+PYTHIA model and the data. The reweighting is applied according to

$$w(t/\bar{t}) = 1 + (p_T^{t/\bar{t}} \pm 100) \times 0.001.$$

Figure 53: A scan of the average global correlation coefficient with respect to the regularisation parameter τ in the e+jets channel. The optimal τ is shown as a red point at the minimum of the scan.

Figure 54: A scan of the average global correlation coefficient with respect to the regularisation parameter τ in the μ+jets channel. The optimal τ is shown as a red point at the minimum of the scan.

Figure 55: Total pull distribution when combining the 5000 pseudo experiments in each bins for the H_{T} variable in the e+jets channel on the left and the μ+jets channel on the right.

Figure 58 shows an example reweighting in the e+jets channel for the H_{T} event variable and the full sets of reweighted distributions are listed in App. TODO.

Unfolding the reweighted distributions using the original response matrix will reveal any bias that is introduced. The bias is defined as the ratio of the unfolded cross sections to the model cross section. The cross section measurements are discussed in Ch. 9. Figures 59 and 60 show the unfolded model cross sections for the reweighted distributions compared against the true model cross sections. The ratios show the bias introduced by using the POWHEG+PYTHIA model in the response matrix, compared to the total systematic uncertainty shown as the grey band. The systematic uncertainties are described in detail in Ch. 8. Any bias seen is small compared to the total systematic uncertainty for the reweighted samples. Bias measurements when unfolding samples of alternate models $\mathrm{t\bar{t}}$ production are shown in Figs. **??** and **??** of App. **??**.

7.4.3 Calculating the cross sections

The yield of $\mathrm{t\bar{t}}$ events is unfolded separately for the electron and muon channels to give the total number of events at particle level in the visible phase space. These are then combined to give the total $N_{\mathrm{t\bar{t}}}$. Individual uncertainties in the e+jets and μ+jets channels, which are explained in more detail in Ch. 8, are added together in quadrature. The normalised cross section with respect to kinematic event variable X, is defined as

$$\frac{1}{\sigma_{\mathrm{t\bar{t}}}^{\mathrm{vis}}}\frac{\mathrm{d}\sigma_{\mathrm{t\bar{t}}}^{i}}{\mathrm{dX}} = \frac{1}{\sum_j N_{\mathrm{t\bar{t}}}^{j}}\frac{N_{\mathrm{t\bar{t}}}^{i}}{\Delta \mathrm{X}^{i}}, \tag{7.1}$$

where $\sigma_{\mathrm{t\bar{t}}}^{\mathrm{vis}}$ is the total $\mathrm{t\bar{t}}$ production cross section in the visible phase space at particle level, $\sigma_{\mathrm{t\bar{t}}}^{i}$ is the $\mathrm{t\bar{t}}$ production cross section in bin i, $N_{\mathrm{t\bar{t}}}^{i(j)}$ is the number of unfolded $\mathrm{t\bar{t}}$ events in bin $i(j)$ and $\Delta \mathrm{X}^{i}$ is the width of bin i. The absolute cross section can be measured as

$$\frac{\mathrm{d}\sigma_{\mathrm{t\bar{t}}}^{i}}{\mathrm{dX}} = \frac{N_{\mathrm{t\bar{t}}}^{i}}{35.9\ \mathrm{fb}^{-1}\Delta \mathrm{X}^{i}}, \tag{7.2}$$

where 35.9 fb^{-1} is the integrated luminosity of the data.

Figure 56: The pull mean and widths in relation to the bin numbers of the event variables in the e+jets channel. The 5000 pseudo experiments are generated from the POWHEG+PYTHIA response matrix.

Figure 57: The pull mean and widths in relation to the bin numbers of the event variables in the μ+jets channel. The 5000 pseudo experiments are generated from the POWHEG+PYTHIA response matrix.

Figure 58: The H_T event distribution of the POWHEG+PYTHIA sample with the top quark p_T reweighted up and down to cover differences to data in the e+jets channel. The distributions are normalised to one.

Figure 59: The cross sections for the reweighted models unfolded using the POWHEG+PYTHIA derived response matrix compared to the true model cross sections are shown for all event variables in the e+jets channel in the upper panels. The lower panels give the ratio of the two cross sections known as the bias.

Figure 60: The cross sections for the reweighted models unfolded using the POWHEG+PYTHIA derived response matrix compared to the true model cross sections are shown for all event variables in the μ+jets channel in the upper panels. The lower panels give the ratio of the two cross sections known as the bias.

CHAPTER 8

Uncertainty

When performing the cross section measurements, there are many factors which can affect the precision of the measurement. These effects are collectively known as *uncertainties*. They are subdivided into statistical and systematic uncertainties, where the statistical uncertainty is reducible by collecting more data and the systematic uncertainty by improvements in theoretical knowledge. Many searches for rare SM processes are dominated by the statistical uncertainty, however due to the multitude of $t\bar{t}$ pairs produced the differential cross sections measured in this thesis are dominated by the systematic uncertainties.

8.1 Statistical uncertainty

The statistical uncertainties measured in the bins of each event variable in each channel are less than 5.8%, however for most variables these values are completely negligible at $< 0.5\%$. The largest statistical uncertainty can be found in the final bins of the $p_{\mathrm{T}}^{\mathrm{miss}}$ variable at 5.8%. The statistical uncertainty from the finite size of the simulated POWHEG+PYTHIA sample used in the construction of the response matrix for unfolding is found to be negligible with no contribution larger than 2.2%.

8.2 Systematic uncertainties

The systematic uncertainties are evaluated and propagated to the final result by recalculating the response matrix using a modified $t\bar{t}$ simulation and/or by modifying the background predictions. The resulting systematic uncertainties are then symmetrised according to the average of the upper and lower variations. The systematic uncertainties can be split into three discrete categories. The first are the experimental uncertainties, which represent our limited understanding on the detector performance. The second and third categories are theoretical uncertainties. One is devoted to the modelling of the parton shower and the other to any remaining theoretical uncertainty.

8.2.1 Experimental uncertainties

The uncertainty in the integrated luminosity of the data was measured to be ±2.5% [130]. The uncertainty in the number of additional inelastic interactions in the same or nearby bunch crossings is measured by varying the total inelastic cross section which is used in the calculation of the pileup distribution as seen in Sec. 6.3, by its uncertainty of ±4.6% [131].

The uncertainty of the b quark jet identification efficiency and mistagging rate in the simulation is taken from the uncertainty in the correction factors dependent on p_{T}, η and quark flavour [120]. The contribution from light jets is calculated independently to that from the b/c quark jets. The upper variations and lower variations are then added in quadrature to give an overall uncertainty before being symmetrised.

The uncertainty in the lepton correction factors is estimated from the simultaneous variation of the individual uncertainties in the lepton identification, reconstruction and trigger efficiencies.

The uncertainty in the JES and JER are estimated as functions of jet p_{T} and η [121] and are propagated to calculation of the $p_{\mathrm{T}}^{\mathrm{miss}}$ variable. Uncertainties on the p_{T} of electron, muon, tau lepton and unclustered PF candidates are also propagated to variables dependent on $p_{\mathrm{T}}^{\mathrm{miss}}$. These uncertainties are found to be negligible.

8.2.2 Parton shower uncertainties

The systematic uncertainties in the modelling of the parton shower of the $\mathrm{t\bar{t}}$ simulation are given by variations in a complete set of associated parameters incorporated in the POWHEG+PYTHIA simulation.

The uncertainty from the parton shower scale used for the ISR modelling is estimated by varying the scale up and down by a factor of two. Similarly the scale used for the FSR modelling is varied up and down by a factor of $\sqrt{2}$. The variation is reduced to $\sqrt{2}$ by constraints measured from the LEP collider [132]. As well as the uncertainties in the parton shower scale, uncertainties in the matrix-element shower scale are also estimated. This is done by varying the factorisation and renormalisation scales independently by a factor of two up and down. Additionally, the scales are varied simultaneously by the same factors. The shower scales uncertainty is defined as the envelope of matrix-element and parton shower scale uncertainties. The dominant component in the shower scales uncertainty originates from the final-state radiation modelling.

The uncertainty in the matching between the matrix-element and parton shower is estimated by varying the h_{damp} parameter by its uncertainty. The h_{damp} parameter is given by $1.58^{+0.66}_{-0.59} \times m_{\mathrm{t}}$ [61]. The parameters controlling the underlying event are varied to estimate the uncertainty in this source.

The uncertainty in the transfer of momentum from b quarks to B hadrons is estimated by varying the tuned parameter $x_b = p_{\mathrm{T}}^{\mathrm{B}}/p_{\mathrm{T}}^{\mathrm{b\,jet}}$ for each tagged particle level b jet up and down by its uncertainty. The variables $p_{\mathrm{T}}^{\mathrm{B}}$ and $p_{\mathrm{T}}^{\mathrm{b\,jet}}$ represent the transverse momenta of the B hadron and the particle-level b jet respectively. The uncertainty is denoted as the fragmentation uncertainty. An additional fragmentation uncertainty is included by using the difference to an alternate model (Peterson fragmentation model [133]). As well as the b quark fragmentation, the b jet energy response is sensitive to the single-lepton branching fractions of B hadrons. The systematic uncertainty introduced by the choice of branching fractions used in the POWHEG+PYTHIA model is estimated by reweighting the fractions to those reported in [8].

The final set of theoretical uncertainties in the parton shower relate to the modelling of the colour reconnection. It is estimated by comparing the cross sections produced when including and excluding the colour reconnection on the top quark decay products (Early resonance decays). Two additional uncertainties are included by using the two different colour reconnection models discussed in Sec. 3.4.1, which are the QCD-based and Gluon move models.

8.2.3 Other theoretical uncertainties

The uncertainties in the cross sections for the single top quark and vector boson backgrounds are taken as ±30% and ±50% respectively. These are based on measurements performed in [134–136] and take into account an extrapolation to the phase space used in this thesis. The uncertainty is typically negligible. The uncertainty in the shape and normalisation of the multijet QCD is taken by predicting the QCD in an alternate control region, leading to uncertainties of up to ±30% and ±60% respectively in any one bin. These typically have a negligible effect on the final results, except at high $|\eta^{\ell}|$.

The uncertainty in the NNPDF30_nlo_as_0118 PDF set used in the POWHEG+PYTHIA simulation is estimated by considering 100 independent replicas. The RMS of the uncertainties, taken as the variations produced from each replica, is defined as the PDF uncertainty. The uncertainty from the choice of $\alpha_S = 0.118$ used in the NNPDF30_nlo_as_0118 set is estimated by varying α_S by ±0.001 and combined in quadrature to the PDF uncertainty.

The uncertainty introduced by the top quark mass used in the $\mathrm{t\bar{t}}$ simulation is estimated by variations from two independent samples simulated with $m_{\mathrm{t}} = 169.5$ GeV and 175.5 GeV. The variations are scaled down to be comparable with the top quark mass uncertainty of ±1 GeV [8].

The uncertainty originating from the mismodelling of the top quark p_{T} spectrum is estimated by reweighting the p_{T} distribution to that measured by [66,67]. The yield of $\mathrm{t\bar{t}}$ events is modified by up to 20% in any one bin and results in a negligible uncertainty.

8.2.4 Reporting the uncertainties

Figures 61 and 62 portray the relative systematic uncertainties in every bin for each variable for the normalised cross section measurements. Similarly, Figs. 63 and 64 show the uncertainty compositions for the absolute cross section measurements. The gold band indicates the total systematic uncertainty and the grey band the total statistical uncertainty. Shown in bold are the systematic uncertainties which are the leading or subleading contributions to any bin.

In the systematic uncertainties for the normalised cross section measurements, the dominant uncertainty for the hadronic based variables is the JES. For the $|\eta^{\ell}|$ variable the normalisation of the background QCD originating from the e+jets channel dominates at high pseudorapidity, however the total uncertainty is small typically at 0.5%. There is no specific systematic uncertainty which dominates in the p_{T}^{ℓ} measurement.

For the absolute cross section measurements the JES is still among the dominant uncertainties present, however more significant contributions are present from the parton shower uncertainties which are fundamentally reduced in the normalised cross section measurements. In particular, the uncertainty from the matrix-element and parton shower scales is large. The b quark tagging efficiency is large at approximately 3% in all bins.

A complete breakdown showing the smallest and largest contributions over all bins of each systematic uncertainty for all cross section measurements are shown in Tabs. 13 and 14. Similar values are shown for the total statistical uncertainty, total systematic uncertainty and total combined uncertainty. Sources of uncertainty that only affect $p_{\mathrm{T}}^{\mathrm{miss}}$ are indicated by — for other variables. The typical uncertainty for the normalised cross section measurements is below 5% but can be as large as 18%. For the measurements of the absolute cross section these are typically 10% but can be as large as 21% in the tails of the distributions.

Figure 61: The composition of the systematic uncertainties for the N_{jets}, H_{T} and S_{T} event variables. Dominant uncertainties are shown in bold. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty.

Figure 62: The composition of the systematic uncertainties for the p_T^{miss}, p_T^{W}, p_T^{ℓ} and $|\eta^{\ell}|$ event variables. Dominant uncertainties are shown in bold. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty.

Figure 63: The composition of the systematic uncertainties for the N_{jets}, H_{T} and S_{T} event variables. Dominant uncertainties are shown in bold. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty.

Figure 64: The composition of the systematic uncertainties for the p_T^{miss}, p_T^W, p_T^ℓ and $|\eta^\ell|$ event variables. Dominant uncertainties are shown in bold. The grey band represents the total statistical uncertainty and the gold band the total systematic uncertainty.

Table 13: The upper and lower bounds, in %, from each source of systematic uncertainty in the normalised differential cross section, over all bins of the measurement for each variable. The bounds of the total relative uncertainty are also shown.

Relative uncertainty source (%)	N_{jets}	H_{T}	S_{T}	$p_{\text{T}}^{\text{miss}}$	p_{T}^{W}	p_{T}^{ℓ}	$\lvert\eta^{\ell}\rvert$
b tagging efficiency	0.1 – 0.8	0.2 – 1.1	0.2 – 1.5	0.1 – 1.2	0.1 – 1.7	0.1 – 1.9	0.1 – 0.5
Electron efficiency	0.1 – 0.2	0.1 – 0.7	0.1 – 0.9	0.1 – 0.7	0.1 – 1.4	0.3 – 2.1	0.1 – 0.8
Muon efficiency	0.1 – 0.3	0.1 – 0.2	0.1 – 0.3	0.1 – 0.2	0.1 – 0.6	0.1 – 1.0	0.1 – 0.1
JER	0.1 – 0.6	0.1 – 0.7	0.2 – 1.8	0.6 – 5.8	0.2 – 2.1	0.1 – 0.2	<0.1
JES	0.1 – 5.5	2.1 – 13.6	2.1 – 15.9	2.1 – 7.1	0.5 – 4.9	0.1 – 2.0	0.1 – 0.2
Electron transverse momentum in $p_{\text{T}}^{\text{miss}}$	—	—	0.1 – 0.3	0.1 – 0.9	0.1 – 0.6	—	—
Muon transverse momentum in $p_{\text{T}}^{\text{miss}}$	—	—	0.1 – 0.9	0.1 – 3.5	0.1 – 0.9	—	—
Tau transverse momentum in $p_{\text{T}}^{\text{miss}}$	—	—	0.1 – 1.4	0.1 – 1.2	0.1 – 1.4	—	—
Unclustered transverse momentum in $p_{\text{T}}^{\text{miss}}$	—	—	0.1 – 1.7	0.2 – 1.9	0.1 – 1.0	—	—
Electron QCD bkg cross section	0.1 – 0.5	0.1 – 0.9	0.1 – 1.5	0.2 – 0.6	0.1 – 0.8	0.1 – 4.5	0.2 – 2.9
Muon QCD bkg cross section	0.1 – 0.2	0.1 – 0.5	0.1 – 0.8	0.1 – 0.2	0.1 – 0.2	0.1 – 0.2	0.1 – 0.1
Electron QCD bkg shape	<0.1	0.1 – 0.4	0.1 – 1.0	0.1 – 0.1	0.1 – 1.5	0.1 – 4.7	0.1 – 1.5
Muon QCD bkg shape	0.1 – 0.1	0.1 – 0.7	0.1 – 0.7	<0.1	<0.1	0.1 – 0.1	0.1 – 0.1
Single top quark cross section	0.1 – 0.4	0.1 – 2.1	0.1 – 4.4	0.1 – 4.9	0.1 – 7.1	0.1 – 6.0	<0.1
V+jets cross section	0.1 – 0.3	0.1 – 2.5	0.1 – 3.7	0.1 – 2.0	0.1 – 3.6	0.1 – 5.4	0.1 – 1.5
PDF	0.1 – 0.3	0.1 – 0.3	0.1 – 0.6	0.1 – 0.3	0.1 – 0.4	0.1 – 0.4	<0.1
Color reconnection (Gluon move)	0.1 – 2.8	0.1 – 4.0	0.1 – 11.7	0.2 – 0.9	0.1 – 1.0	0.2 – 4.8	0.1 – 0.4
Color reconnection (QCD-based)	0.1 – 2.0	0.1 – 4.2	0.1 – 6.6	0.4 – 4.2	0.1 – 3.5	0.1 – 7.6	0.1 – 1.2
Color reconnection (Early resonance decays)	0.2 – 3.9	0.1 – 7.1	0.1 – 4.1	0.1 – 1.6	0.1 – 3.8	0.1 – 5.0	0.1 – 1.0
Fragmentation	0.1 – 0.4	0.1 – 0.5	0.1 – 0.5	0.1 – 0.6	0.1 – 0.6	0.1 – 0.4	<0.1
h_{damp}	0.3 – 3.8	0.1 – 3.1	0.2 – 2.9	0.1 – 2.3	0.1 – 2.7	0.1 – 2.8	0.2 – 1.2
Top quark mass	0.2 – 1.0	0.1 – 3.1	0.2 – 3.5	0.1 – 4.0	0.2 – 1.1	0.2 – 4.5	0.1 – 0.6
Peterson fragmentation model	0.1 – 1.3	0.1 – 0.6	0.1 – 0.9	0.1 – 1.1	0.1 – 1.0	0.1 – 1.3	<0.1
Shower scales	0.4 – 4.3	0.5 – 4.5	0.5 – 4.9	0.2 – 2.4	0.3 – 3.5	0.1 – 4.5	0.1 – 0.7
b hadron decay semileptonic branching fraction	0.1 – 0.1	0.1 – 0.1	0.1 – 0.1	<0.1	<0.1	<0.1	<0.1
Top quark p_{T}	0.1 – 0.7	0.1 – 0.9	0.1 – 1.0	0.1 – 0.8	0.1 – 0.9	0.1 – 1.3	<0.1
Underlying event tune	0.1 – 2.7	0.1 – 5.5	0.2 – 4.4	0.1 – 5.4	0.2 – 2.6	0.2 – 6.1	0.1 – 0.9
Simulated sample size	0.1 – 1.6	0.1 – 1.6	0.1 – 1.9	0.1 – 2.2	0.1 – 1.4	0.1 – 1.7	0.1 – 0.4
Additional interactions	0.1 – 0.4	0.1 – 1.0	0.1 – 1.7	0.1 – 1.5	0.1 – 0.9	0.1 – 1.0	<0.1
Integrated luminosity	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Total systematic uncertainty	0.6 – 9.6	2.7 – 14.1	2.8 – 17.4	2.9 – 11.7	0.8 – 12.6	0.7 – 13.4	0.7 – 4.4
Total statistical uncertainty	0.3 – 3.9	0.3 – 4.5	0.3 – 5.3	0.2 – 5.8	0.2 – 4.1	0.3 – 4.7	0.3 – 1.0
Total uncertainty	0.7 – 10.0	2.8 – 14.1	2.8 – 18.2	3.0 – 13.0	0.8 – 13.2	0.8 – 14.2	0.8 – 4.5

Table 14: The upper and lower bounds, in %, from each source of systematic uncertainty in the absolute differential cross section, over all bins of the measurement for each variable. The bounds of the total relative uncertainty are also shown.

Relative uncertainty source (%)	N_{jets}	H_{T}	S_{T}	$p_{\text{T}}^{\text{miss}}$	p_{T}^{W}	p_{T}^{ℓ}	$\lvert\eta^{\ell}\rvert$
b tagging efficiency	3.1 – 4.0	3.5 – 4.5	3.5 – 5.0	3.6 – 4.8	3.6 – 5.2	3.6 – 5.4	3.6 – 4.2
Electron efficiency	1.6 – 1.9	1.6 – 2.3	1.4 – 2.5	1.6 – 2.4	1.2 – 3.0	1.0 – 3.5	1.0 – 2.1
Muon efficiency	2.0 – 2.5	2.2 – 2.5	2.2 – 2.6	2.3 – 2.6	2.3 – 2.9	2.2 – 3.3	2.1 – 2.3
JER	0.1 – 0.8	0.1 – 1.1	0.2 – 2.3	0.4 – 5.6	0.4 – 1.6	0.1 – 0.3	0.3 – 0.4
JES	1.4 – 9.2	4.3 – 12.8	4.3 – 15.1	2.2 – 10.9	2.2 – 7.8	1.8 – 4.1	3.8 – 4.1
Electron transverse momentum in $p_{\text{T}}^{\text{miss}}$	—	—	0.1 – 0.3	0.1 – 0.9	0.1 – 0.6	—	—
Muon transverse momentum in $p_{\text{T}}^{\text{miss}}$	—	—	0.1 – 0.9	0.1 – 3.5	0.1 – 0.8	—	—
Tau transverse momentum in $p_{\text{T}}^{\text{miss}}$	—	—	0.1 – 1.4	0.1 – 1.2	0.1 – 1.4	—	—
Unclustered transverse momentum in $p_{\text{T}}^{\text{miss}}$	—	—	0.1 – 1.7	0.2 – 1.9	0.1 – 1.0	—	—
Electron QCD bkg cross section	0.1 – 0.8	0.2 – 1.4	0.2 – 2.1	0.1 – 0.8	0.1 – 1.3	0.3 – 5.0	0.1 – 3.6
Muon QCD bkg cross section	0.1 – 0.3	0.1 – 0.7	0.1 – 1.1	0.1 – 0.3	0.1 – 0.4	0.1 – 0.4	0.1 – 0.3
Electron QCD bkg shape	<0.1	0.1 – 0.3	0.1 – 1.0	0.1 – 0.1	0.1 – 1.6	0.1 – 4.8	0.1 – 1.5
Muon QCD bkg shape	0.1 – 0.1	0.1 – 0.7	0.1 – 0.8	<0.1	<0.1	0.1 – 0.1	0.1 – 0.1
Single top quark cross section	1.1 – 1.7	1.1 – 3.5	1.1 – 5.8	1.3 – 6.3	1.1 – 8.4	1.3 – 7.4	1.4 – 1.5
V+jets cross section	0.7 – 1.1	0.6 – 3.4	0.5 – 4.6	0.7 – 2.9	0.7 – 4.5	0.6 – 6.3	0.6 – 2.5
PDF	0.1 – 0.3	0.1 – 0.4	0.1 – 0.6	0.1 – 0.3	0.1 – 0.4	0.1 – 0.5	<0.1
Color reconnection (Gluon move)	0.2 – 2.8	0.2 – 4.1	0.1 – 11.8	0.2 – 1.0	0.1 – 1.1	0.2 – 4.7	0.1 – 0.5
Color reconnection (QCD-based)	0.2 – 2.1	0.1 – 4.3	0.1 – 6.7	0.3 – 4.6	0.3 – 3.8	0.1 – 7.9	0.1 – 1.7
Color reconnection (Early resonance decays)	0.1 – 3.9	0.1 – 7.1	0.1 – 4.1	0.1 – 1.4	0.1 – 3.8	0.1 – 5.0	0.1 – 1.1
Fragmentation	0.1 – 0.7	0.6 – 1.6	0.5 – 1.4	0.1 – 0.7	0.2 – 0.8	0.1 – 0.8	0.3 – 0.4
h_{damp}	0.1 – 3.5	0.2 – 3.1	0.1 – 3.0	0.6 – 2.1	0.2 – 3.2	0.4 – 3.4	0.5 – 2.0
Top quark mass	0.7 – 2.0	0.3 – 3.4	0.3 – 3.7	0.3 – 5.0	0.4 – 2.0	0.3 – 3.5	0.9 – 1.7
Peterson fragmentation model	0.3 – 2.0	1.6 – 2.7	1.9 – 3.2	1.1 – 2.6	1.3 – 2.6	1.2 – 2.9	1.5 – 1.5
Shower scales	2.7 – 6.5	2.6 – 6.4	3.0 – 7.0	4.0 – 6.6	4.6 – 6.2	3.9 – 6.2	4.9 – 5.6
b hadron decay semileptonic branching fraction	0.2 – 0.3	0.1 – 0.3	0.1 – 0.3	0.2 – 0.3	0.2 – 0.3	0.2 – 0.3	0.2 – 0.3
Top quark p_{T}	0.4 – 1.2	0.1 – 0.8	0.1 – 0.9	0.1 – 1.5	0.1 – 1.0	0.1 – 1.0	0.6 – 0.7
Underlying event tune	0.1 – 2.9	0.2 – 5.3	0.2 – 4.4	0.2 – 5.4	0.1 – 2.6	0.1 – 6.0	0.2 – 0.9
Simulated sample size	0.1 – 1.6	0.1 – 1.6	0.1 – 1.9	0.1 – 2.2	0.1 – 1.4	0.1 – 1.7	0.1 – 0.4
Additional interactions	0.1 – 0.3	0.1 – 0.7	0.1 – 1.3	0.3 – 1.3	0.1 – 0.6	0.1 – 0.8	0.1 – 0.3
Integrated luminosity	2.5 – 2.5	2.5 – 2.5	2.5 – 2.5	2.5 – 2.5	2.5 – 2.5	2.5 – 2.5	2.5 – 2.5
Total systematic uncertainty	8.7 – 13.4	9.7 – 15.9	9.5 – 20.0	8.8 – 17.1	8.3 – 17.5	8.6 – 16.1	8.8 – 10.6
Total statistical uncertainty	0.3 – 3.9	0.4 – 4.5	0.3 – 5.3	0.2 – 5.8	0.2 – 4.1	0.3 – 4.7	0.3 – 1.1
Total uncertainty	8.7 – 13.7	9.7 – 16.2	9.5 – 20.7	8.8 – 18.1	8.3 – 18.0	8.6 – 16.7	8.8 – 10.6

CHAPTER 9

Results

As stated previously, measurements of the normalised and absolute cross sections are performed to particle level in a visible phase space. Normalised cross section measurements have greater precision due to the cancellation of some systematics, whereas the absolute measurements contain all information available. They are presented to particle level in order to be useful in comparisons to future models and results from other experiments. It allows models to be compared without wasting valuable computational resources and time on the simulation of the detector and avoids the influence of large theoretical uncertainties introduced extrapolating the results to parton level (or a full phase space). These measurements are presented here along with the results of χ^2 goodness-of-fit tests. Additional studies with respect to the theoretical uncertainties in the $t\bar{t}$ model and the effects of non-regularised unfolding are also shown.

9.1 Cross section measurements

Figure 65 shows the normalised differential cross sections with respect to the N_{jets}, H_{T} and S_{T} event variables and Fig. 66 with respect to the $p_{\text{T}}^{\text{miss}}$, p_{T}^{W}, p_{T}^{ℓ} and $|\eta^{\ell}|$ variables respectively. Figures 67 and 68 show the relevant plots for the absolute cross section measurements. The cross section measurements are compared to four combinations of the matrix-element and parton shower: POWHEG+PYTHIA, POWHEG+HERWIG++, MG5_aMC@NLO-NLO+PYTHIA and MG5_aMC@NLO-LO+PYTHIA. In the lower panels, the ratio of the model cross section to the measured cross section is shown overlaid onto the grey statistical uncertainty band and the gold systematic band. By eye, good agreement can already be seen for the N_{jets} variable between the measured cross sections and the POWHEG+PYTHIA and MG5_aMC@NLO-NLO+PYTHIA models. The trend in the modelling of the p_{T} of the top quark can easily be seen in all the hadronically based variables, which are is softer in data than simulation. Figures 69, 70, 71 and 72 show the normalised and absolute cross section measurements compared to three variations of the POWHEG+PYTHIA simulation. These are variations of the shower scales, variations of the h_{damp} matching parameter and a variation with the application of reweighting to the quark top p_{T}. In the latter case, the slope seen in the ratio is much less pronounced as expected.

While useful information can be inferred from these plots, absolute statements cannot be made regarding how well a simulation models the data. This is due to correlations of the

uncertainties between bins. A χ^2 goodness-of-fit test needs to be performed between each model and the data.

9.2 The χ^2 goodness-of-fit tests

The χ^2 goodness-of-fit test is defined by

$$\chi^2 = u_i.\mathbf{V}_{ij}^{-1}.u_j^T,$$

where u is a vector of the differences between the measured cross section and the model per bin $u_i = (x_i^{\text{meas}} - x_i^{\text{model}})$ and u^T is the transpose of this vector. The correlations between the bins are taken into account by the full covariance matrix $\mathbf{V}$. The full covariance matrix is created by adding together all the covariance matrices for all statistical and systematic uncertainties. Each measurement has a number of degrees of freedom equal to its number of bins. This is reduced by one for the normalised cross section measurements because of the additional constraint from the normalisation. A p-value can be obtained for each χ^2/ndf. A second χ^2 test is performed between the measured cross sections and the POWHEG+PYTHIA model where the theoretical uncertainties in the model as well as in the unfolded data are taken into account. The correlations between the uncertainties in the prediction of the model and the measured cross section are included. All χ^2/ndf and p-values are shown per $t\bar{t}$ model in Tabs. 15 and 16 for the normalised and absolute cross section measurements respectively.

The predictions of the S_{T}, $p_{\text{T}}^{\text{miss}}$ and p_{T}^{ℓ} kinematic event distributions by the POWHEG+PYTHIA model are consistent with the measured cross sections. The N_{jets} variable is well described by the POWHEG+PYTHIA generator with χ^2/ndf of 2.0/5 and 2.2/6 for the normalised and absolute cross section measurements respectively. The tune used in the POWHEG+PYTHIA model is derived using jet multiplicity measurements from 8 TeV, as seen in Fig. 9 and the good description measured here confirms the tune describes well the data for a larger data set at higher $\sqrt{s}$. Tensions are seen in the S_{T}, p_{T}^{W} and $|\eta^{\ell}|$ distributions. The results of the χ^2 tests with the inclusion of the theoretical uncertainties in the POWHEG+PYTHIA model show that any tensions observed are covered.

The POWHEG+HERWIG++ model is broadly consistent with the cross section measurements, particularly for the $p_{\text{T}}^{\text{miss}}$, p_{T}^{ℓ} and p_{T}^{W} event variables, however large tensions are seen for the N_{jets} distributions. Tension is also seen for the $|\eta^{\ell}|$ variable. The MG5_aMC@NLO-NLO+PYTHIA model is consistent with all the kinematic event variables measured, with the exception of $|\eta^{\ell}|$. If the theoretical uncertainties are applied it is expected that it would cover some of the tensions seen in both the POWHEG+HERWIG++ and MG5_aMC@NLO-NLO+PYTHIA models. Without considering the theoretical uncertainties in the MG5_aMC@NLO-LO+PYTHIA model, there is no consistency with measured data for any variable.

No model is consistent for all the cross section measurements without taking the theoretical uncertainties in the model into account. The NLO models are more consistent with the data than the LO model. When taking into account model theoretical uncertainties the tensions are reduced. The $|\eta^{\ell}|$ event variable causes tensions in all models studied.

9.3 Effects of regularisation

A study on the effects of the regularisation in the unfolding on the χ^2/ndf calculation is performed. A lack of regularisation can lead to non-physical effects present in the unfolded distributions, however too much regularisation and the unfolded data is biased towards the input simulation.

Tables 17 and 18 show the χ^2/ndf produced when regularisation is not used. Typically only small changes are seen between the regularised and unregularised measurements. The largest changes are, as expected, linked to event variables which have a low experimental resolution. One of the largest changes in the χ^2/ndf is seen in the $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution of the POWHEG+PYTHIA simulation with model theoretical uncertainties where it decreases from 2.9/5 to 2.1/5. Another is for the H_{T} distribution in the MG5_aMC@NLO-NLO+PYTHIA model where it increases from 11/12 to 12/12. The χ^2/ndf does not change significantly for most other event variables and not at all for p_{T}^{ℓ} and $|\eta^{\ell}|$. This implies that regularisation is not needed in the case of variables with a high experimental resolution.

Figure 65: The normalised cross section measurements for the N_{jets}, H_{T} and S_{T} kinematic event variables. They are compared against different $\mathrm{t\bar{t}}$ simulations. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

Figure 66: The normalised cross section measurements for the p_T^{miss}, p_T^W, p_T^ℓ and $|\eta^\ell|$ kinematic event variables. They are compared against different $t\bar{t}$ simulations. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

Figure 67: The absolute cross section measurements for the N_{jets}, H_T and S_T kinematic event variables. They are compared against different $t\bar{t}$ simulations. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

Figure 68: The absolute cross section measurements for the p_T^{miss}, p_T^{W}, p_T^{ℓ} and $|\eta^{\ell}|$ kinematic event variables. They are compared against different $t\bar{t}$ simulations. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

Figure 69: The normalised cross section measurements for the N_{jets}, H_{T} and S_{T} kinematic event variables. They are compared against the POWHEG+PYTHIA simulation after varying the shower scales, the h_{damp} matching parameter and the application of top quark p_{T} reweighting. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

Figure 70: The normalised cross section measurements for the p_T^{miss}, p_T^W, p_T^ℓ and $|\eta^\ell|$ kinematic event variables. They are compared against the POWHEG+PYTHIA simulation after varying the shower scales, the h_{damp} matching parameter and the application of top quark p_T reweighting. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

Figure 71: The absolute cross section measurements for the N_{jets}, H_{T} and S_{T} kinematic event variables. They are compared against the POWHEG+PYTHIA simulation after varying the shower scales, the h_{damp} matching parameter and the application of top quark p_{T} reweighting. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

Figure 72: The absolute cross section measurements for the p_T^{miss}, p_T^W, p_T^ℓ and $|\eta^\ell|$ kinematic event variables. They are compared against the POWHEG+PYTHIA simulation after varying the shower scales, the h_{damp} matching parameter and the application of top quark p_T reweighting. The verticle bars represent statistical and systematic uncertainties added in quadrature. The ratio of the predictions to the data is shown below.

Table 15: Results of a goodness-of-fit test between the normalised cross sections in data and several models, with values given as χ^2/number of degrees of freedom (ndf).

	POWHEG+PYTHIA		With MC theoretical uncertainties	
	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	2.0 / 5	0.85	1.5 / 5	0.91
H_{T}	26 / 12	< 0.01	4.8 / 12	0.97
S_{T}	22 / 12	0.04	4.2 / 12	0.98
$p_{\text{T}}^{\text{miss}}$	11 / 5	0.06	2.9 / 5	0.72
p_{T}^{W}	16 / 6	0.01	2.5 / 6	0.87
p_{T}^{ℓ}	24 / 16	0.09	14 / 16	0.63
$\lvert\eta^{\ell}\rvert$	19 / 7	< 0.01	15 / 7	0.04

	POWHEG+HERWIG++		MG5_aMC@NLO-NLO+PYTHIA		MG5_aMC@NLO-LO+PYTHIA	
	χ^2/ndf	p-value	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	38 / 5	< 0.01	9.5 / 5	0.09	78 / 5	< 0.01
H_{T}	23 / 12	0.03	11 / 12	0.52	160 / 12	< 0.01
S_{T}	21 / 12	0.04	11 / 12	0.57	110 / 12	< 0.01
$p_{\text{T}}^{\text{miss}}$	1.3 / 5	0.93	5.9 / 5	0.31	23 / 5	< 0.01
p_{T}^{W}	0.81 / 6	0.99	8.9 / 6	0.18	30 / 6	< 0.01
p_{T}^{ℓ}	11 / 16	0.82	16 / 16	0.44	37 / 16	< 0.01
$\lvert\eta^{\ell}\rvert$	19 / 7	< 0.01	24 / 7	< 0.01	30 / 7	< 0.01

Table 16: Results of a goodness-of-fit test between the absolute cross sections in data and several models, with values given as χ^2/number of degrees of freedom (ndf).

	POWHEG+PYTHIA		With MC theoretical uncertainties	
	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	2.2 / 6	0.90	1.7 / 6	0.95
H_{T}	23 / 13	0.05	4.3 / 13	0.99
S_{T}	19 / 13	0.11	4.7 / 13	0.98
$p_{\text{T}}^{\text{miss}}$	13 / 6	0.05	3.1 / 6	0.80
p_{T}^{W}	17 / 7	0.02	2.7 / 7	0.91
p_{T}^{ℓ}	20 / 17	0.28	14 / 17	0.68
$\lvert\eta^{\ell}\rvert$	16 / 8	0.04	15 / 8	0.06

	POWHEG+HERWIG++		MG5_aMC@NLO-NLO+PYTHIA		MG5_aMC@NLO-LO+PYTHIA	
	χ^2/ndf	p-value	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	39 / 6	< 0.01	12 / 6	0.07	93 / 6	< 0.01
H_{T}	21 / 13	0.07	10 / 13	0.66	150 / 13	< 0.01
S_{T}	18 / 13	0.17	9.3 / 13	0.75	110 / 13	< 0.01
$p_{\text{T}}^{\text{miss}}$	1.5 / 6	0.96	6.6 / 6	0.36	26 / 6	< 0.01
p_{T}^{W}	0.90 / 7	1.00	9.2 / 7	0.24	33 / 7	< 0.01
p_{T}^{ℓ}	11 / 17	0.87	15 / 17	0.58	36 / 17	< 0.01
$\lvert\eta^{\ell}\rvert$	17 / 8	0.04	23 / 8	< 0.01	31 / 8	< 0.01

Table 17: Results of a goodness-of-fit test between the unregularised normalised cross sections in data and several models, with values given as χ^2/number of degrees of freedom (ndf).

	POWHEG+PYTHIA		With MC theoretical uncertainties	
	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	1.9 / 5	0.86	1.2 / 5	0.95
H_{T}	25 / 12	0.01	5.6 / 12	0.94
S_{T}	22 / 12	0.04	5.6 / 12	0.94
$p_{\text{T}}^{\text{miss}}$	9.3 / 5	0.10	2.1 / 5	0.84
p_{T}^{W}	16 / 6	0.01	2.7 / 6	0.84
p_{T}^{ℓ}	24 / 16	0.09	13 / 16	0.64
$\lvert\eta^{\ell}\rvert$	19 / 7	< 0.01	15 / 7	0.04

	POWHEG+HERWIG++		MG5_aMC@NLO-NLO+PYTHIA		MG5_aMC@NLO-LO+PYTHIA	
	χ^2/ndf	p-value	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	42 / 5	< 0.01	7.6 / 5	0.18	81 / 5	< 0.01
H_{T}	21 / 12	0.05	12 / 12	0.44	140 / 12	< 0.01
S_{T}	22 / 12	0.04	11 / 12	0.54	98 / 12	< 0.01
$p_{\text{T}}^{\text{miss}}$	0.82 / 5	0.98	5.2 / 5	0.39	21 / 5	< 0.01
p_{T}^{W}	1.7 / 6	0.94	9.4 / 6	0.15	29 / 6	< 0.01
p_{T}^{ℓ}	11 / 16	0.83	16 / 16	0.44	37 / 16	< 0.01
$\lvert\eta^{\ell}\rvert$	19 / 7	< 0.01	24 / 7	< 0.01	30 / 7	< 0.01

Table 18: Results of a goodness-of-fit test between the unregularised absolute cross sections in data and several models, with values given as χ^2/number of degrees of freedom (ndf).

	POWHEG+PYTHIA		With MC theoretical uncertainties	
	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	1.7 / 6	0.94	1.5 / 6	0.96
H_{T}	21 / 13	0.07	4.8 / 13	0.98
S_{T}	18 / 13	0.15	6.0 / 13	0.94
$p_{\text{T}}^{\text{miss}}$	11 / 6	0.10	2.4 / 6	0.88
p_{T}^{W}	16 / 7	0.02	2.9 / 7	0.89
p_{T}^{ℓ}	20 / 17	0.28	14 / 17	0.68
$\lvert\eta^{\ell}\rvert$	16 / 8	0.04	15 / 8	0.06

	POWHEG+HERWIG++		MG5_aMC@NLO-NLO+PYTHIA		MG5_aMC@NLO-LO+PYTHIA	
	χ^2/ndf	p-value	χ^2/ndf	p-value	χ^2/ndf	p-value
N_{jets}	38 / 6	< 0.01	7.2 / 6	0.31	70 / 6	< 0.01
H_{T}	20 / 13	0.09	11 / 13	0.64	140 / 13	< 0.01
S_{T}	19 / 13	0.14	9.8 / 13	0.71	100 / 13	< 0.01
$p_{\text{T}}^{\text{miss}}$	0.86 / 6	0.99	5.7 / 6	0.45	25 / 6	< 0.01
p_{T}^{W}	1.9 / 7	0.96	9.7 / 7	0.21	31 / 7	< 0.01
p_{T}^{ℓ}	10 / 17	0.88	15 / 17	0.58	36 / 17	< 0.01
$\lvert\eta^{\ell}\rvert$	17 / 8	0.04	23 / 8	< 0.01	31 / 8	< 0.01

CHAPTER 10

Conclusions, prospects and outlook

Conclusions HERE

The measurements presented in this analysis have reached what is known as the *systematic wall*. This means without the input of the theory community, improvements... These improvements in the systematic uncertainties are being for example in refTODO

It makes sense then to go beyond the bounds of the current analysis into regions that are no longer dominated by the systematic uncertainty. This could be in the form of double or even triple differential cross sections. The ability to compare a model to two or three different variables simultaneously is an ideal prospect of this analysis for the $\mathrm{t\bar{t}}$ generator modelling and tuning community. Important distributions, for example the $(p_{\mathrm{T}}^{\mathrm{top}}, |\eta^{\mathrm{top}}|)$ and $(p_{\mathrm{T}}^{\mathrm{top}}, M_{\mathrm{t\bar{t}}})$ distributions have already been studied by analyses such as TODO. These allow for comparisons to the full kinematic properties of the top quark as well as information on the discrepancies seen between the top p_{T} spectrum and that of the mass of the $\mathrm{t\bar{t}}$ system. This analysis could provide useful additional distributions with respect to the number of additional jets or missing p_{T} of the system.

Another option is to go beyond the current kinematic event variables used in this thesis. The could include additional variables such as the angular separation between the two b jets with the highest CSV discriminant, the angular separation between the lepton and closest b jet or the p_{T} of additional jets. New and novel event variables, for example the n-jettiness or sphericity of the event TODOREF, could be measured. Of course, all the event variables currently measured can also be extrapolated to the full phase space and parton level, although perhaps more interesting and useful would be to extend them into the boosted regime, i.e. the regime where the top decay products become highly collimated due to being produced at high energy.

Finally, an option is to reinterpret the analysis as a part of an effective field theory. An effective field theory assumes that any new physics is at an energy scale inaccessible to the standard model and as such is a low energy approximation. Any new physics would then be introduced by deviations in the tails of the distributions caused by the interference between the standard model and effective field theory. The RIVET toolkit is a powerful tool in the reinterpretation of analyses.

10.1 Double differential cross sections

Very basic examples for the double differential cross sections are shown in Figs. 73 and 75, which show the distributions with respect to S_{T} and N_{jets}, and p_{T}^{ℓ} and $|\eta^{\ell}|$ respectively. Further distributions for S_{T} and $p_{\mathrm{T}}^{\mathrm{miss}}$, and $p_{\mathrm{T}}^{\mathrm{miss}}$ and N_{jets} are shown in Figs. **??** and **??** in App. **??**. Firstly, the top two panels of each figure show the data-simulation comparison for the e+jets and μ+jets channels. The binning scheme for each variable has been chosen arbitrarily and the yield from multijet QCD measured from simulation. The $\mathrm{t\bar{t}}$ yield is extracted using the background subtraction method, as for the single differential cross section measurements.

The response matrices used in the unfolding of the $\mathrm{t\bar{t}}$ yields are created using the POWHEG+PYTHIA sample and use four times the number of bins at reconstruction level than at generator level, achieved by splitting the binning schemes in two for each variable under study. The central two panels of each figure show the response matrices used in unfolding the e+jets and μ+jets channels. They have been recombined into the bins used at generator level for ease of viewing. The $\mathrm{t\bar{t}}$ yields are unfolded without regularisation and so an unknown amount of bias may be present.

The unfolded $\mathrm{t\bar{t}}$ yields are used to calculate the normalised and absolute cross sections presented in the lower two panels of each figure. A comparison to the POWHEG+PYTHIA $\mathrm{t\bar{t}}$ model is also shown. At no point have any systematic uncertainties been introduced or unfolding checks performed.

10.2 Additional event variables

Some simple comparisons for the data-simulation agreement for new variables can be seen in Fig. TODO. They include the angle of the b quark to the lepton, the

10.3 Effective field theory

The SM is largely assumed to be a valid effective theory up until an energy scale at which new physics is introduced Λ_{NP}. At this point any new field theory describing must satisfy three conditions, namely: The SM gauge group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ should be contained in the new gauge group, all the degrees of freedom of the SM should still be present and at low energies it should reduce into the SM. This reduction is likely believed to originate from the decoupling of the massive new particles at Λ_{NP}.

The new physics can be added to the SM Lagrangian by adding higher-dimensional operators $\mathcal{O}_i$, constructed only out of SM fields which maintains gauge invariance and their associated Wilson coefficients C_i, which describe the strength of the new physics.

$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_{d=5}^{\infty} \frac{1}{\Lambda_{\mathrm{NP}}^{d-4}} \sum_i C_i^d \mathcal{O}_i^d$$

$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \frac{1}{\Lambda_{\mathrm{NP}}} C_1^{(5)} \mathcal{O}_1^{(5)} + \frac{1}{\Lambda_{\mathrm{NP}}^2} \sum_i C_i^{(6)} \mathcal{O}_i^{(6)} + h.c. + \ldots$$

The higher-dimensional terms are suppressed by $\Lambda_{\mathrm{NP}}^{d-4}$, where d is the dimension. The dimension 5 operator only contains one operator after gauge symmetry constraints have been applied which violates lepton number and has no relation to top physics. It is not considered further. At dimension 6 however, there are 16 operators which are relevant for top quark production and decay [137]. These operators can alter the normalisation and/or shape of various distributions measured in relation to the top quark. One particular operator of interest is $\mathcal{O}_{\mathrm{tG}}$ which modifies

Figure 73: Double differential $t\bar{t}$ production cross section measurements with respect to S_T and N_{jets}. The upper two panels show the distributions after full event selection for the e+jets and μ+jets channels respectively. The central two panels show the response matrices in terms of global bin number and the lower two panels show the normalised and absolute cross section measurements in comparison to the POWHEG+PYTHIA model.

Figure 74: Double differential $t\bar{t}$ production cross section measurements with respect to p_T^ℓ and $|\eta^\ell|$. The upper two panels show the distributions after full event selection for the e+jets and μ+jets channels respectively. The central two panels show the response matrices in terms of global bin number and the lower two panels show the normalised and absolute cross section measurements in comparison to the POWHEG+PYTHIA model.

/Users/db0268/Mount/SoolinScratch/DPS/DPSTestingGround/DailyPy

/Users/db0268/Mount/SoolinScratch/DPS/DPSTestingGround/DailyPy

/Users/db0268/Mount/SoolinScratch/DPS/DPSTestingGround/DailyPy

/Users/db0268/Mount/SoolinScratch/DPS/DPSTestingGround/DailyPy

the $t\bar{t}g$ vertex and introduces a new $t\bar{t}gg$ vertex as shown in Fig. 76 and results in a modification to the $t\bar{t}$ production cross section.

Figure 76: The upper two panels show the Feynman diagrams for the modified $t\bar{t}g$ vertices and the lower panel the additional $t\bar{t}gg$ vertex from the inclusion of the $\mathcal{O}_{tG}$ operator. The red dot indicates the effective vertex.

A LO model for the SM and top quark related EFT operators is provided including the $\mathcal{O}_{tG}$ operator and is described in detail in [138]. The parameter C_{tG}/Λ_{NP} is calculated following prescriptions given in [138, 139].

Events are generated using MG5_aMC@NLO (v2.6.2) interfaced with PYTHIA (v8.238), however, it is not reasonable or feasible to generate N different samples depending on N different input values of C_{tG}/Λ_{NP}. Instead, the interference contributions can be scaled to the desired value of C_{tG}/Λ_{NP} using the following method. The $t\bar{t}$ production cross section can be written as a perturbative expansion in terms of C_{tG}/Λ_{NP}

$$\sigma_{t\bar{t}} = \sigma_{t\bar{t}}^{SM} + \frac{C_{tG}}{\Lambda_{NP}} \cdot \beta_1 + \left(\frac{C_{tG}}{\Lambda_{NP}}\right)^2 \cdot \beta_2, \tag{10.1}$$

where the β terms are $\mathcal{O}_{tG}$ contributions to the cross section from linear and quadratic perturbations of C_{tG}/Λ_{NP}. By taking arbitrary values for this perturbation at $\pm X$ this reduces to

$$\beta_1 = \frac{\sigma_{t\bar{t}}(+X) - \sigma_{t\bar{t}}(-X)}{2X}, \tag{10.2}$$

which is the interference contribution for the Feynman diagrams containing one SM vertex and one vertex which can be scaled to the desired C_{tG}/Λ_{NP} and added to the SM prediction ($X = 0$) for the final prediction. The quadratic contributions to the total $t\bar{t}$ production cross section are negligible for $C_{tG}/\Lambda_{NP} < 1$ TeV^{-2} [140]. Thus three simulations are created using $C_{tG}/\Lambda_{NP} = +2,\ 0,\ -2$, the outcomes of which are shown in Fig. TODO, which shows a comparison of all three variations to the unfolded data as seen in this thesis. This is done using the RIVET plugin.

The C_{tG}/Λ_{NP} parameter can then be extracted from a fit of χ^2 goodness-of-fit tests between the unfolded cross sections and predictions for a given a range of C_{tG}/Λ_{NP}. The goodness-of-fit tests proceed in an identical manner as in Sec. 9.2.

Summary

The field of particle physics sets out to answer two of the greatest, most fundamental questions of our age. What is the Universe made of and how does it interact? At its beating heart is the standard model, unparalleled in its success at predicting the existence of and confirming all of the observed particles and their interactions. It is not without issues though, with its inability to explain gravity or to describe more than the baryonic content of the Universe. These and other issues are the lifeblood of particle physicists determined to create a complete picture of our Universe.

The top quark, the most massive particle in the standard model, is expected to be the most sensitive to any new physics beyond the standard model. This new physics will directly manifest itself in the $t\bar{t}$ pair production cross section. For this reason, and others precise measurements of the inclusive and differential standard model top quark cross sections are vital.

This thesis presents measurements of the differential cross section as a function of several kinematic event variables in the single-lepton decay channel. The kinematic event variables are variables that do not require the reconstruction of the complete $t\bar{t}$ system. The event variables considered are the jet multiplicity, N_{jets}, the scalar sum of the jet p_{T}, H_{T}, the scalar sum of the p_{T} of all particles, S_{T}, the magnitudes of the transverse momentum imbalance, $p_{\text{T}}^{\text{miss}}$, the p_{T} of the leptonically decaying W boson, p_{T}^{W}, the p_{T} of the lepton, p_{T}^{ℓ} and lepton pseudorapidity $|\eta^{\ell}|$.

The measurements are performed using collision data collected by the CMS detector at the LHC. The total amount of data used in this thesis corresponds to 35.9 fb^{-1}. Events containing a single, isolated electron or muon and at least four jets, of which two are tagged as originating from the b quark, are considered. Additional selections on the quality of the leptons and jets are applied. The $t\bar{t}$ yield is estimated by subtracting from data the V+jets and single top simulated backgrounds as well as the QCD contribution estimated from data. The measurements are presented to *particle level* where the kinematic distributions are constructed with respect to stable particles in the detector (mean lifetime longer than 30 ps) in a phase space similar to that accessible to the detector. The number of events reconstructed by the detector but not entering the visible phase space are subtracted from the $t\bar{t}$ yield, which are then unfolded to remove the effects of the detector acceptance, efficiency, and migrations between bins induced from the detector resolution.

Both the normalised and absolute cross sections are calculated with respect to the unfolded yields and compared to several state-of-the-art $t\bar{t}$ models, POWHEG+PYTHIA, POWHEG+HERWIG++, MG5_aMC@NLO-LO and MG5_aMC@NLO-NLO. Goodness-of-fit tests between the generators and measured cross section are performed and it is found that the POWHEG+PYTHIA model is generally consistent with the data, with any residual differ-

ences covered by theoretical uncertainties within the model. The POWHEG+HERWIG++ and MG5_aMC@NLO-NLO are consistent with the data for most of the kinematic event variables, whereas the MG5_aMC@NLO-LO model is found to not accurately describe any variable.

It is widely expected for these measurements to be used in the tuning of future $t\bar{t}$ generators and as such the measurements presented here have been implemented in RIVET and are available to the wider community.

Samenvatting

List of abbreviations

CMS	Compact Muon Solenoid
LHC	Large Hadron Collider
QCD	Quantum chromodynamics
SM	Standard model of particle physics
EM	Electromagnetic
BEH	Brout-Englert-Higgs
H	Brout-Englert-Higgs boson
ISR	Initial state radiation
FSR	Final state radiation
QED	Quantum electrodynamics
EWK	Electroweak
CKM	Cabbibo-Kobayashi-Maskawa
CP	Charge-Parity
SNO	Sudbury Neutrino Observatory
CDF	Collider Detector at Fermilab
D0	TODO
NNPDF	TODO
PDF	TODO
DGLAP	TODO
LO	TODO

NLO	TODO
NNLO	TODO
ATLAS	TODO
LHCb	TODO
ALICE	TODO
Powheg	TODO
MPI	TODO
MLM	TODO
FxFX	TODO
CR	TODO
Geant4	TODO
	TODO
	TODO
CERN	TODO
PS	TODO
PSB	TODO
SPS	TODO
LINAC2	TODO
ECAL	TODO
HCAL	TODO
TIB	TODO
TOB	TODO
TEC	TODO
TID	TODO
HB	TODO
HE	TODO
HO	TODO
HF	TODO
RPC	TODO
CSC	TODO

DT	TODO
L1T	TODO
HLT	TODO
WLCG	TODO
PF	TODO
GSF	TODO
	TODO
	TODO
	TODO
	TODO
	TODO

Bibliography

[1] CMS Collaboration, "Measurement of differential top quark pair production cross sections in a fiducial volume as a function of event variables in pp collisions at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-TOP-15-013, 2015.

[2] CMS Collaboration, "Measurements of differential cross sections of top quark pair production as a function of kinematic event variables in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **06** (2018) 002, doi: 10.1007/JHEP06(2018)002, `arXiv:1803.03991`.

[3] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi: 10.1103/PhysRevLett.19.1264.

[4] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588, doi: 10.1016/0029-5582(61)90469-2.

[5] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc.* **C680519** (1968) 367–377.

[6] M. Gell-Mann, "A Schematic Model of Baryons and Mesons", *Phys. Lett.* **8** (1964) 214–215, doi: 10.1016/S0031-9163(64)92001-3.

[7] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking. Version 2", *Developments in the Quark Theory of Hadrons* **1** (1964) 22–101. Edited by D. Lichtenberg and S. P. Rosen.

[8] Particle Data Group, C. Patrignani et al., "Review of particle physics", *Chin. Phys. C.* **40** (2016) 100001, doi: 10.1088/1674-1137/40/10/100001.

[9] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–323, doi: 10.1103/PhysRevLett.13.321.

[10] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, doi: 10.1103/PhysRevLett.13.508.

[11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585–587, doi: 10.1103/PhysRevLett.13.585.

[12] R. P. Feynman, "The Theory of positrons", *Phys. Rev.* **76** (1949) 749–759, doi: 10.1103/PhysRev.76.749.

[13] R. P. Feynman, "Space - time approach to quantum electrodynamics", *Phys. Rev.* **76** (1949) 769–789, doi: 10.1103/PhysRev.76.769.

[14] J. Ellis, "TikZ-Feynman: Feynman diagrams with TikZ", *Comput. Phys. Commun.* **210** (2017) 103–123, doi: 10.1016/j.cpc.2016.08.019, `arXiv:1601.05437`.

[15] E. Noether, "Invariant Variation Problems", *Transport Theory Statistics* (1971) 186–207, doi: 10.1080/00411457108231446, `arXiv:physics/0503066`. Translated by M. Tavel.

[16] CMS Collaboration, "Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV", *JHEP* **03** (2017) 156, doi: 10.1007/JHEP03(2017)156, `arXiv:1609.05331`.

[17] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.* **10** (1963) 531–533, doi: 10.1103/PhysRevLett.10.531.

[18] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.* **49** (1973) 652–657, doi: 10.1143/PTP.49.652.

[19] SNO Collaboration, G. A. McGregor, "First results from the Sudbury Neutrino Observatory", in *Proceedings, 37th Rencontres de Moriond on Electroweak Interactions and Unified Theories*, pp. 249–256. 2002. `arXiv:nucl-ex/0205006`.

[20] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.* **81** (1998) 1562–1567, doi: 10.1103/PhysRevLett.81.1562, `arXiv:hep-ex/9807003`.

[21] V. C. Rubin, N. Thonnard, and W. K. Ford, Jr., "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/", *Astrophys. J.* **238** (1980) 471, doi: 10.1086/158003.

[22] F. Zwicky, "On the Masses of Nebulae and of Clusters of Nebulae", *Astrophys. J.* **86** (1937) 217, doi: 10.1086/143864.

[23] Planck Collaboration, "Planck 2013 results. XVI. Cosmological parameters", *Astron. Astrophys.* **571** (2014) A16, doi: 10.1051/0004-6361/201321591, `arXiv:1303.5076`.

[24] Supernova Search Team Collaboration, "Observational evidence from supernovae for an accelerating universe and a cosmological constant", *Astron. J.* **116** (1998) 1009–1038, doi: 10.1086/300499, `arXiv:astro-ph/9805201`.

[25] M. Dine and A. Kusenko, "The Origin of the matter - antimatter asymmetry", *Rev. Mod. Phys.* **76** (2003) 1, doi: 10.1103/RevModPhys.76.1, `arXiv:hep-ph/0303065`.

[26] J. F. Gunion, "A Simplified summary of supersymmetry", *AIP Conf. Proc.* **397** (1997) 41–64, doi: 10.1063/1.52979, `arXiv:hep-ph/9704349`.

[27] CDF Collaboration, "Observation of top quark production in $\overline{p}p$ collisions", *Phys. Rev. Lett.* **74** (1995) 2626–2631, doi: 10.1103/PhysRevLett.74.2626, `arXiv:hep-ex/9503002`.

[28] D0 Collaboration, "Observation of the top quark", *Phys. Rev. Lett.* **74** (1995) 2632–2637, doi: 10.1103/PhysRevLett.74.2632, `arXiv:hep-ex/9503003`.

[29] 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, doi: 10.1007/JHEP07(2014)079, `arXiv:1405.0301`.

[30] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J. C* **77** (2017) 663, doi: 10.1140/epjc/s10052-017-5199-5, `arXiv:1706.00428`.

[31] 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. [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[32] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 438–450. [Yad. Fiz.15,781(1972)].

[33] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys. B* **126** (1977) 298–318, doi: 10.1016/0550-3213(77)90384-4.

[34] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", *Comput. Phys. Commun.* **185** (2014) 2930, doi: 10.1016/j.cpc.2014.06.021, `arXiv:1112.5675`.

[35] M. Beneke, P. Falgari, S. Klein, and C. Schwinn, "Hadronic top-quark pair production with NNLL threshold resummation", *Nucl. Phys. B* **855** (2012) 695, doi: 10.1016/j.nuclphysb.2011.10.021, `arXiv:1109.1536`.

[36] M. Cacciari et al., "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation", *Phys. Lett. B* **710** (2012) 612, doi: 10.1016/j.physletb.2012.03.013, `arXiv:1111.5869`.

[37] P. Bärnreuther, M. Czakon, and A. Mitov, "Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $\mathrm{q\overline{q}} \to \mathrm{t\overline{t}} + \mathrm{X}$", *Phys. Rev. Lett.* **109** (2012) 132001, doi: 10.1103/PhysRevLett.109.132001, `arXiv:1204.5201`.

[38] M. Czakon and A. Mitov, "NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels", *JHEP* **12** (2012) 054, doi: 10.1007/JHEP12(2012)054, `arXiv:1207.0236`.

[39] M. Czakon and A. Mitov, "NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction", *JHEP* **01** (2013) 080, doi: 10.1007/JHEP01(2013)080, `arXiv:1210.6832`.

[40] M. Czakon, P. Fiedler, and A. Mitov, "Total top-quark pair-production cross section at hadron colliders through $\mathrm{O}(\alpha_S^4)$", *Phys. Rev. Lett.* **110** (2013) 252004, doi: 10.1103/PhysRevLett.110.252004, `arXiv:1303.6254`.

[41] M. Czakon and A. Mitov, "Top++: A program for the calculation of the top-pair cross-section at hadron colliders", *Comput. Phys. Commun.* **185** (2014) 2930, doi: 10.1016/j.cpc.2014.06.021, `arXiv:1112.5675`.

[42] "LHCTopWG: Summary Plots". `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots`.

[43] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section using events in the eμ final state in pp collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **77** (2017) 172, doi: 10.1140/epjc/s10052-017-4718-8, `arXiv:1611.04040`.

[44] 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, doi: 10.1007/JHEP09(2017)051, `arXiv:1701.06228`.

[45] CMS Collaboration, "First Measurement of the Cross Section for Top-Quark Pair Production in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett. B* **695** (2011) 424–443, doi: 10.1016/j.physletb.2010.11.058, `arXiv:1010.5994`.

[46] CMS Collaboration, "Measurement of the Top-antitop Production Cross Section in *pp* Collisions at $\sqrt{s} = 7$ TeV using the Kinematic Properties of Events with Leptons and Jets", *Eur. Phys. J. C* **71** (2011) 1721, doi: 10.1140/epjc/s10052-011-1721-3, `arXiv:1106.0902`.

[47] CMS Collaboration, "Measurement of the $t\bar{t}$ Production Cross Section in *pp* Collisions at 7 TeV in Lepton + Jets Events Using *b*-quark Jet Identification", *Phys. Rev. D* **84** (2011) 092004, doi: 10.1103/PhysRevD.84.092004, `arXiv:1108.3773`.

[48] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section and the top quark mass in the dilepton channel in *pp* collisions at $\sqrt{s} = 7$ TeV", *JHEP* **07** (2011) 049, doi: 10.1007/JHEP07(2011)049, `arXiv:1105.5661`.

[49] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in *pp* collisions at $\sqrt{s} = 7$ TeV with lepton + jets final states", *Phys. Lett. B* **720** (2013) 83–104, doi: 10.1016/j.physletb.2013.02.021, `arXiv:1212.6682`.

[50] CMS Collaboration, "Measurement of the top-antitop production cross section in the tau+jets channel in pp collisions at sqrt(s) = 7 TeV", *Eur. Phys. J. C* **73** (2013) 2386, doi: 10.1140/epjc/s10052-013-2386-x, `arXiv:1301.5755`.

[51] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the dilepton channel in *pp* collisions at $\sqrt{s} = 7$ TeV", *JHEP* **11** (2012) 067, doi: 10.1007/JHEP11(2012)067, `arXiv:1208.2671`.

[52] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **02** (2014) 024, doi: 10.1007/JHEP02(2014)024,10.1007/JHEP02(2014)102, `arXiv:1312.7582`.

[53] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in *pp* collisions at $\sqrt{s} = 8$ TeV in dilepton final states containing one τ lepton", *Phys. Lett. B* **739** (2014) 23–43, doi: 10.1016/j.physletb.2014.10.032, `arXiv:1407.6643`.

[54] CMS Collaboration, "Measurement of $t\bar{t}$ production with additional jet activity, including b quark jets, in the dilepton decay channel using pp collisions at $\sqrt{s} = 8$ TeV", *Eur. Phys. J. C* **76** (2016) 379, doi: 10.1140/epjc/s10052-016-4105-x, `arXiv:1510.03072`.

[55] 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, doi: 10.1007/JHEP08(2016)029, `arXiv:1603.02303`.

[56] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the all-jets final state in pp collisions at $\sqrt{s} = 8$ TeV", *Eur. Phys. J. C* **76** (2016) 128, doi: 10.1140/epjc/s10052-016-3956-5, `arXiv:1509.06076`.

[57] CMS Collaboration, "Measurement of the inclusive $t\bar{t}$ cross section in pp collisions at $\sqrt{s} = 5.02$ TeV using final states with at least one charged lepton", *JHEP* **03** (2018) 115, doi: 10.1007/JHEP03(2018)115, `arXiv:1711.03143`.

[58] CMS Collaboration, "Measurement of the semileptonic $t\bar{t} + \gamma$ production cross section in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **10** (2017) 006, doi: 10.1007/JHEP10(2017)006, `arXiv:1706.08128`.

[59] CMS Collaboration, "Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV", `arXiv:1711.02547`.

[60] CMS Collaboration, "Observation of $t\bar{t}$H production", *Phys. Rev. Lett.* **120** (2018) 231801, doi: 10.1103/PhysRevLett.120.231801,10.1130/PhysRevLett.120.231801, `arXiv:1804.02610`.

[61] CMS Collaboration, "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-16-021, 2016.

[62] CMS Collaboration, "Measurement of differential top-quark pair production cross sections in *pp* colisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J. C* **73** (2013) 2339, doi: 10.1140/epjc/s10052-013-2339-4, `arXiv:1211.2220`.

[63] CMS Collaboration, "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8\,\text{TeV}$", *Eur. Phys. J. C* **75** (2015) 542, doi: 10.1140/epjc/s10052-015-3709-x, `arXiv:1505.04480`.

[64] CMS Collaboration, "Measurement of the integrated and differential $t\bar{t}$ production cross sections for high-p_T top quarks in *pp* collisions at $\sqrt{s} = 8$ TeV", *Phys. Rev. D* **94** (2016) 072002, doi: 10.1103/PhysRevD.94.072002, `arXiv:1605.00116`.

[65] CMS Collaboration, "Measurement of double-differential cross sections for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV and impact on parton distribution functions", *Eur. Phys. J. C* **77** (2017) 459, doi: 10.1140/epjc/s10052-017-4984-5, `arXiv:1703.01630`.

[66] CMS Collaboration, "Measurement of normalized differential $t\bar{t}$ cross sections in the dilepton channel from pp collisions at $\sqrt{s} = 13$ TeV", *JHEP* **04** (2018) 060, doi: 10.1007/JHEP04(2018)060, `arXiv:1708.07638`.

[67] CMS Collaboration, "Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV", *Phys. Rev. D* **95** (2017) 092001, doi: 10.1103/PhysRevD.95.092001, `arXiv:1610.04191`.

[68] CMS Collaboration, "Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* **776** (2018) 355–378, doi: 10.1016/j.physletb.2017.11.043, `arXiv:1705.10141`.

[69] CMS Collaboration, "Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at $\sqrt{s}$ = 13 TeV", *Phys. Rev. D* **97** (2018), no. 11, 112003, doi: 10.1103/PhysRevD.97.112003, arXiv:1803.08856.

[70] CMS Collaboration, "Measurement of the differential cross sections for top quark pair production as a function of kinematic event variables in pp collisions at $\sqrt{s}$=7 and 8 TeV", *Phys. Rev. D* **94** (2016) 052006, doi: 10.1103/PhysRevD.94.052006, arXiv:1607.00837.

[71] "Sherpa and Open Science Grid: Predicting the emergence of jets". https://sciencenode.org/feature/sherpa-and-open-science-grid-predicting-emergence-jets.php.

[72] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi: 10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[73] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi: 10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[74] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi: 10.1007/JHEP06(2010)043, arXiv:1002.2581.

[75] T. Sjöstrand, "The Development of MPI Modelling in PYTHIA", arXiv:1706.02166.

[76] Höche, Stefan, "Introduction to parton-shower event generators", in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*, pp. 235–295. 2015. arXiv:1411.4085. doi: 10.1142/9789814678766_0005.

[77] J. C. Collins, "Sudakov form-factors", *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 573–614, doi: 10.1142/9789814503266_0006, arXiv:hep-ph/0312336.

[78] Sjöstrand, T. and Skands, Peter Z., "Transverse-momentum-ordered showers and interleaved multiple interactions", *Eur. Phys. J. C* **39** (2005) 129–154, doi: 10.1140/epjc/s2004-02084-y, arXiv:hep-ph/0408302.

[79] G. Marchesini and B. R. Webber, "Simulation of QCD Jets Including Soft Gluon Interference", *Nucl. Phys. B* **238** (1984) 1–29, doi: 10.1016/0550-3213(84)90463-2.

[80] T. Sjöstrand et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, doi: 10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[81] M. Bahr et al., "Herwig++ physics and manual", *Eur. Phys. J. C* **58** (2008) 639, doi: 10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.

[82] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473–500, doi: 10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[83] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi: 10.1007/JHEP12(2012)061, arXiv:1209.6215.

[84] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton Fragmentation and String Dynamics", *Phys. Rept.* **97** (1983) 31–145, doi: 10.1016/0370-1573(83)90080-7.

[85] Sjöstrand, Torbjörn, "Status and developments of event generators", *PoS* **LHCP2016** (2016) 007, `arXiv:1608.06425`.

[86] Argyropoulos, Spyros and Sjöstrand, Torbjörn, "Effects of color reconnection on $t\bar{t}$ final states at the LHC", *JHEP* **11** (2014) 043, doi: 10.1007/JHEP11(2014)043, `arXiv:1407.6653`.

[87] J. R. Christiansen and P. Z. Skands, "String Formation Beyond Leading Colour", *JHEP* **08** (2015) 003, doi: 10.1007/JHEP08(2015)003, `arXiv:1505.01681`.

[88] "Event generation and parton shower". `https://www.gk-eichtheorien.physik.uni-mainz.de/Dateien/Zeppenfeld-3.pdf`.

[89] GEANT4 Collaboration, "Geant4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi: 10.1016/S0168-9002(03)01368-8.

[90] S. Frixione, P. Nason, and G. Ridolfi, "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", *JHEP* **09** (2007) 126, doi: 10.1088/1126-6708/2007/09/126, `arXiv:0707.3088`.

[91] S. Gieseke, C. Rohr, and A. Siodmok, "Colour reconnections in Herwig++", *Eur. Phys. J. C* **72** (2012) 2225, doi: 10.1140/epjc/s10052-012-2225-5, `arXiv:1206.0041`.

[92] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, doi: 10.1140/epjc/s10052-016-3988-x, `arXiv:1512.00815`.

[93] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi: 10.1007/JHEP04(2015)040, `arXiv:1410.8849`.

[94] 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, doi: 10.1007/JHEP09(2012)130, `arXiv:1207.5391`.

[95] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J. C* **71** (2011) 1547, doi: 10.1140/epjc/s10052-011-1547-z, `arXiv:1009.2450`.

[96] M. Aliev et al., "HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR", *Comput. Phys. Commun.* **182** (2011) 1034, doi: 10.1016/j.cpc.2010.12.040, `arXiv:1007.1327`.

[97] P. Kant et al., "HATHOR for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions", *Comput. Phys. Commun.* **191** (2015) 74, doi: 10.1016/j.cpc.2015.02.001, `arXiv:1406.4403`.

[98] Y. Li and F. Petriello, "Combining QCD and electroweak corrections to dilepton production in FEWZ", *Phys. Rev. D* **86** (2012) 094034, doi: 10.1103/PhysRevD.86.094034, `arXiv:1208.5967`.

[99] F. Marcastel, "CERN's Accelerator Complex. La chaîne des accélérateurs du CERN", Technical Report OPEN-PHO-CHART-2013-001, 2013.

[100] "The Large Hadron Collider". https://timeline.web.cern.ch/timelines/The-Large-Hadron-Collider.

[101] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, doi: 10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[102] A. Team, "Diagram of an LHC dipole magnet. Schéma d'un aimant dipôle du LHC", (1999).

[103] "CMS Luminosity - Public Results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[104] CMS Collaboration, "The CMS experiment at the CERN LHC.", *JINST* **3** (2008) S08004. 361 p.

[105] T. Sakuma and T. McCauley, "Detector and Event Visualization with SketchUp at the CMS Experiment", *Journal of Physics: Conference Series* **513** (2014) 022032.

[106] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, doi: 10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[107] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". Technical Design Report CMS. CERN, Geneva, 2006.

[108] CMS Collaboration, "Precision crystal calorimetry in LHC Run II with the CMS ECAL", *JINST* **12** (2017) C01069, doi: 10.1088/1748-0221/12/01/C01069.

[109] CMS Collaboration, "The CMS-HF quartz fiber calorimeters", *Journal of Physics: Conference Series* **160** (2009) 012014.

[110] S. Abdullin et al., "The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c", *Eur. Phys. J. C* **60** (2009) 359–373, doi: 10.1140/epjc/s10052-009-0959-5,10.1140/epjc/s10052-009-1024-0.

[111] CMS Collaboration, "The CMS muon system in Run2: preparation, status and first results", *PoS* **EPS-HEP2015** (2015) 237, arXiv:1510.05424.

[112] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade", Technical Report CERN-LHCC-2013-011. CMS-TDR-12, 2013.

[113] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, doi: 10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[114] W. Adam, B. Mangano, T. Speer, and T. Todorov, "Track reconstruction in the CMS tracker",.

[115] Frühwirth, R and Waltenberger, Wolfgang and Vanlaer, Pascal, "Adaptive Vertex Fitting", Technical Report CMS-NOTE-2007-008, CERN, 2007.

[116] CMS Collaboration, "CMS Tracking Performance Results from early LHC Operation", *Eur. Phys. J. C* **70** (2010) 1165–1192, doi: 10.1140/epjc/s10052-010-1491-3, arXiv:1007.1988.

[117] CMS Collaboration, "Studies of Tracker Material", Technical Report CMS-PAS-TRK-10-003, 2010.

[118] Adam, Wolfgang and Frühwirth, R. and Strandlie, Are and Todor, T., "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC",.

[119] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_{T} jet clustering algorithm", *JHEP* **04** (2008) 063, doi: 10.1088/1126-6708/2008/04/063, `arXiv:0802.1189`.

[120] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018) P05011, doi: 10.1088/1748-0221/13/05/P05011, `arXiv:1712.07158`.

[121] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017), no. 02, P02014, doi: 10.1088/1748-0221/12/02/P02014, `arXiv:1607.03663`.

[122] CMS Collaboration, "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016", CMS Detector Performance CMS-DP-2018-028, 2018.

[123] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P06005, doi: 10.1088/1748-0221/10/06/P06005, `arXiv:1502.02701`.

[124] CMS Collaboration, "Electron and photon performance in CMS with the full 2016 data sample",.

[125] CMS Collaboration, "Muon Identification and Isolation efficiency on full 2016 dataset",.

[126] CMS Collaboration, "Muon HLT performance on 2016 data",.

[127] A. Buckley et al., "Rivet user manual", *Comput. Phys. Commun.* **184** (2013) 2803–2819, doi: 10.1016/j.cpc.2013.05.021, `arXiv:1003.0694`.

[128] CMS Collaboration Collaboration, "Object definitions for top quark analyses at the particle level", Technical Report CMS-NOTE-2017-004. CERN-CMS-NOTE-2017-004, 2017.

[129] S. Schmitt, "TUnfold: an algorithm for correcting migration effects in high energy physics", *JINST* **7** (2012) T10003, doi: 10.1088/1748-0221/7/10/T10003, `arXiv:1205.6201`.

[130] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period", CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

[131] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV", (2018). `arXiv:1802.02613`. Submitted to *JHEP*.

[132] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune", *Eur. Phys. J. C* **74** (2014) 3024, doi: 10.1140/epjc/s10052-014-3024-y, `arXiv:1404.5630`.

[133] C. Peterson, D. Schlatter, I. Schmitt, and P. M. Zerwas, "Scaling violations in inclusive e^+e^- annihilation spectra", *Phys. Rev. D* **27** (1983) 105, doi: 10.1103/PhysRevD.27.105.

[134] CMS Collaboration, "Cross section measurement of t-channel single top quark production in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* **772** (2017) 752, doi: 10.1016/j.physletb.2017.07.047, `arXiv:1610.00678`.

[135] CMS Collaboration, "Measurement of the production cross section of a W boson in association with two b jets in pp collisions at $\sqrt{s} = 8\,\text{TeV}$", *Eur. Phys. J. C* **77** (2017) 92, doi: 10.1140/epjc/s10052-016-4573-z, `arXiv:1608.07561`.

[136] CMS Collaboration, "Measurements of the associated production of a Z boson and b jets in pp collisions at $\sqrt{s} = 8\,\text{TeV}$", *Eur. Phys. J. C* **77** (2017) 751, doi: 10.1140/epjc/s10052-017-5140-y, `arXiv:1611.06507`.

[137] C. Zhang and S. Willenbrock, "Effective-Field-Theory Approach to Top-Quark Production and Decay", *Phys. Rev. D* **83** (2011) 034006, doi: 10.1103/PhysRevD.83.034006, `arXiv:1008.3869`.

[138] D. Barducci et al., "Interpreting top-quark LHC measurements in the standard-model effective field theory", `arXiv:1802.07237`.

[139] CMS Collaboration, "Measurements of differential cross sections for $\mathrm{t\bar{t}}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons",.

[140] D. Buarque Franzosi and C. Zhang, "Probing the top-quark chromomagnetic dipole moment at next-to-leading order in QCD", *Phys. Rev. D* **91** (2015) 114010, doi: 10.1103/PhysRevD.91.114010, `arXiv:1503.08841`.